

# Network Inference Analysis Identifies SETDB1 as a Key Regulator for Reverting Colorectal Cancer Cells into Differentiated Normal-Like Cells

Soobeom Lee<sup>1</sup>, Chansu Lee<sup>2</sup>, Chae Young Hwang<sup>1</sup>, Dongsan Kim<sup>1</sup>, Younghyun Han<sup>1</sup>, Sung Noh Hong<sup>2</sup>, Seok-Hyung Kim<sup>3</sup>, and Kwang-Hyun Cho<sup>1</sup>



## ABSTRACT

Cancer cells exhibit properties of cells in a less differentiated state than the adjacent normal cells in the tissue. We explored whether cancer cells can be converted to a differentiated normal-like state by restoring the gene regulatory network (GRN) of normal cells. Here, we report that colorectal cancer cells exhibit a range of developmental states from embryonic and intestinal stem-like cells to differentiated normal-like cells. To identify the transcription factors (TF) that commit stem-like colorectal cancer cells into a differentiated normal-like state, we reconstructed GRNs of normal colon mucosa and identified core TFs (CDX2, ELF3, HNF4G, PPARG, and VDR) that govern the cellular state. We further found that SET Domain Bifurcated 1 (SETDB1), a histone H3 lysine 9-specific

methyltransferase, hinders the function of the identified TFs. SETDB1 depletion effectively converts stem-like colorectal cancer cells into postmitotic cells and restores normal morphology in patient-derived colorectal cancer organoids. RNA-sequencing analyses revealed that SETDB1 depletion recapitulates global gene expression profiles of normal differentiated cells by restoring the transcriptional activity of core TFs on their target genes.

**Implications:** Our study provides insights into the molecular regulatory mechanism underlying the developmental hierarchy of colorectal cancer and suggests that induction of a postmitotic state may be a therapeutic alternative to destruction of cancer cells.

## Introduction

Dysregulation of tissue-specific gene expression programs is a hallmark of cancer (1). Such dysregulation leads to cancer-promoting gene expression programs and, in various types of cancer, the consequent reprogramming of cells into those with stem or progenitor (stem/progenitor) properties (2–8). A study of cancer initiation revealed that normal differentiated cells with oncogenic mutations remain in a nonmalignant state until they undergo cellular reprogramming into a stem/progenitor state (5). This suggests that differentiated cells have an inherent resistance mechanism against malignant transformation and indicates that cellular reprogramming is indispensable for malignancy. Thus, we speculated that malignant properties might be eradicated if the tissue-specific gene expression program is reinstated.

In colorectal cancer, cellular differentiation is impeded through processes involving both oncogenic mutations and microenvironmental alterations (9, 10). This cancer provides a model for exploring whether the malignant cells could be converted to normal-like cells

through restoration of the tissue-specific gene expression program. To address this challenge in a systematic way, we employed a computational framework to identify the core factors to revert cancer cells back to their normal state. A recent computational framework for inferring gene regulatory networks (GRN) has effectively applied to the cell fate conversion study through identification of master regulators of tissue-specific gene expression programs (11–13).

Here, we reconstructed normal colon-specific GRNs and colorectal cancer-specific GRNs, and identified core transcription factors (TF) for differentiation of colorectal cancer cells. We further identified SET Domain Bifurcated 1 (SETDB1) as a key factor that hinders the function of core TFs. We demonstrated that SETDB1 depletion effectively reestablishes the normal colon-specific gene expression profile and induces a postmitotic differentiated state in three stem-like colorectal cancer cell lines and patient-derived colon cancer organoids by recapitulating the transcriptional activities of the core TFs.

## Materials and Methods

### Cell culture

Caco2, HCT116, SW480, and SW620 human colorectal cancer cells were cultured in DMEM supplemented with 10% FBS and antibiotics (100 U/mL of penicillin, 100 µg/mL of streptomycin, and 0.25 µg/mL of Fungizone; Life Technologies) at 37°C in 5% CO<sub>2</sub>. The Caco2 and HCT116 cell lines were obtained from Korean Cell Line Bank, and SW480 and SW620 cell lines were obtained from the ATCC. For retrovirus transduction, we cultured cells in above condition with 2 µg/mL of puromycin for 3 days after transduction and then cultured with 1 µg/mL for the rest of the period. Three days after transduction, to activate PPARG and vitamin D receptor (VDR), we supplemented troglitazone (10 µmol/L; T2573, Sigma-Aldrich) and 1 $\alpha$ ,25-dihydroxyvitamin D<sub>3</sub> (25 nmol/L; D1530, Sigma-Aldrich). We called this supplemented medium differentiation medium in the study. All used cell lines were authenticated by short tandem repeat analysis (Korean Cell Line Bank) and were verified to be *Mycoplasma* negative using a

<sup>1</sup>Laboratory for Systems Biology and Bio-inspired Engineering, Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea. <sup>2</sup>Department of Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea. <sup>3</sup>Department of Pathology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea.

**Note:** Supplementary data for this article are available at Molecular Cancer Research Online (<http://mcr.aacrjournals.org/>).

**Corresponding Author:** Kwang-Hyun Cho, Korea Advanced Institute of Science and Technology (KAIST), 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea. Phone: 824-2350-4325; Fax: 824-2350-4310; E-mail: [ckh@kaist.ac.kr](mailto:ckh@kaist.ac.kr)

Mol Cancer Res 2020;18:118–29

doi: 10.1158/1541-7786.MCR-19-0450

©2020 American Association for Cancer Research.

Mycoplasma PCR detection kit (Intron). The cells were typically used for experiments within 2 weeks after thawing, and retrovirus-mediated experiments were completed within 2 to 3 weeks after transduction.

### Flow cytometry

*In vitro* cultured cells were fixed at 4% PFA and stained with KRT20 antibody (13063S, Cell Signaling Technology) and FITC-anti-rabbit IgG (554020, BD Pharmingen) according to the manufacturer's instruction (catalog no. 554714; BD Biosciences). To isolate RNA from intracellular immunofluorescent stained cells, we performed the experiments under RNase-free condition with 1:100 RNasin Plus RNase Inhibitor (N2615, Promega) following the method for analyzing RNA following intracellular sorting (MARIS) protocol (14). Flow cytometric analysis was conducted with a FACS Jazz (BD Biosciences), and analysis was performed with BD FACSDiva software.

### Immunofluorescence and image analysis

For immunofluorescence (IF) staining, cancer cells were fixed for 15 minutes in 4% (vol/vol) formaldehyde and blocked for 1 hour in 2% BSA with 0.1% Triton-X100 of PBS (blocking and staining buffer), followed by incubation in primary antibody, KRT20 (13063S, Cell Signaling Technology), and Ki-67 (11882S, Cell Signaling Technology) at a dilution of 1:200 for staining buffer for overnight at 4°C. Samples were washed in PBS and incubated with the secondary antibody (ab150113, ab150080) at a dilution of 1:300 in staining buffer for 1 hour. 4',6-Diamidino-2-phenylindole (DAPI; D1306, Invitrogen) was stained for 1 hour before the confocal microscopic analysis. Imaging was performed using a Nikon A1R confocal microscope (Nikon Instruments) with CFI plan Apo objectives under 40× magnification and digital-zooming Nikon imaging software (NIS-element AR). Image data were analyzed by Nikon imaging software (NIS-element AR), and the number of Ki-67-positive cells was counted using MATLAB.

### shRNA and siRNA knockdown

Specific siRNAs against *EHMT2*, *HDAC2*, *KAT2A*, *KDM1A*, *PRMT1*, *SETDB1*, *SMYD2*, and *SMYD3* were custom-synthesized (Bioneer) according to sequences provided from the previous studies. We used Scramble siRNA (Bioneer) as a control, and transfection was carried out using Lipofectamine RNAiMAX (Invitrogen). Retrovirus-mediated short hairpin RNA (shRNA) was purchased from Sigma-Aldrich. Detailed information on siRNA and shRNA experiments is given in Supplementary Table S5.

### Plasmid construction

For extrinsic expression of *CDX2*, *ELF3*, *HNF4G*, *PPARG*, and *VDR*, the full-length cDNAs of each gene were amplified by RT-PCR from human colon cDNA library (Clontech). Each cDNA was ligated into pLentiM1.4 lentiviral vector and confirmed by sequencing.

### Retrovirus production and transduction

For retrovirus-mediated gene transfer, HEK293T cells were transfected with each relevant lentiviral plasmid and packaging plasmids (pLP1, pLP2, and pLP/VSV-G, Invitrogen) using Lipofectamine (Invitrogen) according to the manufacturer's instructions. The medium was harvested after 48 hours of transfection and filtered through a 0.22-μm cellulose acetate filter. Then, virus-containing medium was centrifuged at 25,000 rpm for 90 minutes (Optima L90K; Beckman). Lentiviral-packaged *CDX2*, *ELF3*, *HNF4G*, *PPARG*, and *VDR* were concentrated in 2 mL of complete DMEM 10% supplemented with 4 μg/mL polybrene and infected to HCT116 and SW480 cell lines.

Lentiviral-packaged shSETDB1 was infected to colon cancer organoids with concentration and infected to Caco2, HCT116, and SW480 without concentration.

### Colony-forming assay

For colony-forming assay, 500 cells were seeded in 6-well plate and incubated for 14 days in medium supplemented with troglitazone (10 μmol/L, T2573; Sigma-Aldrich) and 1α,25-dihydroxyvitamin D<sub>3</sub> (25 nmol/L, D1530; Sigma-Aldrich). Cells were fixed with 3.7% paraformaldehyde for 15 minutes, followed by staining with a 0.5% crystal violet solution for 1 hour. After washing with distilled water and air drying, images were obtained.

### Cell growth assay

For assessing cell growth, transfected cells were seeded in 96-well plate (5 × 10<sup>3</sup> cells/well) and imaged every 3 hours using IncuCyte ZOOM (Essen Bioscience). The confluence was analyzed by the software IncuCyte ZOOM 2016A.

### Chromatin immunoprecipitation assays

Chromatin immunoprecipitation (ChIP) assays were performed using a High-Sensitivity ChIP assay kit (Abcam). Briefly, Caco2 cells (2 × 10<sup>6</sup>) were fixed with 1% formaldehyde at room temperature for 10 minutes. To shear chromatin lysates, sonication was carried out with Bioruptor sonicator (20 seconds on, 30 seconds off at 180 W; Biomedical Science). Chromatin samples were incubated with 2 μg anti-H3K9me3 antibody (ab8898, Abcam) or control IgG at 4°C overnight.

### Human colon cancer organoid culture

Human samples were obtained from patients referred to centers in Samsung Medical Center. Patients agreed to donate colon biopsy samples for molecular testing on research base with written informed consent and analyzed. The experiments were conducted according to protocols approved by the Institutional Review Board of Samsung Medical Center (Seoul, Republic of Korea). Human cancer cells were isolated from colon cancer biopsy samples. Briefly, the biopsies were minced under 1-mm fragments with a sterile scalpel or scissors and then incubated in Advanced DMEM/Ham F-12 media (Thermo Fisher Scientific) containing 40 μmol/L collagenase 1 (Thermo Fisher Scientific) and 100 μmol/L Dispase (Thermo Fisher Scientific) with gentle shaking at 37°C water bath for 2 hours. After incubation, the solution was passed through a 40-μm pore cell strainer (BD Biosciences). The cancer cells were cultured in a three-dimensional Matrigel matrix (BD Biosciences) in Advanced DMEM/Ham F-12 containing 1× antibiotic-antimycotic, 2 mmol/L Glutamax (Invitrogen), 10 mmol/L HEPES (Invitrogen), 1 × N2 (Invitrogen), 1 × B27 (Invitrogen), 50 ng/mL EGF, 1 mmol/L N-acetylcysteine (Sigma-Aldrich), 100 ng/mL recombinant murine Noggin (PeproTech), 10 mmol/L Nicotinamide (Sigma-Aldrich), 10 nmol/L SB202190 (Sigma-Aldrich), 10 nmol/L Gastrin 1 (Sigma-Aldrich), 10 nmol/L Prostaglandin E2 (Cayman Chemical Company), 500 nmol/L A-83-01 (Sigma-Aldrich), 500 ng/mL recombinant human R-spondin1 (R&D Systems), and 100 ng/mL recombinant Wnt3a (R&D Systems). Culture medium was replaced every 2 days.

### IHC

A total of 296 tissue samples were obtained from patients with colorectal cancer after ethical approval by the Institutional Review Board of Samsung Medical Center (Seoul, Republic of Korea). The clinical data included overall survival and disease-free survival.

Immunostaining for SETDB1 protein was performed using an anti-SETDB1 polyclonal antibody (HPA018142; Sigma-Aldrich). The IHC scores were determined as follows: weak staining in <20% or moderate staining in <20% of cancer cells (Score 1); weak staining in  $\geq 50\%$ , moderate staining in 20% to 50%, or strong staining in <20% (Score 2); moderate staining in  $\geq 50\%$  or strong staining in  $\geq 20\%$  (Score 3). The experiment and analysis were performed according to the previous study (15).

### Gene set enrichment analysis

Gene set enrichment analysis was performed using “GSVA” package (16) in R and GSEA code provided from Broad Institute. Gene set variation analysis (GSVA) implements a nonparametric, unsupervised method estimating variation of gene set enrichment over the samples without class labels. Each sample was independently assessed, and normal and cancer samples were compared across the datasets. The statistical significance is determined by *P* value of moderated *t* statistics using “limma” package (17) in R. GSEA was performed to analyze RNA-sequencing (RNA-seq) data of three cell lines, evaluating the enrichment scores with distinct sample classes defined by experimental condition.

### Differentially expressed gene analysis

The differentially expressed gene (DEG) analysis was performed by using significance analysis of microarrays method (18) implemented in the “siggenes” package (19) in R. A delta value was set in such a way that FDR became zero in Gene Expression Omnibus (GEO)-integrated (delta = 8.1) and GEO-paired datasets (delta = 15.9), respectively. The common genes that showed higher expression in normal colon mucosa compared with colorectal cancer tissues across all three datasets were defined as normal-specific genes.

### Survival analysis

In the Kaplan–Meier survival analysis, the survival data were available for 633 patients with colorectal cancer from The Cancer Genome Atlas (20). Clinical information about overall survival and disease-free survival was downloaded from cBioPortal. We classified patients into two groups according to the median value of GSVA score of a colon-specific gene set. In addition, we further divided the groups with the third quantile of *SETDB1* gene expression level. The analysis was performed using “survival” package in R (21). Statistical significance was assessed using the log-rank test.

### Pseudotime analysis

To reconstruct cellular trajectories of colorectal cancer, we used Selective Locally Linear Inference of Cellular Expression Relationships (SLICER) algorithm (22) and Monocle2 (23). SLICER performs nonlinear dimensionality reduction using locally linear embedding algorithm which calculates geodesic distances among the samples. The genes measured to more than one tenth of the total single-cell RNA-seq samples were used to conduct dimensionality reduction in SLICER. For Monocle2, the DEGs across normal and cancer samples were determined with 0.01 cutoff of *q* value using the method implemented in “monocle” package. Discriminative Dimensionality Reduction with Trees method was used for performing Monocle2. The first and second dimensions and pseudotime information were used for further data analysis.

### Cell-cycle analysis

Gene sets representing  $G_1$ -S and  $G_2$ -M phases of the cell cycle were defined in previous study (24). As the previous study performed, we

calculated the average expression of each gene set, and the cells of each phase were determined by the average expression value above 3.0 which results in the *P* values of 1.5e-10 and 2.0e-16, respectively, when assigned samples and remained samples were statistically compared.

### Transcription entropy analysis

The single-cell RNA-seq data were used as a read counts format. Median normalization was performed using RaceID2 algorithm (25) in R. The transcription entropy is defined as general formula for the information entropy where each probability indicates the number of transcripts of a single gene in a cell. The analysis was performed according to the previous study (25).

### Principal component analysis of single-cell RNA-seq data

Single-cell RNA-seq data have substantial noise, and it led to inconsistent results of GSEA depending on the gene set composition. Thus, to robustly estimate the state-specific gene expression profiles, we applied imputation algorithm MAGIC (26) to recover the dropouts and performed principal component analysis (PCA) with the union of the state-specific gene sets. In regard to colon-specific, intestinal stem cell (ISC)-specific, and embryonic stem cell (ESC)-specific gene sets, we independently performed PCA and used PC1 as a representative value.

### Network inference and identification of core TFs

To infer context-specific GRNs, we calculated mutual information of each dataset using ARACNE algorithm implemented in the “RTN” package (11, 27). ARACNE was run with 0.01 data processing inequality tolerance, 100 bootstraps. We used a list of curated TFs from a previous study (28) that are included in the list of TFs defined by Functional Annotation of the Mammalian Genome (FANTOM; ref. 29). We inferred GRNs from the GEO-integrated normal dataset and GEO-paired normal dataset, and called them GEO-integrated normal GRN and GEO-paired normal GRN, respectively. To identify the master regulators of normal colon epithelial cells, we used Master Regulator Inference Algorithm (MARINA; ref. 11). For GEO-integrated normal GRN and GEO-paired normal GRN, we identified the TFs whose inferred target genes were significantly enriched in colon tissue-specific genes and normal-specific genes, respectively. We performed MARINA analysis with *P* value cutoff of 1e-10. The common TFs that significantly include the colon-specific genes and normal-specific genes as their target genes were defined as the core TFs of normal colon epithelial cells. The transcriptional activity of the TFs was quantitatively measured using Virtual Inference of Protein activity by Enriched Regulon analysis (VIPER) algorithm (30). We used a normalized enrichment scores calculated by single-sample VIPER methods as a transcriptional activity.

### Identification of modulators using the inferred GRN

To identify the potential modulators, we established a set of criteria: a modulator (i) having negative regulatory interaction with the colon-specific genes, (ii) having protein–protein interaction with the core TFs, and (iii) inducing alteration of regulatory interaction between the core TFs and their target genes depending on the gene expression of a modulator. To measure the negative influence of potential modulators on the gene expression of colon-specific genes, we used the VIPER algorithm (30). We reconstructed the regulatory network composed of proteins (GO: 0006464, cellular protein modification process, *n* = 2,477) using the ARACNE algorithm in GEO-integrated cancer dataset. To define the putative modulators, we used the multiple sample version of VIPER with the colon-specific gene sets. We selected

the proteins that have significant regulatory interaction with colon-specific genes with a *P* value below 0.01. To assess the protein–protein interaction, we used the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database. The interaction scores of each potential modulator were calculated by summing the value of “combined interaction scores” of each modulator with the five core TFs. To examine the changes of regulatory interaction between each potential modulator and the core TFs, we assessed conditional mutual information using Modulator Inference by Network Dynamics (MINDy) algorithm (31) that measures the differences in mutual information by comparing the mutual information calculated from the samples with the highest expression level of a certain gene and that calculated from the samples with the lowest expression of the gene. We used 300 samples among GEO-integrated cancer data with the highest gene expression of potential modulators, and 300 samples with the lowest gene expression of potential modulators. GEO-integrated normal GRN that has the largest number of samples among the datasets was used. We performed the MINDy analysis with *P* value cutoff threshold of 1e-05.

#### RNA isolation, qRT-PCR, and RNA-seq

RNA was extracted using an RNA-spin kit (INTRON), and cDNA was synthesized using a DiaStar RT kit (Solgent) and 2X Taq premix (Solgent) according to the manufacturer's instructions. After reverse transcription, qPCR was performed using the QuantStudio5 (Applied Biosystems) with SYBR Master Mix (GeNet Bio). All conditions were examined in three independent replicates and normalized to  $\beta$ -actin mRNA expression. qPCR primer sequences we used are given in Supplementary Table S5.

#### Microarray analysis

For Caco2, HCT116, SW480, and SW620, microarray analysis was performed on a single sample. Total RNA was extracted using an RNA-spin kit (INTRON), and RNA purity and integrity were evaluated by ND-1000 Spectrophotometer (NanoDrop) and Agilent 2100 Bioanalyzer (Agilent Technologies). The fragmented cRNA was hybridized to the Affymetrix GeneChip Human Genome U133 Plus 2.0 Gene Expression Array at 45°C for 16 hours. Hybridized arrays were washed and stained on a GeneChip Fluidics Station 450 and scanned on a GCS3000 Scanner (Affymetrix). Signal values were computed using the Affymetrix GeneChip Command Console software. The data were normalized with robust multi-average method (32).

#### Bulk RNA-seq analysis

Total RNA was extracted using an RNA-spin kit (INTRON), and RNA quality was assessed by Agilent 2100 bioanalyzer using the RNA 6000 Nano Chip (Agilent Technologies). RNA quantification was performed using ND-2000 Spectrophotometer (Thermo, Inc.), and 500 ng of RNA per sample was used for library construction. The library construction was performed using the QuantSeq 3' mRNA-Seq Library Prep Kit (Lexogen, Inc.), and the library was single-end sequenced on an Illumina NextSeq 500 (Illumina, Inc.). The reads were aligned using Bowtie2 (33), and the read counts estimated by Bedtools (34) were further processed on the basis of Quantile–Quantile normalization method using “edgeR” package (35) in R.

#### Single-cell RNA-seq analysis

For Caco2 cells stably expressed shRNA targeting *SETDB1* and shScramble, we used 10X Genomics Chromium machine to capture single cells and cDNA preparation. Three thousand cells were loaded on a Chromium Single Cell Instrument to generate single-cell nanoliter-scale Gel Bead-In-Emulsions. The barcoded cDNA was then

amplified in PCR, and the single-cell RNA-seq libraries were prepared using version 1 Chromium Single Cell 3' Library, Gel beads, and Multiplex kit (10X Genomics). The libraries were quantified using qPCR according to the qPCR Quantification Protocol Guide (KAPA Library Quantification kits for Illumina Sequencing platforms) and qualified using the TapeStation D1000 ScreenTape (Agilent Technologies). Indexed libraries were then sequenced using the HiSeq2500 platform (Illumina) by the Macrogen Incorporated.

#### Preprocessing and quality control of single-cell RNA-seq data

We aligned raw FASTQ files to hg19 human genome assembly and aggregated two samples using the 10X Genomics Cell Ranger pipeline (<https://support.10xgenomics.com/single-cell-gene-expression/software/downloads/latest>) with standard settings. Cells that had either the number of total counts under 20,000 or the number of expressed genes under 3,000 are precluded. After quality control, we obtained 453 cells of shSETDB1 samples and 654 cells of shScramble samples. To mitigate cell-specific biases, library size normalization was performed by “scater” package (36), and consequent log-transformed normalized expression values were used following analysis.

#### Data availability

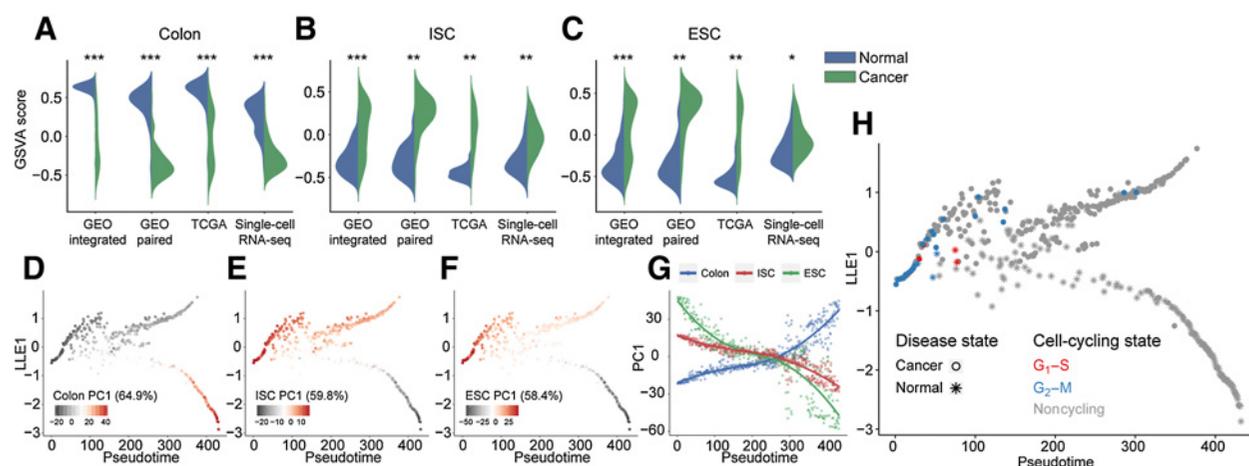
The data accompanying this article were deposited into the GEO database under accession number GSE135461.

For additional information regarding used transcriptomic data, signature gene sets, and sequence information of primers and siRNAs, see Supplementary Materials and Methods.

## Results

### Colorectal cancer comprises embryonic stem-like cells, intestinal stem-like cells, and differentiated normal-like cells

To determine the differences in gene expression programs between normal colonic epithelial cells and colorectal cancer cells, we collected four different public transcriptome datasets (Supplementary Table S1) and compared normal and cancer samples using gene sets representing distinct states of cellular development (Supplementary Table S2). We scored individual samples for their match with each state-specific expression profile derived from three large consortiums on tissue-specific gene expression: Human Protein Atlas, Genotype-Tissue Expression Project, and FANTOM5 project (Supplementary Fig. S1A). The most apparent feature of the cancer samples across all datasets was a reduction in the expression of the colon-specific program (Fig. 1A). We also found a significant correlation between the gene set enrichment score for colon-specific genes and tumor size and prognosis of patients with colorectal cancer (Supplementary Fig. S1B and S1C), suggesting that loss of this profile is clinically relevant and may produce histopathologic features of poorly differentiated cancer (37). To further examine the cellular reprogramming in colorectal cancer, we derived stem/progenitor-associated gene sets from previous studies, ISCs, primitive guts, neural crest, embryonic endoderm, and ESCs (Supplementary Fig. S2A; Supplementary Table S2; Supplementary Materials and Methods). We found that, compared with the other stem/progenitor gene sets, gene sets associated with ISC and ESC are significantly enriched in colorectal cancer (Fig. 1B and C; Supplementary Fig. S2B and S2C), suggesting that these two stem/progenitor expression programs may predominantly drive colorectal cancer-associated dedifferentiation. To examine the correlation among the gene expression programs, we assessed Spearman correlation with the gene set enrichment scores of colorectal cancer epithelial cells in



**Figure 1.**

A developmental hierarchy in colorectal cancer. **A–C**, Violin plots showing gene set enrichment score of each state-specific gene set in normal tissues and cancer tissues across the datasets. Colon-specific gene set presented here is from Human Protein Atlas, ISC-specific gene set is from Li and colleagues (38), and ESC-specific gene set is from Wong and colleagues (3). *P* values are by two-sample proportion test (\*,  $P < 1e-10$ ; \*\*,  $P < 1e-30$ ; and \*\*\*,  $P < 1e-50$ ). **D–F**, Inferred pseudotime (*x*-axis) and first dimension (*y*-axis) of linear local embedding (LLE), which is a nonlinear dimensionality reduction implemented by the SLICER algorithm. Points are colored by PC1 of each state-specific gene set. The proportion of variance explained by PC1 is indicated. **G**, Smoothed curve of PC1 of each state-specific gene set along pseudotime. **H**, Scatter plot showing cycling cells among total cell population. Point shapes are based on the disease states annotated by Li and colleagues (38).

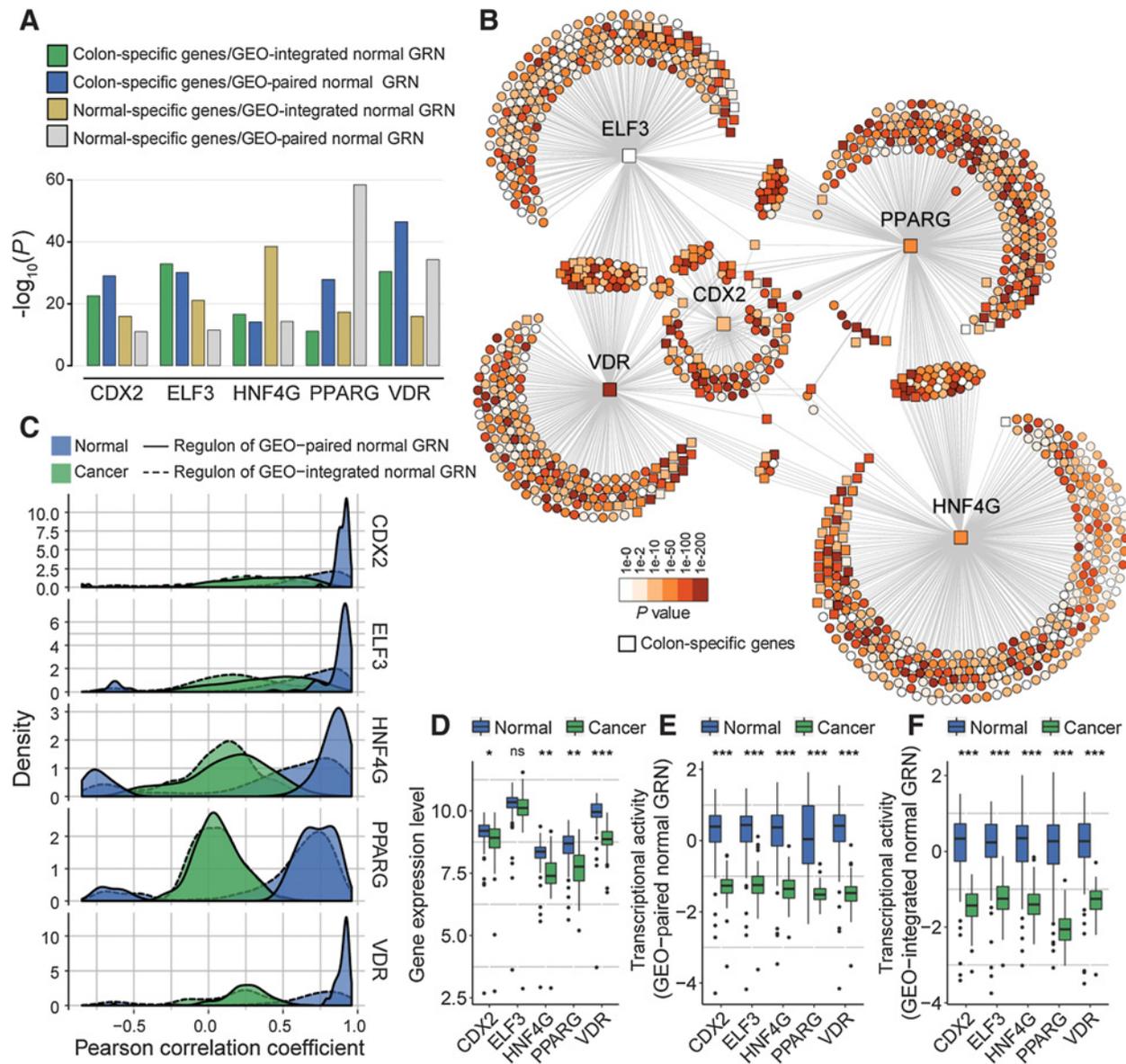
single-cell RNA-seq data. We found that ESC-specific genes are not significantly correlated with ISC-specific genes (Spearman correlation coefficient = 0.112; Supplementary Fig. S3A), and that only a small number of cancer cells exhibit high expression level of ESC-specific genes (Supplementary Fig. S3B).

Because dedifferentiation is considered a continuous process, we performed pseudotime analysis on RNA-seq data of both normal and colon cancer cells using the previously annotated differentiation state (38). As expected, stem-like and differentiated-like cellular identities were well divided along with the pseudotime axis (Supplementary Fig. S3C). We also performed PCA and used principal component 1 (PC1) as a representative value of state-specific gene expression. The colon-specific gene profile increased over the pseudotime trajectory, whereas both stem cell profiles decreased (Fig. 1D–F; Supplementary Fig. S2B and S2C). However, only a few cancer cells at the beginning of trajectory exhibited markedly elevated ESC-specific gene expression (Fig. 1G; Supplementary Figs. S3D and S4E). Moreover, transcriptional entropy, the diversity of transcriptome from the concept of Shannon's entropy which has been used in the study of cellular differentiation (25), also particularly increased at the beginning of trajectory (Supplementary Fig. S3F), suggesting that such distinct cell population has a high developmental potential. We found that expressions of *MYC* and the ISC markers (39), *TERT*, *LRIG1*, and *BM11*, are increased in cells early in the pseudotime trajectory, whereas expressions of *LGR5* and *CD44*, well-known markers of ISCs and cancer stem cells (40, 41), are not particularly increased in cells early in the pseudotime trajectory (Supplementary Fig. S3G). Also enriched at the forefront of pseudotime trajectory were cells undergoing cell division (Fig. 1H), which implies that these cells might be proliferative cancer stem cells within the sample population. We also found a few cancer cells near the end of the trajectory where most normal cells were found. These colorectal cancer cells exhibited a high level of colon-specific gene expression (Fig. 1D; Supplementary Fig. S3D and S3E) and were noncycling (Fig. 1H), indicating that these are postmitotic differentiated cancer cells with a normal-like gene expression profile. Collectively, these results indicated

that colorectal cells undergoing oncogenic transformation span a developmental continuum. Furthermore, the data suggested that colorectal cancer cells could be differentiated into a normal-like postmitotic state through such a trajectory.

#### The function of core TFs of the normal colon-specific gene expression program is disrupted in colorectal cancer

Overexpression of a few core TFs governing the expression of a tissue-specific gene profile can override the original developmental state and convert the cells into another desired state (42). Therefore, to find core TFs that can convert a cancer stem-like state to a normal differentiated state, we reconstructed GRNs of normal colon epithelial cells (Supplementary Fig. S4A) and defined colon-specific genes (Supplementary Fig. S4B and S4C) and normal-specific genes that are highly expressed in normal colon mucosa compared with colorectal cancer (Supplementary Fig. S4D) as the signature gene sets of normal colon epithelial cells. From these, we established a set of criteria for master regulators that show high statistical significance with which a TF regulates the defined signature genes as its target genes (Materials and Methods). We identified five TFs (*CDX2*, *ELF3*, *HNF4G*, *PPARG*, and *VDR*) that dominantly regulate the signature genes in both inferred GRNs (Fig. 2A and B; Supplementary Fig. S4E–S4H). Interestingly, we found that the correlation between the core TFs and their inferred target genes is markedly reduced in colorectal cancer cells (Fig. 2C), yet the expression of the genes encoding these core TFs was much less reduced (Fig. 2D; Supplementary Fig. S4I–S4K and Supplementary Table S3). In addition, transcriptional activity of the core TFs estimated by the VIPER algorithm (ref. 30; Fig. 2E and F) and the average expression of their target genes were also significantly decreased (Supplementary Fig. S4I–S4K and Supplementary Table S3), suggesting that not only expression of the genes encoding the core TFs but also transcriptional activity of the TFs might be dysregulated. Thus, we hypothesized that posttranslational modifications contribute to the dysfunction of the core TFs in a direct or indirect manner.


**Figure 2.**

Identification of core TFs of normal colon epithelial cells. **A**, Core TFs of normal colon epithelial cells with  $P$  value below  $1e-10$  determined by MARINA analysis. GRNs inferred from GEO-integrated normal dataset (GEO-integrated normal GRN) and from GEO-paired normal dataset (GEO-paired normal GRN) were used for identifying the core TFs regulating colon-specific genes and normal-specific genes, respectively. **B**, Visualization of the core TFs and their target genes based on GEO-integrated normal GRN. The nodes are colored by statistical significance of differences between normal and cancer examined by unpaired two-tailed Student  $t$  test. The square nodes represent colon-specific genes. **C**, Histograms showing Pearson correlation coefficient between each core TF and its regulon (a group of inferred target genes of TF) in GEO-paired dataset. Pearson correlation coefficient between the core TFs and the regulons from GEO-paired normal GRN (solid line), and the regulons from GEO-integrated normal GRN (dashed line) was calculated, respectively. The color of histogram indicates the disease states. **D**, Comparison of gene expression level of core TFs in GEO-paired dataset. **E** and **F**, Transcriptional activity of the core TFs inferred by the VIPER algorithm in GEO-paired normal GRN (**E**) and GEO-integrated normal GRN (**F**; \*,  $P < 5e-2$ ; \*\*,  $P < 1e-10$ ; and \*\*\*,  $P < 1e-20$ ).

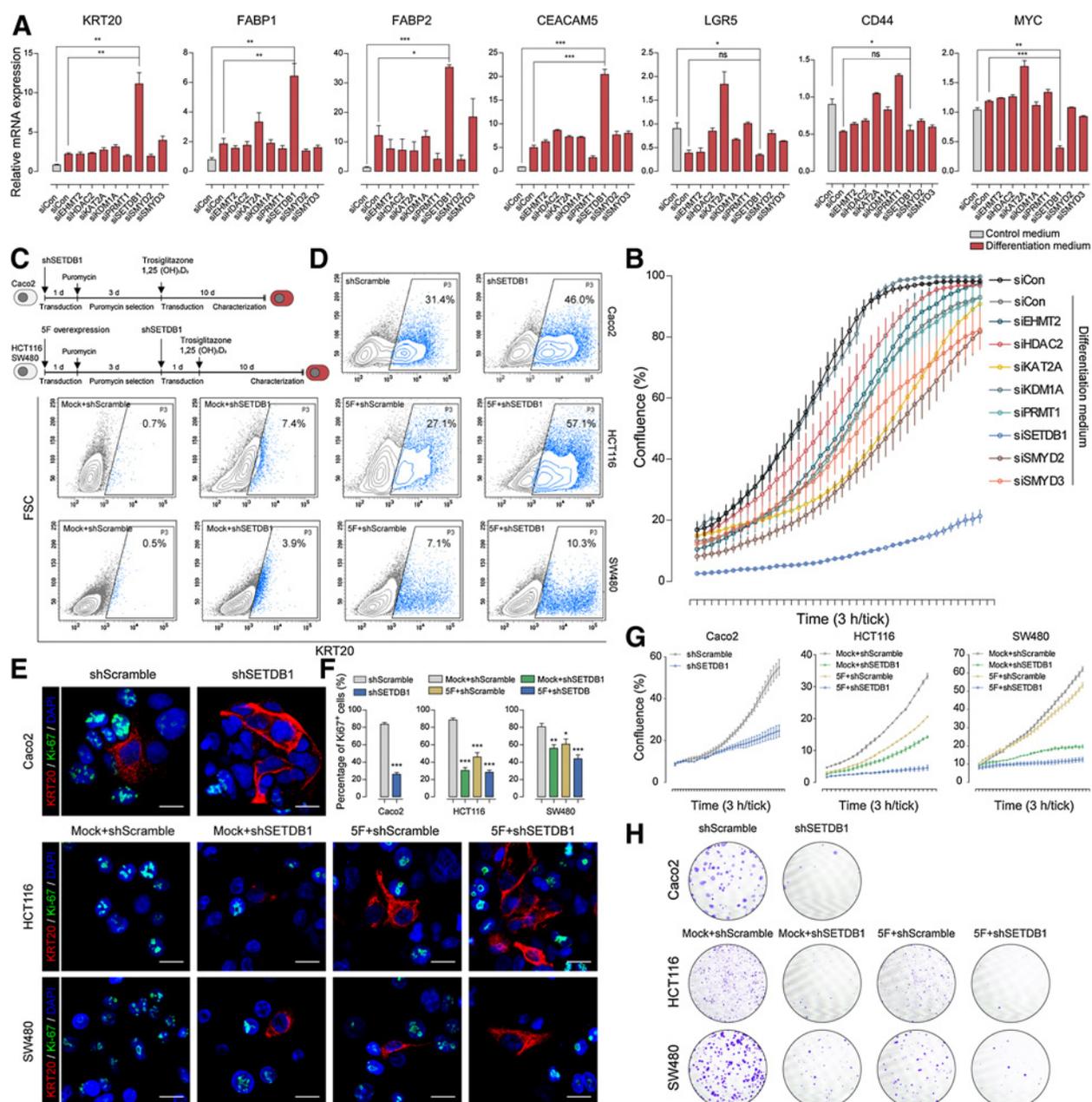
### SETDB1 causes functional disruption of core TFs and alteration of the gene expression of the differentiated state

To identify the hidden factor(s) that hinders the transcriptional activity of the core TFs, we screened the proteins (GO: 0006464, cellular protein modification process,  $n = 2,477$ ) that had negative regulatory interactions with colon-specific genes. Among these proteins (Supplementary Table S4,  $P < 0.01$ ), we focused on chromatin-modifying enzymes that have significantly enriched GO terms (Sup-

plementary Fig. S5A). We chose chromatin-modifying enzymes, because a restrictive chromatin structure would impair access of TFs to their target genes (7). We selected eight chromatin-modifying enzymes that have high interaction scores with core TFs and show significant differences in regulatory interactions between core TFs and target genes depending on their expression (Supplementary Fig. S5B). To determine the role of the selected enzymes on the activity of the core TFs, we performed siRNA experiments in Caco2, which are stem

cell-like colorectal cancer cells and exhibited a relatively higher expression of core TFs than other stem cell-like colorectal cancer cell lines (Supplementary Fig. S5C). We confirmed knockdown after transfection of the siRNAs and 48-hour incubation in differentiation medium (Supplementary Fig. S5D). Of the eight chromatin-modifying enzymes tested, knockdown of *SETDB1*, a histone H3

lysine 9-specific methyltransferase, increased the expression of *KRT20*, *FABP1*, *FABP2*, and *CEACAM5*, which are markers of differentiated colonic epithelial cells, and reduced expression of *MYC* (Fig. 3A). Consistent with the reduction in *MYC* expression, the cell proliferation rate was markedly reduced upon *SETDB1* knockdown (Fig. 3B).



**Figure 3.**

Identification of a chromatin regulator that hinders transcriptional activity of the core TFs. **A**, Transcript analysis by qRT-PCR of gene expression in Caco2 with each siRNA transfection. Graph bars are cells grown in control medium, and red are cells grown in differentiation medium ( $n = 3$ , mean  $\pm$  SEM). **B**, Percent growth over time in Caco2 after transfection with the indicated siRNAs. **C**, Schematic for the experiments with retrovirus-mediated transduction. **D**, Flow cytometry data showing KRT20 abundance (x-axis) and forward-scattered light (FSC; y-axis;  $n = 3$ ). **E**, IF of KRT20 (red) and Ki-67 (green) with DAPI nuclear counterstain (blue). Scale bar, 50  $\mu$ m. **F**, Quantification of Ki-67-positive cells in low-magnification confocal IF images across 10 fields to 20 fields for each group (mean  $\pm$  SEM). **G**, Percent growth of three cell lines after *SETDB1* knockdown ( $n = 3$ , mean  $\pm$  SEM). **H**, Colony-forming assay over 14 days. Data are representative of three independent experiments (\*,  $P < 0.05$ ; \*\*,  $P < 0.01$ ; and \*\*\*,  $P < 0.001$ ).

Single-cell RNA-seq analysis showed higher expression of *SETDB1* in stem cell-like cancer cells than in normal stem cells (Supplementary Fig. S5E and S5F), and significant correlation of each PC1 value of colon-, ISC-, and ESC-specific genes with *SETDB1* expression (Supplementary Fig. S5G–S5I). A previous IHC study also reported that *SETDB1* is only expressed in colorectal cancer but not in adjacent normal colon tissues (43). The expression of *SETDB1* was also correlated with poor prognosis, especially for disease-free survival (Supplementary Fig. S5J). Furthermore, the prognosis was particularly poor for those patients with low gene set enrichment score of colon-specific genes and high *SETDB1* expression (Supplementary Fig. S5K).

### **SETDB1 depletion induces terminal differentiation of stem-like colorectal cancer cells into postmitotic cells**

Because cellular reprogramming requires several days, we performed prolonged depletion of *SETDB1* by retroviral-mediated shRNA interference (Fig. 3C; Supplementary Fig. S6A). We also tested the effect of *SETDB1* depletion in the stem cell-like colorectal cancer cell lines, HCT116 and SW480, in which we overexpressed all five core TFs (Supplementary Fig. S6C), because these cells had lower expression of the core TFs than do Caco2 cells (Supplementary Figs. S5C and S6B). Using flow cytometry analysis, we assessed the effect of *SETDB1* depletion on the cell differentiation status using the abundance of KRT20 as a marker of differentiated colon cells (10). *SETDB1* depletion increased the KRT20<sup>+</sup> cell population in Caco2, as well as in TF-overexpressing HCT116 and SW480 cells (Fig. 3D). Of note, concomitant *SETDB1* depletion and overexpression of the core TFs synergistically induced KRT20<sup>+</sup> cell population to 57.1% of HCT116 cells which originally had very few KRT20<sup>+</sup> cells (Fig. 3D). We sorted the cells based on the abundance of KRT20 and analyzed the transcript abundance from genes indicative of normal colon epithelial cells. These transcriptional markers were highly increased in KRT20<sup>+</sup> cells and also moderately increased in KRT20<sup>-</sup> cells (Supplementary Fig. S6D), suggesting that *SETDB1* depletion profoundly restores normal-like gene expression profiles even in KRT20<sup>-</sup> colorectal cancer cells. To characterize phenotypic properties of differentiated normal-like cells, we performed IF analysis using antibodies against KRT20 and Ki-67 (a marker of cell proliferation). We observed that most KRT20<sup>+</sup> cells in all three cell lines appeared Ki-67 negative (Fig. 3E and F; Supplementary Fig. S6E), indicating that the differentiated cells are not actively proliferating. At the population level, the proliferation rate was remarkably decreased (Fig. 3G), and most cells remained in postmitotic states in long-term culture (Fig. 3H). Taken together, the results showed that *SETDB1* knockdown in colorectal cancer cell lines expressing the five core TFs effectively induces a differentiated postmitotic state from cells with a stem cell-like phenotype.

### **SETDB1 depletion recapitulates the function of core TFs and restores the normal colon-specific gene expression profile in colorectal cancer cells**

To determine whether *SETDB1* depletion produced a global gene expression profile like that of normal differentiated colon cells, we performed single-cell RNA-seq and bulk RNA-seq analysis. For single-cell RNA-seq analysis, we used Caco2 cells transduced with sh*SETDB1* and shScramble, and we obtained 453 cells and 654 cells, respectively (Fig. 4A). We found that a single differentiation trajectory resulted from *SETDB1* depletion (Fig. 4B; Supplementary Fig. S7A), and that the colon-specific gene expression increased cellular differentiation progressed, whereas both ISC- and ESC-specific gene expressions decreased (Fig. 4C and D; Supplementary Fig. S7B–S7D). The bulk RNA-seq data also showed that *SETDB1* depletion reinstates the global

gene expression of normal colon mucosa across the three cell lines (Fig. 4E and F; Supplementary Fig. S8A and S8B). In addition, we found that intestinal differentiation signature genes are profoundly upregulated in *SETDB1*-downregulated cells, whereas ISC signature genes and Wnt signaling signature genes are downregulated (Supplementary Fig. S8C). We also found that the expression level of E-cadherin as well as ZO-1 and ZO2 is upregulated upon *SETDB1* depletion, suggesting that reestablishment of normal colon gene expression programs includes restoration of an epithelial cell state (Supplementary Fig. S8D). The expressions of target genes of the five core TFs were also increased (Supplementary Fig. S7E–S7H). *SETDB1* mediates the trimethylation of histone H3 lysine 9 (H3K9me3) and promotes transcriptional suppression of its target genes. To examine the H3K9me3 level at the promoter loci of the key differentiation genes, we performed ChIP-qPCR analysis. We found that the H3K9me3 levels at the promoter loci of KRT20, FABP1, and E-cadherin are decreased upon *SETDB1* depletion (Fig. 4G). It implies that the core TFs might restore their transcription function at their target genes due to the chromatin rearrangement mediated by *SETDB1* downregulation. Also noteworthy, numerous TFs previously known to regulate colon differentiation such as KLF4 (44), SATB2 (45), and PRDM1 (46) were reactivated upon *SETDB1* depletion, whereas the TFs that regulate proliferation and stemness such as TCF3 (47) and FOXM1 (48) were significantly inactivated (Fig. 4H).

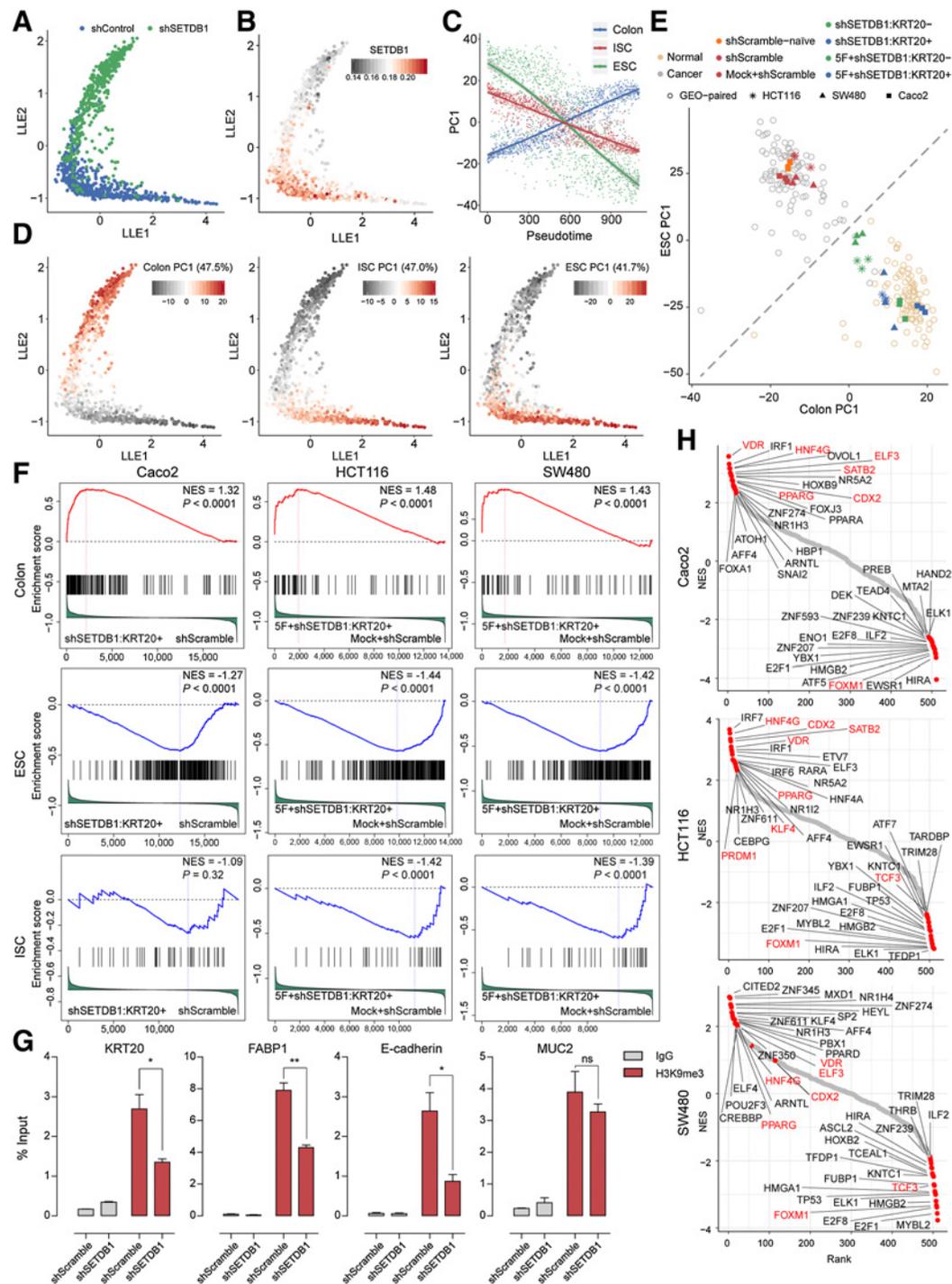
### **SETDB1 is highly expressed in poorly differentiated cells in human colorectal cancer, and its depletion induces cellular differentiation of patient-derived colon cancer organoids**

To examine potential clinical implications, we performed IHC for *SETDB1* on tissue microarrays of patients with human colorectal cancer. We discovered that *SETDB1* is increased in cancer cells compared with that in normal cells (Fig. 5A). Moreover, we found that *SETDB1* is highly expressed in poorly differentiated colorectal cancer cells that lose conventional gland-forming histology (Fig. 5A). Indeed, the prognosis of patients who have cancer cells with high *SETDB1* expression showed significantly worse than those who have cancer cells with relatively low *SETDB1* expression (Fig. 5B).

To evaluate the function of *SETDB1* in human colon cancer organoids, we established patient-derived colon cancer organoid lines and assessed the effect of *SETDB1* depletion. *SETDB1*-depleted cancer organoids exhibited differentiated morphology contrary to spheroid-like morphology of typical colon cancer organoids (Fig. 5C and D). We also found that these *SETDB1*-depleted organoids with differentiated morphology have an increased number of KRT20<sup>+</sup> cells compared with control organoids (Fig. 5E). Together, these data suggest that the ablation of *SETDB1* can override the oncogenic dedifferentiation process in colorectal cancer and successfully redifferentiate cancer cells into normal-like cells.

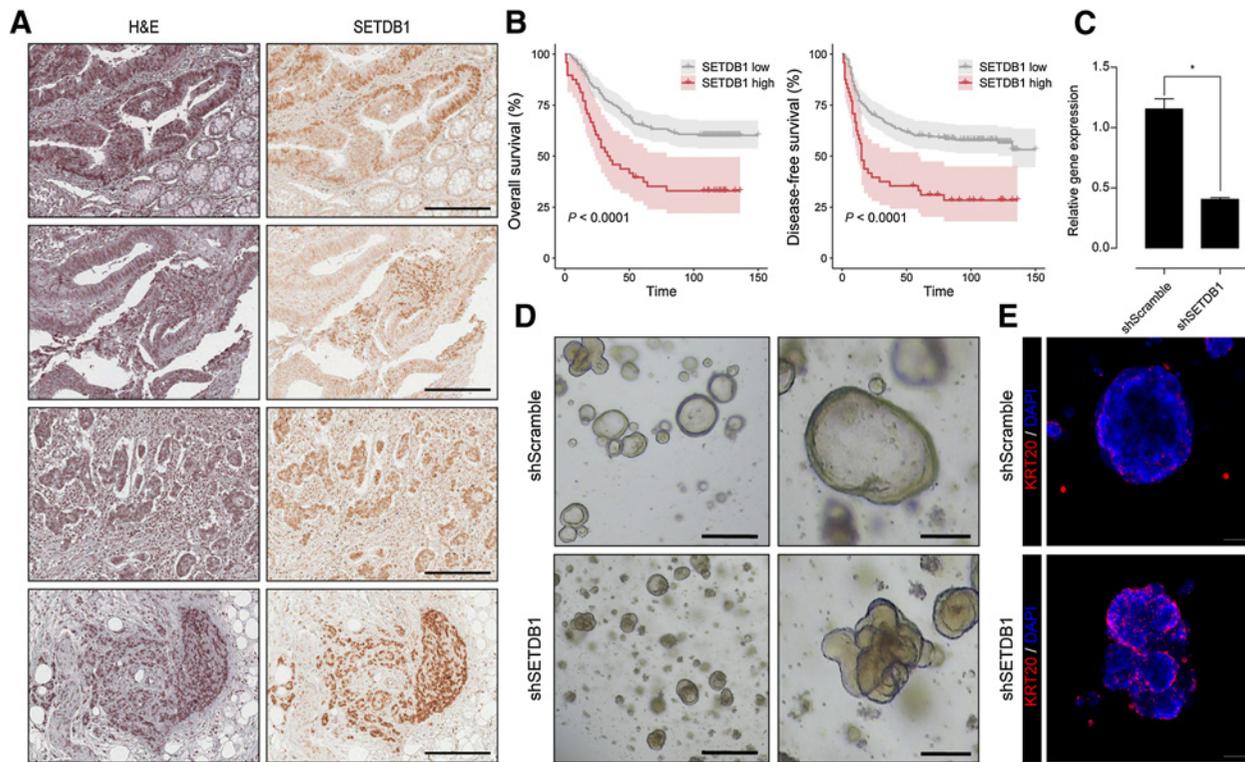
## **Discussion**

The observation of sudden loss of tumorigenicity in malignant cancer has been sporadically reported (49–51). Such observation often corresponded to cellular differentiation or transdifferentiation in which cancer cells spontaneously convert into postmitotic normal-like cells. Although the loss of tumorigenicity upon terminal differentiation has a great potential as a new anticancer therapy, the application of differentiation therapy has been restricted to only acute promyelocytic leukemia (52). This is primarily because the solid cancer has more complex mechanism of differentiation than leukemia (53). With the progress of computational frameworks to direct



**Figure 4.**

Global gene expression of colorectal cancer cells converted to differentiated normal-like cells. **A**, Scatter plot showing the first dimension of LLE constructed by SLICER for Caco2 cells. Points are colored by sample types. **B**, Imputed gene expression level of *SETDB1*. **C**, Scatter plot with smoothed profiles of PC1 of each state-specific gene set. **D**, Scatter plots colored by state-specific gene set expression. **E**, Bulk RNA-seq data with PC1 of colon-specific genes (x-axis) and PC1 of ESC-specific genes (y-axis) are shown. GEO-paired dataset and *in vitro* samples are shown with different shapes. Except for shScramble-naïve sample, all *in vitro* samples were cultured in differentiation condition. **F**, GSEA on bulk RNA-seq data with colon-specific genes defined by Human Protein Atlas (top), ESC-specific genes defined by Wong and colleagues (ref. 3; middle), and ISC defined by Li and colleagues (ref. 38; bottom). **G**, ChIP-qPCR analysis of H3K9me3 at the promoter loci of KRT20, FABP1, E-cadherin, and MUC2 in Caco2 cells. The relative amounts of immunoprecipitated DNA are depicted as a percentage of input DNA ( $n = 3$ , mean  $\pm$  SEM). **H**, Normalized enrichment scores (NES) of all TFs calculated using the VIPER algorithm are presented in a rank-order. NES was calculated on the basis of GEO-paired normal GRN and DEGs between unsorted cells and sorted KRT20<sup>+</sup> cells in Caco2 (top), HCT116 (middle), and SW480 (bottom) cell lines, respectively.



**Figure 5.**

Expression of SETDB1 in colorectal cancer tissues and cellular differentiation of patient-derived colon cancer organoids upon *SETDB1* depletion. **A**, Hematoxylin and eosin (H&E) staining (left) and IHC staining of SETDB1 (right). Scale bar, 200  $\mu$ m. **B**, Kaplan-Meier analysis of overall survival ( $n = 248$ , the number of samples of SETDB1 high = 48, the number of samples of SETDB1 low = 200,  $P = 3.66e-05$ ) and disease-free-survival ( $n = 249$ , the number of samples of SETDB1 high = 48, the number of samples of SETDB1 low = 201,  $P = 2.37e-05$ ) with SETDB1 expression. Samples were dichotomized into high and low according to the presence of cancer cells which exhibit very high expression of SETDB1. **C**, *SETDB1* mRNA level was measured by qRT-PCR in human colon cancer organoids. Lentiviral infection of shScramble and shSETDB1 was proceeded with puromycin (2  $\mu$ g/mL) selection, and stable knockdown organoids were maintained for 2 weeks ( $P = 0.0102$ ). **D**, Phase contrast images of stable knockdown organoids. Two-hundred cells per well in 48-well plates were cultured for 8 days. Scale bar, 200  $\mu$ m (left) and 50  $\mu$ m (right). **E**, IF of KRT20 with DAPI nuclear counterstain. Scale bar, 50  $\mu$ m.

reprogramming (11–13), we explored a systems approach to identify causal factors that can induce normalizing transitions of cancer cells. In our study, we discovered the factors using a network-based inference approach and demonstrated that SETDB1 constitutes an epigenetic barrier against colorectal cancer differentiation by impeding the function of lineage-specifying TFs of normal colon epithelial cells. Both bulk RNA-seq and single-cell RNA-seq analyses revealed that, upon SETDB1 depletion, colon-specific genes regulated by the master regulators CDX2, ELF3, HNF4G, PPARG, and VDR are reactivated, and stem cell-associated genes are suppressed. Such restoration of a normal gene expression profile effectively reestablished the postmitotic state in malignant colorectal cancer cells.

Previous studies reported that SETDB1 contributes to suppression of genes encoding lineage-specifying regulators in ESCs (54) and melanoma (55), supporting our discovery. These studies demonstrated that SETDB1 mediates H3K9 methylation on differentiation-associated genes such as HOX genes, preventing ESCs and melanoma cells from differentiation. We also found that the transcriptional activity of some HOX genes is reactivated upon SETDB1 depletion, but the colon-specific TFs, including the five core TFs, KLF4, and SATB2 are more evidently reactivated (Fig. 4F). Thus, the detailed mechanism of SETDB1 in colorectal cancer as well as in other cancer types should be examined in a further study.

It is noteworthy that SETDB1 would be an effective therapeutic target for eradicating cancer stem cells because it is upregulated in colorectal cancer stem-like cells but not in normal stem cells (Supplementary Fig. S6E). In colorectal cancer, the absence of distinguishing characteristics of cancer stem cells compared with normal stem cells makes it difficult to selectively ablate cancer stem cells without damages on normal stem cells. In addition, cancer stem cells are known to have strong drug resistance to anticancer drugs; thereby, current chemotherapy drugs have a limited effect in treating them. Our experiment shows that SETDB1 depletion combined with cytotoxic drugs might also be potentially beneficial to anticancer treatment (Supplementary Fig. S8E and S8F). In this regard, inducing differentiation through *SETDB1* depletion may provide a promising therapeutic strategy for colorectal cancer. Taken together, our results highlight not only the possibility of reverting colorectal cancer cells into normal-like postmitotic cells by restoring the tissue-specific gene expression program, which represents a therapeutic intervention that might overcome current limitations of cancer therapy, which mostly focuses on inducing cancer cell apoptosis.

#### Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

## Authors' Contributions

**Conception and design:** K.-H. Cho

**Development of methodology:** S. Lee

**Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.):** S. Lee, S.N. Hong, S.-H. Kim

**Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis):** S. Lee, D. Kim, K.-H. Cho

**Writing, review, and/or revision of the manuscript:** S. Lee, C. Lee, C.Y. Hwang, S.-H. Kim, K.-H. Cho

**Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases):** C. Lee, Y. Han, S.-H. Kim, K.-H. Cho

**Study supervision:** K.-H. Cho

## References

- Bradner JE, Hnisz D, Young RA. Transcriptional addiction in cancer. *Cell* 2017; 168:629–43.
- Ben-Porath I, Thomson MW, Carey VJ, Ge R, Bell GW, Regev A, et al. An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. *Nat Genet* 2008;40:499–507.
- Wong DJ, Liu H, Ridky TW, Cassarino D, Segal E, Chang HY. Module map of stem cell genes guides creation of epithelial cancer stem cells. *Cell Stem Cell* 2008; 2:333–44.
- Suva ML, Rheinbay E, Gillespie SM, Patel AP, Wakimoto H, Rabkin SD, et al. Reconstructing and reprogramming the tumor-propagating potential of glioblastoma stem-like cells. *Cell* 2014;157:580–94.
- Kaufman CK, Mosimann C, Fan ZP, Yang S, Thomas AJ, Ablain J, et al. A zebrafish melanoma model reveals emergence of neural crest identity during melanoma initiation. *Science* 2016;351:aad2197.
- Suva ML, Riggi N, Bernstein BE. Epigenetic reprogramming in cancer. *Science* 2013;339:1567–70.
- Flavahan WA, Gaskell E, Bernstein BE. Epigenetic plasticity and the hallmarks of cancer. *Science* 2017;357:eaal2380.
- Califano A, Alvarez MJ. The recurrent architecture of tumour initiation, progression and drug sensitivity. *Nat Rev Cancer* 2017;17:116–30.
- Schwitala S, Fingerle AA, Cammareri P, Nebelsiek T, Goktuna SI, Ziegler PK, et al. Intestinal tumorigenesis initiated by dedifferentiation and acquisition of stem-cell-like properties. *Cell* 2013;152:25–38.
- Shimokawa M, Ohta Y, Nishikori S, Matano M, Takano A, Fujii M, et al. Visualization and targeting of LGR5(+) human colon cancer stem cells. *Nature* 2017;545:187–92.
- Basso K, Margolin AA, Stolovitzky G, Klein U, Dalla-Favera R, Califano A. Reverse engineering of regulatory networks in human B cells. *Nat Genet* 2005;37: 382–90.
- Cahan P, Li H, Morris SA, Lummertz da Rocha E, Daley GQ, Collins JJ. CellNet: network biology applied to stem cell engineering. *Cell* 2014;158:903–15.
- Rackham OJ, Firas J, Fang H, Oates ME, Holmes ML, Knaupp AS, et al. A predictive computational framework for direct reprogramming between human cell types. *Nat Genet* 2016;48:331–5.
- Hrvatn S, Deng F, O'Donnell CW, Gifford DK, Melton DA. MARIS: method for analyzing RNA following intracellular sorting. *PLoS One* 2014;9:e89459.
- Yeo SY, Lee KW, Shin D, An S, Cho KH, Kim SH. A positive feedback loop bi-stably activates fibroblasts. *Nat Commun* 2018;9:3016.
- Hanzelmann S, Castelo R, Guinney J. GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* 2013;14:7.
- Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015;43:e47.
- Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* 2001;98:5116–21.
- Schwender H, Ickstadt K. Empirical Bayes analysis of single nucleotide polymorphisms. *BMC Bioinformatics* 2008;9:144.
- Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 2013;6:pl1.
- Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 2004;5:R80.
- Welch JD, Hartemink AJ, Prins JF. SLICER: inferring branched, nonlinear cellular trajectories from single cell RNA-seq data. *Genome Biol* 2016;17:106.
- Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, et al. Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods* 2017;14: 979–82.
- Tirosh I, Venteicher AS, Hebert C, Escalante LE, Patel AP, Yizhak K, et al. Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. *Nature* 2016;539:309–13.
- Grun D, Muraro MJ, Boisset JC, Wiebrands K, Lyubimova A, Dharmadhikari G, et al. De novo prediction of stem cell identity using single-cell transcriptome data. *Cell Stem Cell* 2016;19:266–77.
- van Dijk D, Sharma R, Nainys J, Yin K, Kathail P, Carr AJ, et al. Recovering gene interactions from single-cell data using data diffusion. *Cell* 2018;174:716–29.
- Fletcher MN, Castro MA, Wang X, de Santiago I, O'Reilly M, Chin SF, et al. Master regulators of FGFR2 signalling and breast cancer risk. *Nat Commun* 2013;4:2464.
- Carro MS, Lim WK, Alvarez MJ, Bollo RJ, Zhao X, Snyder EY, et al. The transcriptional network for mesenchymal transformation of brain tumours. *Nature* 2010;463:318–25.
- FANTOM Consortium and the RIKEN PMI and CLST (DGT), Forrest AR, Kawaji H, Rehli M, Baillie JK, de Hoon MJ, et al. A promoter-level mammalian expression atlas. *Nature* 2014;507:462–70.
- Alvarez MJ, Shen Y, Giorgi FM, Lachmann A, Ding BB, Ye BH, et al. Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat Genet* 2016;48:838–47.
- Wang K, Saito M, Bisikirska BC, Alvarez MJ, Lim WK, Rajbhandari P, et al. Genome-wide identification of post-translational modulators of transcription factor activity in human B cells. *Nat Biotechnol* 2009;27:829–39.
- Gautier L, Cope L, Bolstad BM, Irizarry RA. affy-analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* 2004;20:307–15.
- Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;9:357–9.
- Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010;26:841–2.
- Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010;26:139–40.
- McCarthy DJ, Campbell KR, Lun AT, Wills QF. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* 2017;33:1179–86.
- O'Connell JB, Maggard MA, Ko CY. Colon cancer survival rates with the new American Joint Committee on Cancer sixth edition staging. *J Natl Cancer Inst* 2004;96:1420–5.
- Li H, Courtois ET, Sengupta D, Tan Y, Chen KH, Goh JLL, et al. Reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors. *Nat Genet* 2017;49:708–18.
- Munoz J, Stange DE, Schepers AG, van de Wetering M, Koo BK, Itzkovitz S, et al. The Lgr5 intestinal stem cell signature: robust expression of proposed quiescent '+4' cell markers. *EMBO J* 2012;31:3079–91.
- Barker N, Ridgway RA, van Es JH, van de Wetering M, Begthel H, van den Born M, et al. Crypt stem cells as the cells-of-origin of intestinal cancer. *Nature* 2009; 457:608–11.
- Du L, Wang H, He L, Zhang J, Ni B, Wang X, et al. CD44 is of functional importance for colorectal cancer stem cells. *Clin Cancer Res* 2008;14:6751–60.
- Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 2006;126: 663–76.

## Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grants funded by the Korea Government, the Ministry of Science and ICT (2017R1A2A1A17069642 and 2015M3A9A7067220). It was also partially supported by KAIST Grand Challenge 30 Project.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received April 27, 2019; revised August 7, 2019; accepted September 23, 2019; published first January 2, 2020.

43. Chen K, Zhang F, Ding J, Liang Y, Zhan Z, Zhan Y, et al. Histone methyltransferase SETDB1 promotes the progression of colorectal cancer by inhibiting the expression of TP53. *J Cancer* 2017;8:3318–30.
44. Wei D, Kanai M, Huang S, Xie K. Emerging role of KLF4 in human gastrointestinal cancer. *Carcinogenesis* 2006;27:23–31.
45. Munera JO, Sundaram N, Rankin SA, Hill D, Watson C, Mahe M, et al. Differentiation of human pluripotent stem cells into colonic organoids via transient activation of BMP signaling. *Cell Stem Cell* 2017;21:51–64.
46. Liu C, Banister CE, Weige CC, Altomare D, Richardson JH, Contreras CM, et al. PRDM1 silences stem cell-related genes and inhibits proliferation of human colon tumor organoids. *Proc Natl Acad Sci U S A* 2018;115:E5066–75.
47. Fodde R, Brabletz T. Wnt/beta-catenin signaling in cancer stemness and malignant behavior. *Curr Opin Cell Biol* 2007;19:150–8.
48. Malta TM, Sokolov A, Gentles AJ, Burzykowski T, Poisson L, Weinstein JN, et al. Machine learning identifies stemness features associated with oncogenic dedifferentiation. *Cell* 2018;173:338–54.
49. Dow LE, O'Rourke KP, Simon J, Tschaharganeh DF, van Es JH, Clevers H, et al. Apc restoration promotes cellular differentiation and reestablishes crypt homeostasis in colorectal cancer. *Cell* 2015;161:1539–52.
50. Cho KH, Lee S, Kim D, Shin D, Joo JI, Park SM. Cancer reversion, a renewed challenge in systems biology. *Curr Opin Syst Biol* 2017;2:49–58.
51. Cho KH, Joo JI, Shin D, Kim D, Park SM. The reverse control of irreversible biological processes. *Wiley Interdiscip Rev Syst Biol Med* 2016;8:366–77.
52. de Thé H. Differentiation therapy revisited. *Nat Rev Cancer* 2018;18:117–27.
53. Cruz FD, Matushansky I. Solid tumor differentiation therapy - is it possible? *Oncotarget* 2012;3:559–67.
54. Bilodeau S, Kagey MH, Frampton GM, Rahl PB, Young RA. SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state. *Genes Dev* 2009;23:2484–9.
55. Ceol CJ, Houvras Y, Jane-Valbuena J, Bilodeau S, Orlando DA, Battisti V, et al. The histone methyltransferase SETDB1 is recurrently amplified in melanoma and accelerates its onset. *Nature* 2011;471:513–8.

# Molecular Cancer Research

## Network Inference Analysis Identifies SETDB1 as a Key Regulator for Reverting Colorectal Cancer Cells into Differentiated Normal-Like Cells

Soobeom Lee, Chansu Lee, Chae Young Hwang, et al.

*Mol Cancer Res* 2020;18:118-129.

**Updated version** Access the most recent version of this article at:  
<http://mcr.aacrjournals.org/content/18/1/118>

**Cited articles** This article cites 55 articles, 9 of which you can access for free at:  
<http://mcr.aacrjournals.org/content/18/1/118.full#ref-list-1>

**E-mail alerts** [Sign up to receive free email-alerts](#) related to this article or journal.

**Reprints and Subscriptions** To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at [pubs@aacr.org](mailto:pubs@aacr.org).

**Permissions** To request permission to re-use all or part of this article, use this link <http://mcr.aacrjournals.org/content/18/1/118>. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.