Utilizing a dual endogenous reporter system to identify functional regulators of aberrant stem cell and differentiation activity in colorectal cancer

Sandor Spisak[1,2#], David Chen[1,3#], Pornlada Likasitwatanakul[1,4,5,6#], Paul Doan[1,5#], Zhixin Li[1,5], Pratyusha Bala[1,5], Laura Vizkeleti[7], Viktoria Tisza[2], Pushpamail De Silva[1], Marios Giannakis[1,4,5,8], Brian Wolpin[1,4,8], Jun Qi[9], Nilay S. Sethi[1,4,5,8 *]

[1] Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA;

[2] Institute of Enzymology, HUN-REN Research Centre for Natural Sciences, Budapest, Hungary

[3] Temerty Faculty of Medicine, University of Toronto, Toronto, ON, Canada;

[4] Department of Medicine, Harvard Medical School, Boston, MA, USA;

[5] Broad Institute of Massachusetts Institute of Technology (MIT) and Harvard University, Cambridge, MA, USA;

[6] Department of Medicine, Faculty of Medicine Siriraj Hospital, Bangkok, Thailand;

[7] Department of Bioinformatics, Faculty of Medicine, Semmelweis University, 1094 Budapest, Hungary

[8] Gastrointestinal Cancer Center, Dana-Farber Cancer Institute, Boston, MA, USA;

[9] Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA, USA;

[#]These authors contribute equally

*Materials and Correspondence:

Nilay Sethi, M.D., Ph.D.

Department of Medical Oncology

450 Brookline Avenue, Dana 740

Dana-Farber Cancer Institute, Boston, MA 02215

Phone: (617) 632 3693; Fax: (617) 632 3072;

E-mail: nilay_sethi@dfci.harvard.edu

Keywords: stem cell program, intestinal differentiation, genome editing, colorectal cancer

Abstract

Aberrant stem cell-like activity and impaired differentiation are central to the development of colorectal cancer (CRC). To identify functional mediators that regulate these key cellular programs in CRC, we developed an endogenous reporter system by genome-editing human CRC cell lines with knock-in fluorescent reporters at the SOX9 and KRT20 locus to report aberrant stem cell-like activity and differentiation, respectively, and then performed pooled genetic perturbation screens. Constructing a dual reporter system that simultaneously monitored aberrant stem cell-like and differentiation activity in the same CRC cell line improved our signal to noise discrimination. Using a focused-library CRISPR screen targeting 78 epigenetic regulators with 542 sgRNAs, we identified factors that contribute to stem cell-like activity and differentiation in CRC. Perturbation single cell RNA sequencing (Perturb-seq) of validated hits nominated SMARCB1 of the BAF complex (also known as SWI/SNF) as a negative regulator of differentiation across an array of neoplastic colon models. SMARCB1 is a dependency in CRC and required for *in vivo* growth of human CRC models. These studies highlight the utility of a biologically designed endogenous reporter system to uncover novel therapeutic targets for drug development.

Introduction

Genetic screening is a powerful tool to identify functional components of biological processes in a specific and unbiased manner. CRISPR-Cas9 technology has been the method of choice and initially deployed to reveal cancer dependencies[1-4]. As such, CRISPR-Cas9 screens in different cell types and conditions have typically focused on viability[5,6] and drug sensitivity as a phenotypic readout[7-12]. For example, a genome-wide genetic screen detected GRB7 as a driver of resistance to MEK inhibitor in KRAS-mutant colon cancer[13]. While viability remains an important readout, it can also lead to identification of essential genes and/or nonspecific hits[14].

Applying genetic screens to functional pathway readouts allow for identification of factors that regulate disease-specific phenotypes[15,16]. Aberrant activation of WNT signaling, which supports stem cell behavior in the intestines, is a hallmark of CRC initiation, most often through loss-of-function *APC* mutations[17-19]. Genetic screens in cancer cell lines expressing an exogenous canonical Wnt/β-catenin reporter identified novel regulators such as CDK8[20,21]. Since exogenous reporters may not fully recapitulate the aberrant rewiring in cancer, an endogenous WNT reporter by knocking in a fluorescent probe into the genomic locus of *CTNNB1* was employed for a genetic screen, revealing unique hits when compared an exogenous (i.e., TOPFLASH) WNT reporter system, including epigenetic regulators, cell cycle factors and G protein-coupled receptor pathway components [22]. These functional screens relied on capturing a loss of signal, which has a greater potential for false positives compared to screens using a gain of signal[23], and a single exogenous or endogenous reporter readout, which may be limited in its ability to faithfully capture pathway activity.

Disrupting the balance between stem cell and differentiation programs is a defining property of CRC[24-26]. As such, genomic alterations that hinder intestinal differentiation, either by activating stem cell-like programs or inactivating pro-differentiation pathways, are central to CRC development. Our group recently identified SOX9 as a key mediator of impaired differentiation by mediating aberrant stem cell-like activity in CRC[27]. Keratin 20 (KRT20) is a well-recognized marker of differentiated intestinal cells that is notably absent in normal stem cells and suppressed in models of cancer initiation[27-31]. To translate this understanding into a cellular platform for discovery, we engineered single and dual endogenous reporter systems by knocking-in fluorescent probes at both the SOX9 and KRT20 genomic loci to measure the aberrant stem cell-like and differentiation activity in CRC, respectively[32]. Applying CRISPR screens to our system, we nominate new epigenetic factors that regulate stem cell and differentiation programs in CRC.

Results

Development of an endogenous reporter by genome editing *SOX9* locus

SOX9 functionally blocks differentiation by activating an aberrant stem cell-like transcriptional program in human CRC[27,33]. SOX9 suppression induces differentiation and impairs CRC growth. These results implicate a dependency of CRC on specific stem cell programs and inspire therapeutic approaches directed at promoting intestinal differentiation. Based on these findings, we reasoned that genetic perturbations disrupting aberrant stem cell-like signaling and inducing intestinal differentiation will serve as targets for therapeutic development in CRC. To this end, we engineered an endogenous reporter system by introducing a fluorescent probe into genomic loci that readout aberrant stem cell-like (i.e., SOX9) and intestinal differentiation (i.e., KRT20) activity in CRC cells (Supplementary Table 1). To establish the SOX9 reporter, we knocked-in a cassette containing GFP and neomycin antibiotic resistance in-frame at the end of the SOX9 coding region of a CRC cell line using a combination of CRISPR/Cas9 technology and template-based homologous recombination [34] (Fig.1a). Engineered CRC cells were propagated in media containing neomycin to select populations with in-frame integration. Accurate genomic integration was confirmed using site-specific PCR with primers against genomic locus and cassette (Supplementary Fig. 1a). Compared to parental LS180 CRC cells, engineered GFP knock-in LS180 (LS180$^{SOX9\text{-}GFP}$) displayed GFP expression (Fig. 1b and Supplementary Fig. 1b). Although 13% of the LS180$^{SOX9\text{-}GFP}$ did not express GFP, this distribution remained consistent following culture of pure GFP(+) and GFP(-) fractions, suggesting plasticity among these populations (data not shown). To validate that GFP faithfully reflects SOX9 expression, two constitutively expressed SOX9 shRNAs were independently introduced into LS180$^{SOX9\text{-}GFP}$ cells, which led to a greater than 50% reduction of GFP levels (Fig. 1c and Supplementary Fig. 1c, d). These studies established an endogenous reporter CRC cell line that reflects aberrant stem cell-like activity through GFP expression.

To evaluate the performance of our platform in a pooled genetic perturbation format, we performed sgRNA and shRNA screens using custom libraries consisting of 76 sgRNAs or 154 shRNAs targeting positive and negative control genes, including *SOX9* and *GFP* (Supplementary Table 3). After introducing the libraries and sorting cells based on GFP expression at two different time points, genomic DNA was sequenced to determine sgRNA or shRNA representation in the high, low, and negative GFP fractions (Fig. 1d). To account for variation in sample preparation, sequencing depth, we normalized by 1) total read count, 2) plasmid library pool, and 3) population before comparing the sorted cell fractions (Fig. S1e). As expected, GFP and SOX9 sgRNAs and shRNAs were consistently depleted in GFP high fractions compared to GFP low and negative

fractions (Fig. 1e, f), validating the SOX9 reporter line in the genetic screening format. Confirming a consensus among genetic screens, we noted that CRISPR perturbations provided stronger and more consistent discriminatory power system compared to shRNA-mediated suppression (Fig. 1e). We also observed that screening performance at 7 to 10 days of culture following library introduction was superior to 3 days of cultures (Fig. 1f).

After validating the platform in a pooled screening format, we next performed a discovery CRISPR screen focused on druggable epigenetic regulators and their family members (Supplementary Table 2) [35-37], asking which genes when perturbed reduce SOX9 expression. To demonstrate the broad utility of the reporter system, we utilized another CRC cell line and different fluorescent probe. 542 sgRNAs targeting 78 genes associated with epigenetic regulation were introduced into HT29$^{SOX9\text{-}mKate2}$ cells (Supplementary Table 3). Cells were sorted into four even quartiles based on mKate2 expression and comparisons of normalized sgRNA abundance was made between each quartile (25-50%, 50-75%, 75-100%) and the quartile with the lowest mKate2 expression (0-25%). As expected, we found that sgRNAs targeting *SOX9* and *mKate2* were depleted from all three (25-50%, 50-75% and 75-100%) quartiles compared to the low (0-25%) mKate2 expressing subpopulation (Fig. 1*g*).

To estimate the degree of selection after a gene perturbation, we used MaGeCK Maximum Likelihood Estimation (MLE) to generate a beta score based on the differences in normalized sgRNA abundance when comparing mKate2 fractions. We identified several candidate genes, including but not limited to *SUZ12*, *SMARCD2*, *DNMT1*, and *KMT2A*, which when perturbed suppressed aberrant stem cell-like activity in CRC (Fig. 1h). We then applied a rank sum scoring method to identify consistently depleted sgRNAs in the mKate2 high fraction across replicates. We considered genes that had at least 2 or more targeting sgRNA within the bottom 15% percentile of rank sum scores as stronger candidates (Fig. 1i). As expected, sgRNAs targeting *SOX9* and *mKate2* were depleted in the mKate2 high fraction, along with sgRNAs targeting members of the histone deacetylase family (*HDAC2*, *HDAC4*), the sirtuin family (*SIRT1*, *SIRT2*, *SIRT4*), and the SWI/SNF complex (*SMARCA2, SMARCB1, SMARCD2*, *SMARCE1*). Together, these data support the use of an endogenous SOX9 reporter system for functional genetic screens that may lead to the identification novel regulators of aberrant stem cell-like behavior in CRC.

Endogenous differentiation reporter by genome editing *KRT20* locus

A principle reason that aberrant stem cell-like activity is selected as an early event in CRC pathogenesis is to block differentiation and prevent cell death [25,27,29]. In other words, aberrant stem cell-like signaling antagonizes intestinal differentiation in colorectal cancer initiation, enabling the persistence of neoplastic colon cells rather than their turnover in a rapidly renewing epithelium. We therefore sought to develop an endogenous reporter platform that faithfully captures differentiation activity. KRT20 is a well-recognized marker of differentiated intestinal cells[29,30] that is suppressed upon neoplastic initiation as shown in an Apc^{KO} mouse model (Fig. 2a). We integrated the GFP cassette into the KRT20 genomic locus of HT-29 and LS180 CRC cell lines (Fig. 2b). To biologically validate the system, we asked whether disrupting SOX9 induced differentiation reporter activity based on our previous results[27,33]. HT-29$^{KRT20\text{-}GFP}$ reporter line stably expressing a shRNA against SOX9 led to a shift in the number of GFP+ cells (Supplementary Fig. 2a, 0.3% to 62.4%), confirming that induction of differentiation via SOX9 suppression is captured by our endogenous differentiation reporter system.

To validate the endogenous KRT20 reporter in a pooled genetic perturbation format, we applied our control shRNA and sgRNA libraries to the engineered cell lines (Supplementary Table 3). After introducing the libraries into HT-29$^{KRT20\text{-}GFP}$ cells, we determined shRNA and sgRNA distribution in GFP high versus low sorted fractions using 25% gating. Similar to the endogenous SOX9 reporter, the endogenous KRT20 reporter demonstrated greater discrimination with the CRISPR library at day 7 (Fig. 2c, d and Supplementary Fig. 2b, c). As expected, sgRNAs targeting *KRT20* and *GFP* were depleted in GFP high fraction compared to GFP low fraction, providing technical validation, whereas control sgRNAs showed a normal distribution among fractions (Fig. 2c-e). Notably, SOX9 sgRNAs were enriched in the GFP high fraction (Supplementary Fig. 2d, e), providing biological validation that disrupting SOX9 activity induces the differentiation reporter. The enrichment of SOX9 sgRNAs gradually improved by gating a more stringent GFP high population (Supplementary Fig. 2e). These results provide strong evidence that the KRT20 reporter system is powered to detect genetic perturbations that promote intestinal differentiation of CRC.

We next applied the sgRNA library targeting epigenetic regulators to the HT29$^{KRT20\text{-}GFP}$reporter cell line. By comparing MLE beta scores of sgRNA abundance in GFP high versus low fractions, we identified several epigenetic regulators that support differentiation upon perturbation, including but not limited to *HDAC3, SET, AEBP2, and RBBP4* (Fig. 2f). We then applied the rank sum scoring method to replicates of the epigenetic regulator screen in HT29$^{KRT20\text{-}GFP}$ cells (Fig. 2g). In addition to confirming our positive and negative controls, we found that perturbations in core histone binding

proteins (*RBBP4*), DNA methyltransferase family members (*DNMT3B*), and the histone lysine methyltransferase family members (*SETD1A*, *SETD1B*) induced differentiation by reporter activity. These results underscore the utility of coupling functional CRISPR screens with endogenous reporters to identify functional regulators of cancer-specific properties.

Dual stem cell and differentiation reporter system

While the single-reporter systems were successful, we observed considerable variability in individual sgRNA scores (Fig. 1i and 2g) and few genes that once perturbed disrupted both stem cell-like activity and induced differentiation. We therefore sought to improve signal to noise and identify perturbations that induced differentiation by potentially disrupting stem cell-like activity. To this end, we engineered a dual endogenous reporter system to simultaneously monitor aberrant stem cell-like and differentiation programs within the same cell (Fig. 3a). HT29 CRC cells were edited to express mKate2 from the SOX9 locus and GFP from the KRT20 locus (HT29$^{\text{SOX9-mKate2/}\textit{KRT20-GFP}}$). By applying bulk RNA-sequencing to flow-sorted cells, we validated that the SOX9-mK2^{high} fraction demonstrates greater expression of stem cell genes, whereas the KRT20-GFP^{high} fraction demonstrates elevated expression of differentiation genes (Figure 3b,c). SOX9-mK2^{high}/KRT20^{-}GFP^{low} demonstrated elevated expression of stem cell markers *ASCL2*, *SMOC2* and *LGR5* (Fig 3c). KRT20-GFP^{high}/SOX9-mK2^{low} (highly differentiated) cells displayed elevated differentiation markers as indicated by increased *KRT20*, *MUC2* and *DPP4* expression. These findings were further confirmed by gene set enrichment analyses (GSEA) (Fig 3b). After introduction of our control sgRNA library, we performed two-color flow at day 7, setting the threshold at 2.5% (Supplementary Fig. 3a). We then compared the enrichment of sgRNAs between the mKate2^{low}/GFP^{high} fraction and the mKate2^{high}/GFP^{low} fractions (i.e., GFP vs. mKate2). We observed that sgRNAs against *SOX9* and *mKate2* were the most enriched, whereas sgRNAs against *KRT20* and *GFP* were the most depleted (Fig. 3d and Supplementary Fig. 3b). Reassuringly, the negative control sgRNAs showed no preference for either bin (Fig. 3d).

We established two additional dual endogenous reporter cell lines, HT115$^{\text{SOX9-mKate2/}\textit{KRT20-GFP}}$ and LS180$^{\text{SOX9-mKate2/}\textit{KRT20-GFP}}$, confirming utility of this system in other CRC cell lines for functional genetic screens (Supplementary Fig. 3c,d). Notably, the magnitude of sgRNA discrimination was greater in the dual reporter compared to single reporter system, improving our ability to identify genes that regulate this specific aberrant stem cell-differentiation axis in CRC (Supplementary Fig. 3b compared to Supplementary Fig. 2b). To control for potential technical fluorescent color bias, we engineered dual reporters by swapping fluorescent probes: *GFP* was expressed from the *SOX9* locus

and *mKate2* from the *KRT20* locus (Supplementary Fig. 3e-g). Consistent with the original dual reporter CRC cell lines, the ones with reversed fluorescent probes showed the expected distribution of *SOX9*, *GFP, KRT20*, and *mKate2* sgRNAs relative to controls. These data demonstrated the successful development of an endogenous dual stem cell-like and differentiation reporter system in CRC.

Identifying epigenetic regulators of aberrant stem cell-like and differentiation activity in CRC using dual endogenous reporter system

To identify new regulators of aberrant stem cell-like and differentiation activity, we applied the epigenetic sgRNA library to the HT29$^{\text{SOX9-mKate2/}KRT20\text{-}GFP}$ reporter cell line given its superior performance in discriminating control sgRNAs (SOX9, KRT20, mKate2, GFP) (Supplementary Fig. 3h-j). The top-scoring perturbations that induced differentiation/suppressed aberrant stem-cell activity (therapeutic targets) included *SOX9* as well as genes associated with the BAF chromatin remodeling complex (*SMARCA4, SMARCB1, SMARCD1, SMARCA5*), histone lysine methyltransferase (*SET, KMT2A, SETD2, DOT1L, EHMT2, EZH1*), and DNA methyltransferase (DNMT1) (Fig. 3e). Conversely, perturbations that may promote aberrant stem cell-like activity and impair differentiation (cancer-promoting genes) included bromodomain-containing proteins (*BRD1, BRD2*), histone acetyltransferases (*EP300*), and histone lysine methyltransferases (*KMT2C, NSD1, SETD1B*).

We analyzed perturbations that promote differentiation captured by dual and single-channel comparisons, demonstrating that dual-channel comparison (GFP vs. mKate2) consistently identified both control gRNA (*SOX9 , mKate2)* as well as exclusive candidate perturbations that induce differentiation compared to either single reporter system (Fig. 3g,h). Given the biological variability of gRNA knockout efficiency and the superior performance of the dual endogenous reporter system to discriminate candidate hits, we observed that different systems may identify different gRNA perturbation hits that target the same gene. Interestingly, the hits uniquely attributed to the mKate2^{low}/GFP^{high} and mKate2^{high}/GFP^{low} fraction comparison of the dual reporter system primarily consisted of genes associated with the SWI/SNF chromatin remodeling complex. Indeed, the rank sum scoring method to identify consensus perturbations across the three replicates revealed genes associated with the BAF chromatin remodeling complex (*SMARCA4*, *SMARCA5*, *SMARCB1*, *SMARCC1*, *SMARCD1*), DNA methyltransferase (*DNMT1*), histone lysine methyltransferase (*DOT1L*, *SETD2*), lysine methyltransferase (*KMT2A*), and sirtuin (*SIRT4*) (Fig. 3i).

SMARCB1 restricts intestinal differentiation in CRC

Based on our screen results, we focused on 12 genes that upon perturbation suppressed aberrant stem cell-like and activated differentiation reporter activity. Of note, most of the candidate genes encode proteins that are part of the BAF complex and histone methyl transferase family, save for 2 DNMTs and SIRT4 (Fig. 4a). Using the best performing sgRNAs, we examined whether individual gene knockout (KO) induced KRT20 and/or suppressed SOX9 in HT29 cells by immunoblot (Supplementary Fig. 4a,b). Perturbation of three genes – *KMT2A*, *SMARCA4*, and *SMARCB1* – increased KRT20 protein expression by immunoblot (Supplementary Fig. 4a, b).

To contextualize the effect of these genetic perturbations on differentiation beyond KRT20 expression, we performed single cell transcriptome analysis coupled with genetic perturbations (Perturb-seq)[38] for our validated hits in HT-29 CRC cells. Compared to non-targeting controls, cells with sgRNAs targeting *SOX9* and *CTNNB1* (encodes β-catenin) clustered near one another (Fig. 4b, left) and demonstrated similar transcriptional states involving activation of intestinal differentiation (Fig. 4b, right). While *KMT2A* KO did not meet technical standards, *SMARCA4* and *SMARCB1* perturbation led to distinct cell states with differentiation induction, albeit to a lesser extent than *SOX9* and *CTNNB1* inactivation (Fig. 4b, c). Further analysis revealed that broad multi-lineage differentiation was activated upon perturbation corresponding to suppression of a canonical stem cell program (Fig. 4d). These results indicate that specific BAF complex components may participate in CRC pathogenesis by restricting differentiation and perturbation of these factors promote differentiation, nominating new therapeutic targets.

We next focused on SMARCB1 given its consistent and potent ability to induce differentiation upon perturbation relative to KMT2A and SMARCA4 (Fig. 4c-d and Supplementary Fig. 4a-c). Indeed, shRNA-mediated SMARCB1 knockdown (KD) and individual clones with partial SMARCB1 KO (homozygous deletion was not observed) demonstrated induction of differentiation as indicated by elevated KRT20 expression (Fig. 4e). SMARCB1 KD reduced SOX9 and induced KRT20 mRNA expression in colon premalignant organoids from a patient with familial adenomatous polyposis (FAP), albeit to a lesser extent, likely due to strong stem cell cues from WNT/R-spondin/Noggin containing conditioned media (Supplementary Fig. 4d); similar results were achieved with KMT2A KD but not SMARCA4 KD (Supplementary Fig. 4d). SMARCB1 KD in a patient-derived CRC organoid (PDO) harboring *APC* and *KRAS* mutations reduced SOX9 and induced KRT20 mRNA and protein expression (Fig. 4f). Bulk RNA-sequencing followed by GSEA indicated the SMARCB1 KD induced a broad differentiation program (Fig. 4g); stem cell programs were not reduced, however, diverging from our cell line PERTURB-seq data. Furthermore, SMARCB1 KD enhanced the ability of pro-differentiation drug all-trans retinoic acid (ATRA)[39,40] to induce *KRT20* expression in human CRC

(Supplementary Fig. 4e). These results suggest that SMARCB1 restricts differentiation in different models of CRC.

To understand the functional relationship between SMARCB1 and SOX9, we started by measuring SOX9 expression levels in SMARCB1 KD lines, which showed a consistent ~50% reduction in *SOX9* mRNA transcripts in HT-29, HT-115, and PDO CRC models (Fig. 4f, h and Supplementary Fig. 4f). These results raised the possibility that SMARCB1 is part of the transcriptional regulation of SOX9 in CRC. Since SOX9 overexpression suppresses *KRT20* expression in CRC[27] (Supplementary Fig. 4g), we asked whether SMARCB1 was required for the SOX9-mediated differentiation block. SMARCB1 KD prevented the ability of SOX9 overexpression to suppress *KRT20* (Fig. 4i and Supplementary Fig. 4g), suggesting that SMARCB1 is required for SOX9-mediated impairment of differentiation.

SMARCB1 suppression impairs CRC growth *in vitro* and *in vivo*

Since promoting differentiation has anti-tumor effects in CRC, we next sought to characterize SMARCB1 dependency. Using Broad's Dependency Map portal[9], we analyzed the relative dependency for *CTNNB1*, *SOX9*, *SMARCB1*, *KMT2A*, and *SMARCA4* in microsatellite stable (MSS, n = 30) and instability (MSI, n = 12) CRC cell lines (Fig. 5a). Not surprisingly, *CTNNB1* and *SOX9* showed the strongest dependency across CRC cell lines, consistent with previous findings[27,30,33,41]. Among the newly implicated genes, *SMARCB1* showed the strongest dependency (M = -0.643, 95% CI = [-0.753, -0.533]) compared to *KMT2A* (P = 0.0001, M = -0.225, 95% CI = [-0.288, -0.163]) and *SMARCA4* (P = 0.0005, M = -0.265, 95% CI = [-0.358, -0.163]) (Fig. 5a), in proportion to their relative effects on differentiation (Fig. 4). While SOX9 shows a near-preferential dependency in MSS relative to MSI CRC cell lines ($t(40) = -1.814$, $p = 0.0772$), the impact of SMARCB1 loss was similar in CRC cell lines irrespective of MSI status ($t(40) = -0.656$, $p = 0.516$) (Fig. 5a). Among 30 MSS CRC cell lines, *SOX9* and *SMARCB1* showed comparable distribution of dependency strengths ($p = 0.190$, 95% CI = [-0.0764, 0.633]) (Supplementary Fig 5a). We next demonstrated that shRNA-mediated SMARCB1 KD significantly reduced *in vitro* proliferation (Fig. 5b) and *in vivo* tumor growth using a human CRC cell line xenograft (Fig. 5c and Supplementary Fig. 5b). Tumor xenografts maintained SMARCB1 KD throughout the experiment as shown by SMARCB1 IHC (Fig. 5c).

As PDOs are increasingly adopted as the favored pre-clinical model for their capacity to faithfully recapitulate patient response to small-molecule chemical perturbations[42,43], we next evaluated the requirement for SMARCB1 in this human CRC model ($APC^{c.835\text{-}8A>G}$, $KRAS^{G12V}$). SMARCB1 KD reduced organoid proliferation by 50% *in vitro*, which was associated with

morphological changes (Fig. 5d). SMARCB1 KD also impaired tumor growth *in vivo* using a xenograft model (Fig. 5e and Supplementary Fig. 5b). Two organoid xenografts escaped SMARCB1 KD targeted by their shRNA and showed greater tumor growth compared to controls, reinforcing selective pressure against SMARCB1 loss (Supplementary Fig. 5c). Tumor xenografts at experimental endpoint showed reduced *SMARCB1* and elevated *KRT20* mRNA expression by RT-PCR (Fig. 5e). Consistently, SMARCB1 KD xenografts showed reduced SMARCB1 and elevated KRT20 protein levels by histopathology and immunoblot (Fig. 5f,g, paired xenografts marked in different shades of blue). A second differentiation marker, DPP4, was also induced in SMARCB1 KD PDO xenografts (Fig 5f, g). Notably, two crypts that escaped SMARCB1 KD did not induce KRT20 or DPP4 expression (Fig. 5f, arrows). While stem cell marker LGR5 was modestly reduced, SOX9 levels did not change in SMARCB1 KD xenografts (Fig 5f). Interestingly, SMARCC1, a core member of all 3 BAF complexes, was only expressed in SMARCB1 KD xenografts, suggesting compensation for SMARCB1 loss (Fig 5g). Collectively, these results demonstrate a requirement for SMARCB1 in restricting differentiation and tumor growth in CRC.

Discussion

Genetic screens that utilized functional readouts have helped identify new regulators of biological processes and pathological conditions[44]. Impaired differentiation is a hallmark of cancer[24], perhaps most recognized and well-characterized in hematological malignancies. Acute myeloid leukemia (AML) displays an arrest in differentiation, which leads to expansion of an immature and self-renewing population of neoplastic cells. Of translational importance, therapies that promote myeloid differentiation have shown great efficacy in acute promyelocytic and IDH-mutant subtypes of AML[26]. These translational successes have motivated genetic and chemical screens to identify new regulators and drugs targeting differentiation blocks, respectively. By applying a chemical compound library to a conditional HoxA9-mediated differentiation block[45], inhibition of the enzyme dihydroorotate dehydrogenase was identified as potential pro-differentiation therapeutic strategy[46]. A genome-wide CRISPR screen in acute myeloid leukemia using CD14 expression as a marker of AML blasts demonstrated that ZFP36L2 promotes differentiation blockade by destabilizing myeloid maturation transcripts[47]. Similarly, using CD11b surface marker expression as a readout of myeloid differentiation, a CRISPR screen targeting chromatin remodelers and RNA binding proteins revealed that KAT6A and ENL collaborate in a pro-leukemogenicity transcription circuit that stalls differentiation[48].

Aberrant stem cell-like activity imparts differentiation defects in CRC development[27,29,31,33]. Given that unrestricted WNT activation is a critical initiating event in CRC pathogenesis, most frequently via loss-of-function *APC* mutations[17-19], it is natural that therapeutic discovery efforts and screening approaches have been directed towards disrupting this potent developmental pathway[30,49-51]. In DLD1 and HCT116 CRC cell lines engineered to express an exogenous β-catenin-responsive, luciferase-based reporter, *CDK8* was identified as being necessary for β-catenin-driven malignancy from a focused shRNA screen[52]. With a different exogenous WNT-responsive reporter approach on the same CRC cell lines, *TCF7* was identified as a positive regulator of CRC while *TCF7L2* was shown to have tumor suppressor properties in CRC[20]. When DLD1 cells were engineered to express a similarly designed, exogenous β-catenin-responsive reporter, *AGGF1* was found to regulate WNT/β-catenin signaling by a genome-wide siRNA screen[53]. An alternative genetic perturbation screening strategy by gene-trap retrovirus employed on a haploid CRC cell line engineered to express an exogenous Wnt-responsive, fluorescence-based reporter nominated a different set of regulators of WNT pathway in CRC[54]. Due to these broad results using exogenous reporter-based screens, efforts have been made to engineer endogenous reporters[31] compatible with forward genetic

screening. The histone modifier KMT2A was recently nominated to be a regulator of aberrant WNT/β-catenin signaling by a genome-scale CRISPR screen in DLD1 CRC cells engineered to express fluorescence-based reporter at the endogenous locus of either *AXIN2* or *cMYC*, direct targets of β-catenin[22]. They also found that there was a disconnect between exogenous and endogenous reporter results. These screens have relied on capturing a loss of signal, which has the potential for false positives compared a gain in signal [23], from a single exogenous or endogenous reporter readout. We therefore decided to manufacture a dual endogenous reporter system capable of capturing information from either pole of the stem cell-differentiation axis.

Our results indicate that utilizing the dual reporter improved signal-to-noise and consistency among replicates while reducing sgRNA variance in the CRISPR screen compared to each single reporter, providing greater confidence in nominated functional candidates. By evaluating gain and loss of signal from opposing pathological processes, we aimed to identify regulators along a specific axis and reduce the chances of false positive results. Another approach would have been to use both gain- and loss-of-function screening, which was recently shown using parallel CRISPRi and CRISPRa screens to investigate functional regulators of interferon and IL-2 signaling in T-cell cells[55]. The dual reporter can also facilitate identification of factors that only impact one reporter (e.g., differentiation) without influencing the other (e.g., stem cell activity). This platform can be scaled up to perform genome-wide CRISPR screens and is also suitable for small molecule chemical screens, which can be combined with high content imaging to accurately measure phenotypic readouts of stem cell and differentiation activity.

We focused our validation studies on three genes - *KMT2A*, *SMARCA4*, and *SMARCB1* - that are part of two functional modules (Fig. 4a). KMT2A (lysine methyltransferase 2A; also known as mixed lineage leukemia 1 or MLL1) is a methyl transferase that characterized for methylating lysine 4 on histone 3 (H3K4), which regulates enhancers that function early development and hematopoiesis [56]. Our data indicated that KMT2A negatively regulates differentiation in the colon. Disruption of KMT2A in zebrafish resulted in premature differentiation of neurons and decrease proliferation[57], suggesting that it may restrict differentiation across multiple tissue. With respective to cancer biology, KMT2A/MLL1 is altered by chromosomal rearrangement in poor prognosis leukemias[58]. In a mouse model of colon cancer, Kmt2a/Mll1 was required for WNT-driven adenomas by sustaining activating H3K4 tri-methylation at stem cell genes, including Gata4/6 transcription factors[59]. Consistently, a genome-scale CRISPR screen using an endogenous WNT reporter identified KMT2A as a regulator of canonical β-catenin transcriptional activity, and the loss of *KMT2A* by CRISPR-mediated knockout

led to induction of KRT20 and suppression of intestinal stem cell markers (*LGR5* and *ASCL2*) in CRC cells[22]. These studies suggest that suppression of KMT2A leads to intestinal differentiation by disrupting β-catenin/WNT signaling in our CRC models.

SMARCA4 and *SMARCB1* both encode DNA-binding subunits of the chromatin remodeling BAF complex[60]. *SMARCA4* encodes BRG1, which functions as an ATPase in the BAF complex. High expression of *SMARCA4* is observed in most cancers and correlates with poorer survival in a subset [61]. In a model of colon cancer, SMARCA4 and PRMT1 were shown to promote CRC progression cooperatively via EGFR signaling[62]. There are multiple drugs inhibiting SMARCA4 under preclinical evaluation[63,64]. Most recently, a SMARCA2/4 degrader disrupted promoter and enhancer interactions and suppressed tumor growth in prostate xenograft models[65].

SMARCB1, which encodes SNF5, is well-known for its tumor suppressor function across multiple cancers including pediatric CNS cancers and sarcomas[66,67]. Embryonic rhabdoid tumors harbor homozygous *SMARCB1* mutations[68]; SMARCB1 binds to and maintains enhancers involved in differentiation in these tumors[69]. Indeed, SMARCB1 engages lineage-specific enhancers to maintain developmental and differentiation programs[70]. Gene expression profiling of induced pluripotent stem cells and neural progenitors revealed that the transcriptional effect of SMARCB1 is context dependent[71]. Mutations in SMARCB1 also lead to neurodevelopmental diseases[72] and suppression of BAF complex components promote premature neuronal differentiation[73]. SMARCB1 loss led to induction of differentiation in our CRC models, which may be due to enhancer specificity in colon epithelium[69]; the underlying molecular mechanism through requires further elucidation. SMARCB1 loss has also been shown to induce β-catenin activity[74], which would promote stem cell behavior in colonic tissue and contradict our model. While a SMARCB1 inhibitor does not exist, SMARCB1-deficient tumors appear to be co-dependent on EZH2, which has an oral-form small molecule inhibitor. EZH2 inhibitors induced differentiation and apoptosis in SMARCB1-deleted malignant rhabdoid tumor xenograft[67,75]. If SMARCB1 inhibitors are developed, their therapeutical potential in CRC may be realized in combination with EZH2 inhibitors.

Study Limitations: The dual reporter relies on SOX9 and KRT20 as markers for aberrant stem cell-like and differentiation activity. While these are faithful markers, they may not capture the full complexity of these transcriptional programs. Perturbation of epigenetic regulators did not induce

differentiation to the same magnitude as SOX9 suppression. This may reflect that many of these factors function in complexes rather than individual proteins and their loss may lead a partial defect in complex activity. The impact of losing these factors on differentiation should also be examined using dynamic assays in addition to single timepoint experiment. The molecular mechanism(s) by which SMARCB1 restricts differentiation in CRC has not been addressed by this study and requires further investigation.

Materials and Methods

Cell Culture

All cell lines were maintained at 37 °C with 5% CO2. The human colorectal cancer cell lines were obtained from the CCLE core facility and used at early passage for the experiments. HT-115, HT-29, and LS180 cells were maintained in DMEM medium supplemented with 10% FBS and 1% penicillin/streptomycin.

Engineering endogenous stem cell and differentiation reporter system

We knocked-in mKate2 (mK2, a next-generation RFP probe) and GFP fluorescent markers in-frame at the end of the SOX9 and KRT20 coding regions using a combination of CRISPR/Cas9 and template-based homologous recombination to establish single and dual endogenous stem cell and differentiation reporter lines according to the detailed protocols (PCR template-mediated HDR, RNP complex assembly and transient transfection) below.

We designed and tested sgRNAs to target the last exon of *SOX9* in closest proximity to the stop codon. Cas9 pre-preloaded *in vitro* with the best performing sgRNA (see methods for T7 assay) was introduced into CRC cells for precise and efficient genome editing by electroporation. A double stranded break (DSB) followed by homology directed repair (HDR) facilitated integration of the GFP fluorescent reporter cassette.

PCR template-mediated HDR. Fluorophore reporter genes and antibiotic selection markers were amplified by universal primers flanked with SOX9 and KRT20 specific flaking primers to ensure the site-specific integration of the PCR products. Primer sequences are listed in Supplementary Table 1.

RNP complex assembly

AltR RNA oligos were ordered from IDT. After reconstitution (100 uM) tracrRNA and crRNA were mixed equimolarly and incubated for 5 min at 95C. After denaturation, products were incubated for 5 min at room temperature for complex formation.

Transfection and KI cell line preparation HT-29 cells were plated two days before transfection, to reach 70–80% confluency at the time of transfection. 0.5×10^6 cells per transfection were collected for each condition. Cells were transfected with RNP complex and HDR templates by nucleofection with SF Cell Line 4D-Nucleofector X Kit (Lonza) using 20-ml Nucleocuvette Strips, as described by the manufacturer (Program FF137). Cells were immediately resuspended in 100 ul of culturing

medium and plated into 1.5 ml of pre-warmed culturing medium in 24-well tissue culture plates. T7E1 assay, and site-specific PCR reactions were performed 72 h after nucleofection in order to check cleavage efficiency and integration, respectively. After confirming the correct integration and one-week regeneration cells were plated into 15 cm dish. The next day media was changed to antibiotic containing media (500 ug/ml neomycin and 20 ug/ul blasticidin) and antibiotic selection was performed. After an additional week, formed colonies were collected by trypsinization and resuspended in culture media. After regeneration culturing period (3-5 days), cell lines were tested by FACS analysis and site-specific PCR.

Composition of control sgRNA and shRNA library

Six sgRNAs/shRNAs per gene targeting GFP, mKate2, SOX9, KRT20 and sgRNAs/shRNAs targeting other genes including control genes in a total of 76 sgRNAs and 154 shRNAs (Supplementary Table 3).

Design of focused epigenetic sgRNA library

For focused epigenetic regulator screens, we created a small library from selected members of epigenetic regulator families (Histone Acetylation (BRD), Histone deacetylation (HDAC & SIRT), Histone Lysine Methyl transferase (SET & KMT), Chromatin remodeling BAF(SWI/SNF) complex, and DNA Methyltransferases (DNMT)) for a total of 78 genes. 466 sgRNAs were selected from the H3 knock out library (Addgene # 133914) from Brown/Liu labs.

Regarding the control sgRNAs, we used SOX9, KRT20, GFP, mKate2, PROM1, LRIG1 as positive controls and non-targeting sgRNAs as negative controls. sgRNAs were designed using the Chop-Chop algorithm [76]. The total of 542 sgRNA were included in the library (Supplementary Table 2). Its abundance was determined by deep amplicon sequencing.

Oligo design, cloning and library preparation

The pCC_01 - hU6-BsmBI-sgRNA(E+F)-barcode-EFS-Cas9-NLS-2A-Puro-WPRE plasmid (Addgene #139086) was used for CRISPR screen and validation. Briefly the plasmid was digested with Esp3I (NEB#R0734) at 37°c overnight. The 13 kb band was purified with gel purification kit (Takara Nucleospin® Gel and PCR Clean-up Midi). Eighty-two base pair single stranded oligos were designed according to the following structure:

AGGCACTTGCTCGTACGACGCGTCTCGCACC - gNNNNNNNNNNNNNNNNNNN - GTTTAGAGACGTTAAGGTGCCGGGCCCACAT

31 base pair 5` constant sequence compatible with the linearized (***Esp3I***) pCC_01 vector (3` end of the U6 promoter) followed by the 19 bp sgRNA sequence adding an extra g base for the proper sgRNA transcription, followed by the 3`vector compatible (5` end of the sgRNA scaffold) sequence. Single stranded oligo pools were ordered from IDT (oPools™ Oligo Pools) and built into the digested PCC_01 vector using NEBuilder HiFi DNA Assembly Master Mix (E2621L, New England Biolabs) according to the manufacturer`s instruction. The list of all oligos can be found in the Supplementary Table 3. Assembled gRNA libraries were purified by AMPure XP magnetic beads (Beckman Coulter).

Flow cytometry and cell sorting (FACS)

FACS was performed on the FACSAria III platform (BD) at DFCI Flow-core facility. For GFP labelled cell collection, 488 nm laser and 585/42 filter were used. For mKate2 651 nm laser and 610/20 filter were used. For single channels, 4 fractions were collected based on histogram profile, using 4-way sorting, approximately 25% per each of under curve area. We named these fractions: low (0-25%), mid low (25-50%), mid high (50%-75%), high (75-100%). For dual reporter, 4 fractions were collected, 2.5% of each, including dual negative, dual positive, GFP positive and mKate2 positive fractions. Cells were collected into 5 ml tubes. After collecting cell pellet, direct PCR was performed using Phire Tissue master mix (F170L, Thermo), according to the manufacturer recommendations. Guide RNA abundance was determined from each fraction by deep amplicon sequencing. Library preparation was performed according to the Broad Institute recommended protocol at the GPP portal using Argon primers for the demultiplexing. (https://portals.broadinstitute.org/gpp/public/resources/protocols)

Individual gRNA cloning for validation

All the sequences associated with the validation are listed in the Supplementary Table 4.

For the validation, two individual oligos, top and bottom were ordered from IDT as designed.

Top strand: protospacer sequence 5’ – CACCGNNNNNNNNNNNNNNNNNNNN—3’

Bottom strand: reverse complement 5’—AAACNNNNNNNNNNNNNNNNNNNNC—3’

The oligos were phosphorylated with T4 PNK enzyme (NEB #M0201) and annealed after denaturation. The digested plasmid and phosphorylated oligo cassettes were ligated with T4 DNA ligase (NEB #M0202L) at room temperature for 10 minutes. The ligated vectors were introduced by heat shock transformation into NEB Stable (NEB #C3040H) chemically competent *E. coli* and propagated on selective agar plates. Colonies were picked and the protospacer sequences were

confirmed with Sanger sequencing (Genewiz). Plasmid DNAs were extracted with QIAGEN Plasmid Plus MIDI kit (QIAGEN 12941) for downstream applications.

pLKO.1-TRC plasmid (Addgene #10878) was used for shRNA constitutive knockdown in cancer cells. The plasmid was digested with AgeI (NEB #R3552) and EcoRI (NEB #R3101L) at 37°c overnight. The 7kb band was purified as above. The single-stranded pool oligos were order from IDT.

5' – CTTTATATATCTTGTGGAAAGGACGAAACA CCGG NNNNNNNNNNNNNNNNNNNNNN CTCGAG nnnnnnnnnnnnnnnnnnn TTTTT AATTCTCGACCTCGAGACAAATGGCAGTATT—3'

The digested plasmid and pool oligos were ligated with Gibson reaction (NEBuilder HiFi DNA Assembly Master Mix E2621) at 50°c for 1 hour and purified with 1.0x AmPure Bead (AmPure XP A63881) before the transformation. Each bacterial colony was picked and Sanger sequenced before proceeding to plasmid propagation.

pRSITEP-U6Tet-sh-noHTS-EF1-TetPep-2A-Puro plasmid (Cellecta #SVSHU6TEP-L-CT) was used for shRNA inducible knockdown in colon organoids. The plasmid was digested with BbsI (NEB R3539L) at 37°c overnight and purified with 1.0x AmPure Bead. Two individual oligos, top and bottom were ordered from IDT as designed.

Top strand: hairpin sequence 5' – G NNNNNNNNNNNNNNNNNNNN CTCGAG nnnnnnnnnnnnnnnnnnn TTTT —3'

Bottom strand: reverse complement 5'—CGAA AAAA NNNNNNNNNNNNNNNNNNNN CTCGAG nnnnnnnnnnnnnnnnnnn C —3'

The oligos were phosphorylated and annealed with T4 PNK enzyme. The digested plasmid and phosphorylated oligos were ligated with T4 DNA ligase at room temperature for 10 minutes. The ligated vectors were heat-shock transformed to NEB Stable competent *E. coli.* The plasmids were extracted with QIAGEN Plasmid Plus MIDI kit and the hairpin sequences were confirmed with Sanger sequencing.

pLIX403-Blasticidin (Addgene #158560) was used for doxycycline-inducible SOX9 overexpression. SOX9 full length and GFP in donor plasmids were swapped with CcdB toxin using gateway cloning as described in the paper [27].

Lentivirus transfection for validation

To generate lentiviruses for validation experiments, HEK293T cells were co-transfected with sgRNA expression vectors and lentiviral packaging constructs psPAX2 and pMD2.G (VSV-G) in a 2:1:1 ratio using X-tremeGENE 9 DNA Transfection Reagent (Roche) according to the manufacturer's instructions. Cell culture media was changed the following day and lentiviral supernatant was

harvested 24 h and 48 h later and filtered through a 0.22 µm filter (Celltreat). Lentiviruses were aliquoted and stored at - 80 °C until use.

To perform lentiviral infection, the CRC cells were plated in a 6-well flat bottom plate and infected with 0.5 mL virus in media containing 8 mg/mL polybrene overnight. The media was changed the next day. After 2 days, puromycin or blasticidin were started at 5-8 ug/mL and 20 ug/mL respectively and continued until the negative control cells died.

CRISPR knockout single clone derivation

HT29 cell line with SMARCB1 CRISPR knock down was trypsinized and filtered through 20 um then 10 um filter in order to make single cells. Single cells were counted and plated at 1,000 cells in 10 cm culture dish. After 1-2 weeks, multiple clones were picked and collected DNA to confirm genomic sequence around the cut site.

ATRA treatment

HT29 was treated with all-trans retinoic acid (ATRA) at 10 uM concentration for 48 hours.

Immunoblot analysis

Cells were lysed in RIPA buffer supplemented with a protease inhibitor cocktail (Roche) and phosphatase inhibitor (Cell Signaling 5870s). Protein were measured with Pierce™ Rapid Gold BCA Protein Assay Kit (Thermo Fisher A53227) and denatured in 4x Laemmli buffer (Biorad 1610747). Whole cell extracts were resolved by 8-16% Tris-glycine polyacrylamide gel (Invitrogen XP08165BOX), transferred to PVDF membranes with iBlot™ 2 transfer device (Thermo Fisher IB21001) using P0 protocol, and probed with indicated primary antibodies. Bound antibodies were detected with LI-COR IRDye ® 680/800CW anti-rabbit/mouse secondary antibodies. Primary antibodies listed in the Supplementary Table 5.

RNA isolation, RT-qPCR

Total RNA was isolated using the RNeasy Mini Kit (Qiagen 74004) and cDNA was synthesized using the iScriptTM Reverse Transcription Supermix for RT-qPCR (Bio-Rad 1708840). Gene-specific primers for SYBR Green real-time PCR were synthesized by Integrated DNA Technologies. RT-qPCR was performed and analyzed using CFX96 Real-Time PCR Detection System (Bio-Rad 1845096) and using Power SYBR Green PCR Master Mix (Thermo Fisher 4368577). Relative mRNA expression was determined by normalizing to ACTB expression.

Pooled CRISPR screen analysis

Abundance of sgRNA reads was counted using a custom Python-based adaptation of the Aho-Corasick string-search algorithm to 1) demultiplex FASTQ files and assign each read to a known cell fraction by searching for the associated barcode in each read and to 2) read count the number of sgRNA in each cell fraction by searching for the sgRNA sequence in each read without allowing for any mismatches.

Normalization of sgRNA representation

First, total read count normalization assumes that the total expression of sgRNA reads is the same under the different experimental conditions. The total read count normalization step divides the read count for each sgRNA by the total number of read counts across all sgRNA in the sorted cell fraction. After dividing by total sample size, the output normalized read counts represent the proportion of each sgRNA read's abundance relative to the total sgRNA read count Second, library pool normalization considers the uneven sgRNA representation within the library and assumes that each sgRNA read within the pool should have similar read counts on average in each sorted cell fraction. The library pool normalization step divides the read count for each sgRNA in every sample by the mean of the read counts for each sgRNA from 3 independent library pool control populations. Third, population normalization assumes that most sgRNA should have similar read counts on average between different sorted cell fractions from the same cell population. The population normalization step divides the read count for each sgRNA in each fraction by the total read count for each gRNA across all fractions of the cell population. Normalized sgRNA read counts are compared between different cell fractions sorted based on fluorescence intensity using the single and dual endogenous reporter systems.

Flow sorting using single endogenous reporter system

Cell populations with the knock-in single endogenous reporter system were sorted based on fluorescence intensity into four equal fractions (0-25**%,** 25-50%, 50-75%, and 75%-100% fluorescence intensity) to represent the gradual progression of cell state differentiation (using GFP) or the stem cell program (using mKate2) in colorectal cancer cell lines. Three informative pairwise comparisons of sgRNA read count abundance to the cell fraction with the lowest fluorescence intensity range (0-25%) were made: 25-50% vs. 0-25%, 50%-75% vs. 0-25%, and 75-100% vs. 0-25**%.**

Flow sorting using dual endogenous reporter system

Four fractions for each sample were obtained using flow sorting based on fluorescent intensity of green fluorescent protein and red fluorescent protein using the dual endogenous reporter system. For cell lines with knock-in of the dual reporter system, the mKate2^{low}/GFP^{high} fraction represents cells within the population which exhibit relatively high green fluorescent intensity and the mKate2^{high}/GFP^{low} fraction represents cells within the population which exhibit relatively high red fluorescent intensity. The mKate2^{low}/GFP^{low} fraction and mKate2^{high}/GFP^{high} fraction represents cells within the population which exhibit relatively low and high green and red fluorescent intensity respectively. Between the four sorted cell fractions, five possible informative pairwise comparisons can be made to assess the differential fold change of normalized sgRNA abundance in different cell states: mKate2^{low}/GFP^{high} vs. mKate2^{high}/GFP^{low}, mKate2^{low}/GFP^{high} vs. mKate2^{low}/GFP^{low} , mKate2^{high}/GFP^{low} vs. mKate2^{low}/GFP^{low} , mKate2^{low}/GFP^{high} vs. mKate2^{high}/GFP^{high} , and mKate2^{high}/GFP^{low} vs. mKate2^{high}/GFP^{high} .

Differential sgRNA abundance analysis

Log_2 fold change of normalized sgRNA abundance and Z-score of normalized sgRNA abundance was used to compare differences in sgRNA abundance between different sorted cell fractions. The Model-based Analysis of Genome-wide CRISPR/Cas9 Knockout (MaGeCK) tool with default parameters was used to test for differential abundance of normalized sgRNA read counts between different sorted fractions of multiple technical replicates and estimate the beta score, a measure of gene essentiality, using MaGeCK Maximum Likelihood Estimation (MLE). Beta scores generated from differential sgRNA abundance between two different sorted cell fractions represent the extent of positive or negative gene selection across multiple technical replicates.

Identifying enriched/depleted sgRNA in CRISPR screens using the rank sum scoring method

Three technical replicates of the focused epigenetic screen using the dual endogenous reporter system in the HT-29$^{\text{SOX9-mKate2/}\textit{KRT20}\text{-GFP}}$ genome-edited cell line were conducted. Three technical replicates of the focused epigenetic screen using the single endogenous reporter system in the HT29$^{\textit{SOX9-mKate2}}$ and HT29$^{\textit{KRT20-EGFP}}$ genome-edited cell line were also conducted. Comparing the mKate2^{low}/GFP^{high} fraction and mKate2^{high}/GFP^{low} fractions in the dual reporter system was done since this comparison yielded the largest range of log_2 fold change of normalized sgRNA abundance. Comparing the fraction with the highest fluorescence intensity range (75-100%) and the fraction with

the lowest fluorescence intensity range (0-25%) in the single reporter system was done since this comparison yielded the largest range of $\log_2$ fold change of normalized sgRNA abundance.

To identify consistently enriched and depleted sgRNA hits in the focused epigenetic library across different replicates with different effect sizes, a rank sum method was used. First, the $\log_2$ fold change of normalized read count abundance between two fractions was determined for each sgRNA in each replicate. Second, within each replicate, the sgRNA were rank ordered from highest fold change to lowest fold change, such that sgRNA assigned a lower rank value (closer to 0) had a higher fold change value. Third, the ranks of each sgRNA were summed across each replicate to generate the rank sum score of each sgRNA. For sgRNAs targeting a gene to be considered enriched or depleted using their assigned rank sum scores, there must be at least 2 sgRNAs targeting the same gene that are within the top or bottom 15% of all sgRNA rank sum scores respectively. This rule was used to exclude single outlier sgRNA hits that were enriched or depleted despite other sgRNA targeting the same gene not being enriched or depleted.

PERTURB-seq

pCC_01 and pCC_09 (Addgene #139094) were modified to incorporate Capture Sequences 1 and 2 at the 3' end of guide sequences. Oligo cloning and library preparation were done as described above. The Capture Sequences allowed sgRNAs to be captured within the same cell when cells were processed according to the Chromium Single Cell 3' Reagent Kits v3.1 with Feature Barcoding technology for CRISPR screening protocol (10X Genomics, CG000205). For analysis,
FASTQ files of GEX and CRISPR libraries were processed with default parameters on 10X Genomics *Cell Ranger* pipeline (v7.0.0) using GRCh38-2020-A reference. Processed data were then analyzed with *Seurat* (v4.1.1) [77]and its extension *Mixscape* [78].

Depmap gene dependency analysis

Depmap is a cancer dependency map that systematically identified known genetic dependencies using genome-scale CRISPR knockout screens of cancer cell lines in the Cancer Cell Line Encyclopedia (CCLE). For each cell line in the CCLE, genes annotated with dependency scores below –1.0 are considered essential. To characterize the dependency of validated screen hits along with known dependencies involved in the initiation and progression of colorectal cancer (CTNNB1, SOX9), we compared the dependency score distribution of the validated gene hits in the 31 microsatellite stable (MSS) and 11 microsatellite instable (MSI) colorectal cancer cell lines of the CCLE with available dependency score data.

Patient-derived CRC organoid model

A tumor specimen from an unidentifiable patient with CRC was collected post colectomy under approval (protocol 13-189) by the Internal Review Board of the Dana Farber Cancer Institute, Boston, Massachusetts, USA.

Tumor tissue was treated with EDTA and then resuspended in 30-50 ul of Matrigel (BD Bioscience) and plated in 24-well plates. WNT/R-spondin/Noggin (WRN) containing DMEM/F12 with HEPES (Sigma-Aldrich) containing 20% FBS, 1% penicillin/streptomycin and 50ng/ml recombinant mouse EGF (Life Technologies) was used for culturing colon organoids. For the first 2-3 days after seeding, the media was also supplemented with 10 mM ROCK inhibitor Y-27632 (Sigma Aldrich) and 10 mM SB431542 (Sigma Aldrich), an inhibitor for the transforming growth factor (TGF)-β type I receptor to avoid anoikis. For passage, colon organoids were dispersed by trypsin-EDTA and transferred to fresh Matrigel. Passage was performed every 3-4 days with a 1:3–1:5 split ratio. For human colon organoid culture, the previous media was supplemented with antibiotics 100 ug/ml Primocin (Invivogen), 100 ug/ml Normocin (Invivogen); serum-free supplements 1X B27 (Thermo Fisher (Gibco)), 1X N2 (Thermo Fisher (Gibco)); chemical supplements 10mM Nicotinamide (Sigma), 500mM N-acetylcysteine (Sigma), hormone 50mM [Leu15]-Gastrin (Sigma), growth factor 100ug/ml FGF10 (recombinant human) (Thermo Fisher) and 500nM A-83-01 (Sigma), which is an inhibitor of the TGF-β Receptors ALK4, 5, and 7.

Notable genomic alterations in the patient-derived CRC organoid includes APC: c.835-8A>G (pathogenic intronic splice variant, rs1064793022) and KRAS: G12V. All genetically manipulated colon organoid lines were generated using the protocol described here[7]. shRNAs against *SMARCB1* were cloned into PLKO.1 vector. To generate lentiviruses, expression vectors were co-transfected into HEK293T cells with the lentiviral packaging constructs psPAX2 and pMD2.G (VSV-G) in a 1:1:1 ratio using X-tremeGENE9 DNA Transfection Reagent (Roche) according to the manufacturer's instructions. Cell culture media was changed the following day and lentiviral supernatant was harvested 48 h and 72 h later and filtered through a 0.45 µm filter (Millipore). Lentiviruses were aliquoted and stored at - 80 °C until use.

To transduce human CRC organoids, spheroids in one well (6-well plate) were trypsinized and used for each infection. Cells were resuspended in 500 µl lentiviral supernatant with 8 µg/mL polybrene and 10 mM Y-27632, centrifuged at 600g 37 °C 1 hours, and incubated for 6 hours in cell culture incubator. The infected cells were suspended in 30-50 ul of Matrigel and cultured with Wnt/R-spondin-deprived medium containing 10 mM Y-27632 and 10 mM SB431542. Colon organoids were

selected with 3 µg/ml puromycin at 24 hours post-infection. Xenograft experiments were performed following standard procedures involving athymic nude mice.

CRC cell line xenograft in vivo tumor model

Related animal procedures were conducted at Kineto Lab Ltd. (Budapest, Hungary) in compliance with national Hungarian legislation. Experiments were performed according to FELASA recommendations for animal use under license number PE/EA/401-7/2020. Female NOD-SCID mice approximately 8 weeks old (Charles River Laboratories) were used for the in vivo xenograft experiment. HT-29 human colon tumor cell line derivatives (HT-29 non-template control (NTC), HT-29 SMARCB1_shRNA1# (sh#1) and HT-29_SMARCB1_shRNA#4 (sh#4)) were cultured in RPMI-1640 medium with 10% fetal bovine serum (FBS) at 37º C with 5% CO2. Before injection cells were trypsinized and washed two times with FBS free medium. One and half million cells per injection was recovered in 50ul final volume and mixed the same volume of Matrigel. Hundred ul cell suspension were injected into each flanks subcutaneously on both side. Tumor volumes were obtained on every 3-4 days using handheld calipers to measure tumor length, width and height to calculate cubic millimeters; mice were weighed using a digital balance. Tumors were collected at the end of the study and placed in neutral buffered formalin (NBF) for fixation. Comparison of tumor volume over the course of treatment between either SMARCB1 knockdown or non-targeting control was conducted using two-way ANOVA.

Histopathology

Paraffin-embedded xenograft tumors were serially sectioned and mounted on microscopic glass slides. For morphological analysis, sections were serially dehydrated in xylene and ethanol, stained with H&E for histological assessment. Slides were digitized with a Pannoramic Midi slide scanner using a 20× objective (3D Histech, Budapest, Hungary).

Statistical analysis and reproducibility

Experiments were performed in triplicate. Data are represented as mean±s.d unless indicated otherwise. For each experiment, either independent biological or technical replicates were conducted as noted in the figure legends and were repeated with similar results. Statistical analysis was performed using Microsoft Office or Prism 8.0 (GraphPad) statistical tools. Pairwise comparisons between group means were performed using an unpaired two-tailed Student's *t*-test or Kruskal–Wallis test as appropriate unless otherwise indicated. Multi-group comparison of means was performed

using one-way ANOVA with post-hoc Tukey's HSD test as appropriate unless otherwise indicated. For all experiments, the variance between comparison groups was found to be equivalent.

Competing Interest Statement

S.S. and N.S.S. are co-inventors on an unpublished PCT patent application involving part of this work filed on June 8th, 2022. N.S.S. is on the scientific advisory board for Astrin Biosciences.

Authors contributions

S.S. designed the study, performed experiments, and revised the manuscript. D.C. performed computational analyses, developed methodology for CRISPR screening data, wrote/revised manuscript. P.L. designed and performed validation experiments and wrote/revised manuscript. P.D. performed experimental validation, wrote/revised manuscript. Z.L. analyzed perturb-seq experiment. P.B., L.V., T.V. performed in vivo experiments. P.S. and M.G. provided the human CRC organoid. B.W. provided funding and revised the manuscript. J.Q. provided resources and revised the manuscript. N.S.S. designed and supervised the study, obtained funding, analyzed data, assembled manuscript, and wrote/revised the manuscript.

Code and Data Availability

Code associated with the CRISPR screens using endogenous reporter system are available at https://github.com/davidchen0420/Endogenous_Reporter and raw data are available from the authors upon request.

Acknowledgements

We thank Jacob Stewart-Ornstein for generously sharing plasmids from which PCR-amplified HDR templates were utilized. We thank Pranshu Sahgal in addition to other Sethi lab members for thoughtful suggestions. Harvard Digestive Disease Center and NIH grant P30DK034854 for core services, resources, technology, and expertise. Dana-Farber/Harvard Cancer Center is supported in part by an NCI Cancer Center Support Grant # NIH 5 P30 CA06516. S.S. was supported by Bolyai (HAS) award and received funding from National Research Development and Innovation Office Hungary (FK142835). P.L. is a recipient of the Prince Mahidol Award Youth Program, Thailand. This work was funded by the Karen Grunebaum Award and Virtual Scholar Award from the Department of Defense (CA201084) to N.S.S and philanthropic support from the Jimmy Fund Walk (Opiela Family), Craig Baskin, and Howard & Wendy Cox to the Sethi Lab.

References

1 Jackson, A. L. *et al.* Widespread siRNA "off-target" transcript silencing mediated by seed region sequence complementarity. *Rna* **12**, 1179-1187, doi:10.1261/rna.25706 (2006).

2 Boettcher, M. & McManus, M. T. Choosing the Right Tool for the Job: RNAi, TALEN, or CRISPR. *Mol Cell* **58**, 575-585, doi:10.1016/j.molcel.2015.04.028 (2015).

3 Mohr, S. E., Smith, J. A., Shamu, C. E., Neumuller, R. A. & Perrimon, N. RNAi screening comes of age: improved techniques and complementary approaches. *Nat Rev Mol Cell Biol* **15**, 591-600, doi:10.1038/nrm3860 (2014).

4 Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**, 80-84, doi:10.1126/science.1246981 (2014).

5 Takeda, D. Y. *et al.* A Somatically Acquired Enhancer of the Androgen Receptor Is a Noncoding Driver in Advanced Prostate Cancer. *Cell* **174**, 422-432 e413, doi:10.1016/j.cell.2018.05.037 (2018).

6 Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603-607, doi:10.1038/nature11003 (2012).

7 Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84-87, doi:10.1126/science.1247005 (2014).

8 Chen, S. *et al.* Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. *Cell* **160**, 1246-1260, doi:10.1016/j.cell.2015.02.038 (2015).

9 Tsherniak, A. *et al.* Defining a Cancer Dependency Map. *Cell* **170**, 564-576 e516, doi:10.1016/j.cell.2017.06.010 (2017).

10 Behan, F. M. *et al.* Prioritization of cancer therapeutic targets using CRISPR–Cas9 screens. *Nature* **568**, 511-516, doi:10.1038/s41586-019-1103-9 (2019).

11 Cowley, G. S. *et al.* Parallel genome-scale loss of function screens in 216 cancer cell lines for the identification of context-specific genetic dependencies. *Sci Data* **1**, 140035, doi:10.1038/sdata.2014.35 (2014).

12 Olivieri, M. *et al.* A Genetic Map of the Response to DNA Damage in Human Cells. *Cell* **182**, 481-496 e421, doi:10.1016/j.cell.2020.05.040 (2020).

13 Yu, C. *et al.* Genome-wide CRISPR-cas9 knockout screening identifies GRB7 as a driver for MEK inhibitor resistance in KRAS mutant colon cancer. *Oncogene* **41**, 191-203, doi:10.1038/s41388-021-02077-w (2022).

14 Ford, K., McDonald, D. & Mali, P. Functional Genomics via CRISPR-Cas. *J Mol Biol* **431**, 48-65, doi:10.1016/j.jmb.2018.06.034 (2019).

15 Bock, C. *et al.* High-content CRISPR screening. *Nature Reviews Methods Primers* **2**, 8, doi:10.1038/s43586-021-00093-4 (2022).

16 Cortez, J. T. *et al.* CRISPR screen in regulatory T cells reveals modulators of Foxp3. *Nature* **582**, 416-420, doi:10.1038/s41586-020-2246-4 (2020).

17 Fearon, E. R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **61**, 759-767, doi:10.1016/0092-8674(90)90186-i (1990).

18 Kinzler, K. W. & Vogelstein, B. Lessons from hereditary colorectal cancer. *Cell* **87**, 159-170, doi:10.1016/s0092-8674(00)81333-1 (1996).

19 Morin, P. J. *et al.* Activation of beta-catenin-Tcf signaling in colon cancer by mutations in beta-catenin or APC. *Science* **275**, 1787-1790, doi:10.1126/science.275.5307.1787 (1997).

20 Tang, W. *et al.* A genome-wide RNAi screen for Wnt/beta-catenin pathway components identifies unexpected roles for TCF transcription factors in cancer. *Proc Natl Acad Sci U S A* **105**, 9697-9702, doi:10.1073/pnas.0804709105 (2008).

21 Firestein, R. *et al.* CDK8 is a colorectal cancer oncogene that regulates β-catenin activity. *Nature* **455**, 547-551, doi:10.1038/nature07179 (2008).

22 Wan, C. *et al.* Genome-scale CRISPR-Cas9 screen of Wnt/β-catenin signaling identifies therapeutic targets for colorectal cancer. *Sci Adv* **7**, doi:10.1126/sciadv.abf2567 (2021).

23 Kaelin, W. G., Jr. Common pitfalls in preclinical cancer target validation. *Nat Rev Cancer* **17**, 425-440, doi:10.1038/nrc.2017.32 (2017).

24 Hanahan, D. Hallmarks of Cancer: New Dimensions. *Cancer Discov* **12**, 31-46, doi:10.1158/2159-8290.CD-21-1059 (2022).

25 de Sousa, E. M. F. & de Sauvage, F. J. Cellular Plasticity in Intestinal Homeostasis and Disease. *Cell Stem Cell* **24**, 54-64, doi:10.1016/j.stem.2018.11.019 (2019).

26 de The, H. Differentiation therapy revisited. *Nat Rev Cancer* **18**, 117-127, doi:10.1038/nrc.2017.103 (2018).

27 Liang, X. *et al.* An Enhancer-Driven Stem Cell-Like Program Mediated by SOX9 Blocks Intestinal Differentiation in Colorectal Cancer. *Gastroenterology* **162**, 209-222, doi:10.1053/j.gastro.2021.09.044 (2022).

28 Calnek, D. & Quaroni, A. Differential localization by in situ hybridization of distinct keratin mRNA species during intestinal epithelial cell development and differentiation. *Differentiation* **53**, 95-104, doi:10.1111/j.1432-0436.1993.tb00649.x (1993).

29 Dow, L. E. *et al.* Apc Restoration Promotes Cellular Differentiation and Reestablishes Crypt Homeostasis in Colorectal Cancer. *Cell* **161**, 1539-1552, doi:10.1016/j.cell.2015.05.033 (2015).

30 Clevers, H. Wnt/beta-catenin signaling in development and disease. *Cell* **127**, 469-480, doi:10.1016/j.cell.2006.10.018 (2006).

31 Shimokawa, M. *et al.* Visualization and targeting of LGR5(+) human colon cancer stem cells. *Nature* **545**, 187-192, doi:10.1038/nature22081 (2017).

32 Stewart-Ornstein, J. & Lahav, G. Dynamics of CDKN1A in Single Cells Defined by an Endogenous Fluorescent Tagging Toolkit. *Cell Reports* **14**, 1800-1811, doi:https://doi.org/10.1016/j.celrep.2016.01.045 (2016).

33 Bala, P. *et al.* Aberrant cell state plasticity mediated by developmental reprogramming precedes colorectal cancer initiation. *Sci Adv* **9**, eadf0927, doi:10.1126/sciadv.adf0927 (2023).

34 Stewart-Ornstein, J. & Lahav, G. Dynamics of CDKN1A in Single Cells Defined by an Endogenous Fluorescent Tagging Toolkit. *Cell Rep* **14**, 1800-1811, doi:10.1016/j.celrep.2016.01.045 (2016).

35 Ng, J. M. & Yu, J. Promoter hypermethylation of tumour suppressor genes as potential biomarkers in colorectal cancer. *Int J Mol Sci* **16**, 2472-2496, doi:10.3390/ijms16022472 (2015).

36 Weisenberger, D. J. *et al.* CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nat Genet* **38**, 787-793, doi:10.1038/ng1834 (2006).

37 Audia, J. E. & Campbell, R. M. Histone Modifications and Cancer. *Cold Spring Harb Perspect Biol* **8**, a019521, doi:10.1101/cshperspect.a019521 (2016).

38 Dixit, A. *et al.* Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell* **167**, 1853-1866 e1817, doi:10.1016/j.cell.2016.11.038 (2016).

39 Reynolds, S., Rajagopal, S. & Chakrabarty, S. Differentiation-inducing effect of retinoic acid, difluoromethylornithine, sodium butyrate and sodium suramin in human colon cancer cells. *Cancer Lett* **134**, 53-60, doi:10.1016/s0304-3835(98)00242-0 (1998).

40 Wester, R. A. *et al.* Retinoic acid signaling drives differentiation toward the absorptive lineage in colorectal cancer. *iScience* **24**, 103444, doi:10.1016/j.isci.2021.103444 (2021).

41 Qian, Y., Xia, S. & Feng, Z. Sox9 mediated transcriptional activation of FOXK2 is critical for colorectal cancer cells proliferation. *Biochem Biophys Res Commun* **483**, 475-481, doi:10.1016/j.bbrc.2016.12.119 (2017).

42 Mertens, S. *et al.* Drug-repurposing screen on patient-derived organoids identifies therapy-induced vulnerability in KRAS-mutant colon cancer. *Cell Rep* **42**, 112324, doi:10.1016/j.celrep.2023.112324 (2023).

43 He, X. *et al.* Patient-derived organoids as a platform for drug screening in metastatic colorectal cancer. *Front Bioeng Biotechnol* **11**, 1190637, doi:10.3389/fbioe.2023.1190637 (2023).

44 Agarwal, N. *et al.* TRIM28 is a transcriptional activator of the mutant TERT promoter in human bladder cancer. *Proc Natl Acad Sci U S A* **118**, doi:10.1073/pnas.2102423118 (2021).

45 Blanco, M. A. *et al.* Chromatin-state barriers enforce an irreversible mammalian cell fate decision. *Cell Rep* **37**, 109967, doi:10.1016/j.celrep.2021.109967 (2021).

46 Sykes, D. B. *et al.* Inhibition of Dihydroorotate Dehydrogenase Overcomes Differentiation Blockade in Acute Myeloid Leukemia. *Cell* **167**, 171-186 e115, doi:10.1016/j.cell.2016.08.057 (2016).

47 Wang, E. *et al.* Surface antigen-guided CRISPR screens identify regulators of myeloid leukemia differentiation. *Cell Stem Cell* **28**, 718-731.e716, doi:https://doi.org/10.1016/j.stem.2020.12.005 (2021).

48 Yan, F. *et al.* KAT6A and ENL Form an Epigenetic Transcriptional Control Module to Drive Critical Leukemogenic Gene-Expression Programs. *Cancer Discov* **12**, 792-811, doi:10.1158/2159-8290.CD-20-1459 (2022).

49 Barker, N. & Clevers, H. Mining the Wnt pathway for cancer therapeutics. *Nat Rev Drug Discov* **5**, 997-1014, doi:10.1038/nrd2154 (2006).

50 Cancer Genome Atlas, N. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330-337, doi:10.1038/nature11252 (2012).

51 Cheng, X., Xu, X., Chen, D., Zhao, F. & Wang, W. Therapeutic potential of targeting the Wnt/beta-catenin signaling pathway in colorectal cancer. *Biomed Pharmacother* **110**, 473-481, doi:10.1016/j.biopha.2018.11.082 (2019).

52 Firestein, R. *et al.* CDK8 is a colorectal cancer oncogene that regulates beta-catenin activity. *Nature* **455**, 547-551, doi:10.1038/nature07179 (2008).

53 Major, M. B. *et al.* New regulators of Wnt/beta-catenin signaling revealed by integrative molecular screening. *Sci Signal* **1**, ra12, doi:10.1126/scisignal.2000037 (2008).

54 Lebensohn, A. M. *et al.* Comparative genetic screens in human cells reveal new regulatory mechanisms in WNT signaling. *Elife* **5**, doi:10.7554/eLife.21459 (2016).

55 Schmidt, R. *et al.* CRISPR activation and interference screens decode stimulation responses in primary human T cells. *Science* **375**, eabj4008, doi:10.1126/science.abj4008 (2022).

56 Wiersma, M. *et al.* Protein kinase Msk1 physically and functionally interacts with the KMT2A/MLL1 methyltransferase complex and contributes to the regulation of multiple target genes. *Epigenetics Chromatin* **9**, 52, doi:10.1186/s13072-016-0103-3 (2016).

57 Huang, Y. C. *et al.* The epigenetic factor Kmt2a/Mll1 regulates neural progenitor proliferation and neuronal and glial differentiation. *Dev Neurobiol* **75**, 452-462, doi:10.1002/dneu.22235 (2015).

58 Meyer, C. *et al.* The MLL recombinome of acute leukemias in 2017. *Leukemia* **32**, 273-284, doi:10.1038/leu.2017.213 (2018).

59 Grinat, J. *et al.* The epigenetic regulator Mll1 is required for Wnt-driven intestinal tumorigenesis and cancer stemness. *Nat Commun* **11**, 6422, doi:10.1038/s41467-020-20222-z (2020).

60 Mashtalir, N. *et al.* A Structural Model of the Endogenous Human BAF Complex Informs Disease Mechanisms. *Cell* **183**, 802-817.e824, doi:10.1016/j.cell.2020.09.051 (2020).

61 Tang, Z., Kang, B., Li, C., Chen, T. & Zhang, Z. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res* **47**, W556-w560, doi:10.1093/nar/gkz430 (2019).

62 Yao, B. *et al.* PRMT1-mediated H4R3me2a recruits SMARCA4 to promote colorectal cancer progression by enhancing EGFR signaling. *Genome Med* **13**, 58, doi:10.1186/s13073-021-00871-5 (2021).

63 Rago, F. *et al.* Exquisite Sensitivity to Dual BRG1/BRM ATPase Inhibitors Reveals Broad SWI/SNF Dependencies in Acute Myeloid Leukemia. *Molecular Cancer Research* **20**, 361-372, doi:10.1158/1541-7786.Mcr-21-0390 (2022).

64 Lee, D. *et al.* The Bromodomain Inhibitor PFI-3 Sensitizes Cancer Cells to DNA Damage by Targeting SWI/SNF. *Mol Cancer Res* **19**, 900-912, doi:10.1158/1541-7786.Mcr-20-0289 (2021).

65 Xiao, L. *et al.* Targeting SWI/SNF ATPases in enhancer-addicted prostate cancer. *Nature* **601**, 434-439, doi:10.1038/s41586-021-04246-z (2022).

66 Wang, X., Haswell, J. R. & Roberts, C. W. Molecular pathways: SWI/SNF (BAF) complexes are frequently mutated in cancer--mechanisms and potential therapeutic insights. *Clin Cancer Res* **20**, 21-27, doi:10.1158/1078-0432.CCR-13-0280 (2014).

67 Wilson, B. G. *et al.* Epigenetic antagonism between polycomb and SWI/SNF complexes during oncogenic transformation. *Cancer Cell* **18**, 316-328, doi:10.1016/j.ccr.2010.09.006 (2010).

68 Roberts, C. W., Galusha, S. A., McMenamin, M. E., Fletcher, C. D. & Orkin, S. H. Haploinsufficiency of Snf5 (integrase interactor 1) predisposes to malignant rhabdoid tumors in mice. *Proc Natl Acad Sci U S A* **97**, 13796-13800, doi:10.1073/pnas.250492697 (2000).

69 Wang, X. *et al.* SMARCB1-mediated SWI/SNF complex function is essential for enhancer regulation. *Nat Genet* **49**, 289-295, doi:10.1038/ng.3746 (2017).

70 Alver, B. H. *et al.* The SWI/SNF chromatin remodelling complex is required for maintenance of lineage specific enhancers. *Nat Commun* **8**, 14648, doi:10.1038/ncomms14648 (2017).

71 Parisian, A. D. *et al.* SMARCB1 loss interacts with neuronal differentiation state to block maturation and impact cell stability. *Genes Dev* **34**, 1316-1329, doi:10.1101/gad.339978.120 (2020).

72 Sekiguchi, F. *et al.* Genetic abnormalities in a large cohort of Coffin-Siris syndrome patients. *J Hum Genet* **64**, 1173-1186, doi:10.1038/s10038-019-0667-4 (2019).

73 Zhu, Z. *et al.* Mitotic bookmarking by SWI/SNF subunits. *Nature*, doi:10.1038/s41586-023-06085-6 (2023).

74 Mora-Blanco, E. L. *et al.* Activation of beta-catenin/TCF targets following loss of the tumor suppressor SNF5. *Oncogene* **33**, 933-938, doi:10.1038/onc.2013.37 (2014).

75 Knutson, S. K. *et al.* Durable tumor regression in genetically altered malignant rhabdoid tumors by inhibition of methyltransferase EZH2. *Proc Natl Acad Sci U S A* **110**, 7922-7927, doi:10.1073/pnas.1303800110 (2013).

76 Labun, K. *et al.* CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Res* **47**, W171-W174, doi:10.1093/nar/gkz365 (2019).

77 Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587 e3529, doi:10.1016/j.cell.2021.04.048 (2021).

78 Papalexi, E. *et al.* Characterizing the molecular regulation of inhibitory immune checkpoints with multimodal single-cell screens. *Nat Genet* **53**, 322-331, doi:10.1038/s41588-021-00778-2 (2021).

79 Haber, A. L. *et al.* A single-cell survey of the small intestinal epithelium. *Nature* **551**, 333-339, doi:10.1038/nature24489 (2017).

MAIN FIGURES

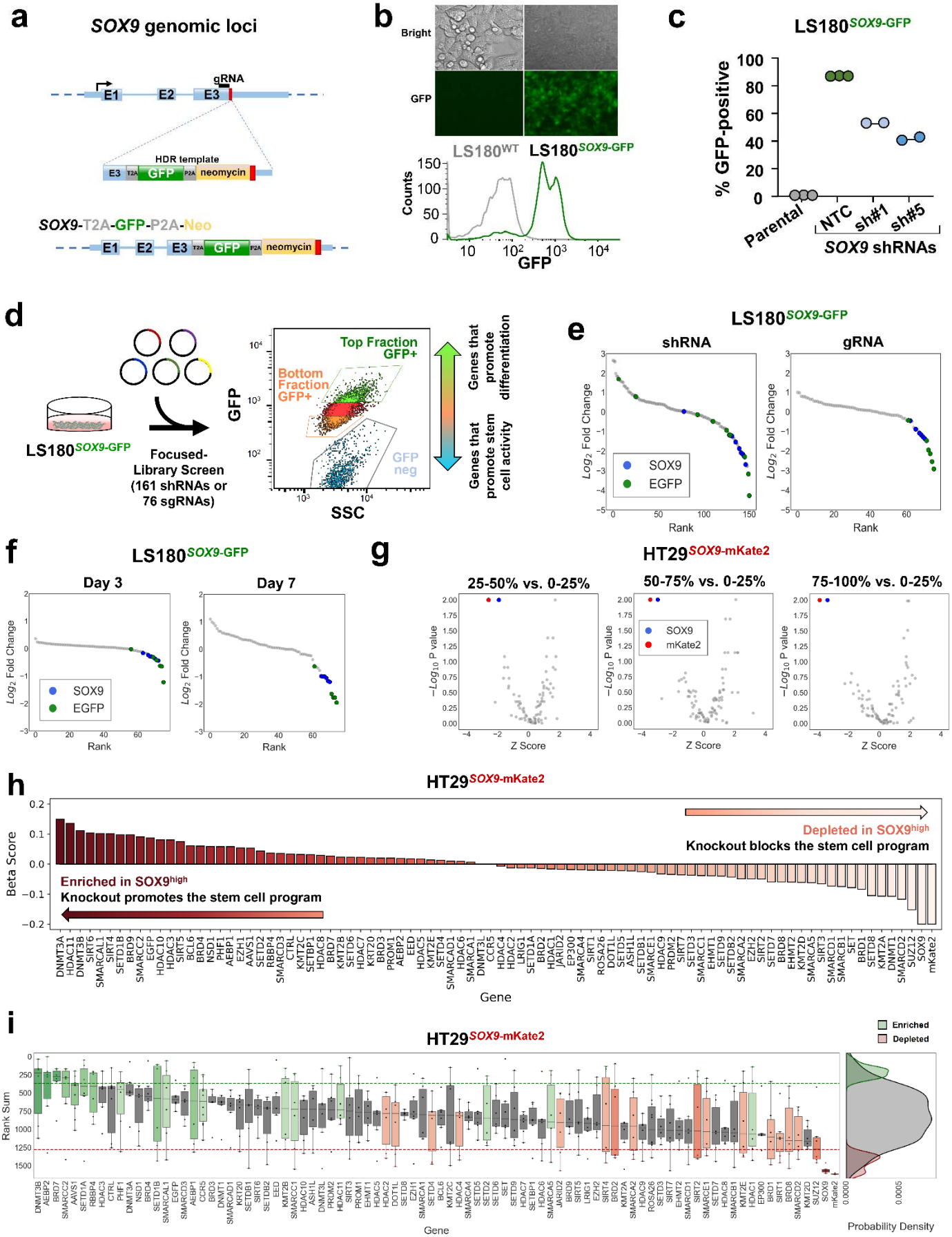

Figure 1. Development of an endogenous reporter by genome editing *SOX9* locus

a. Schematic of *SOX9* genomic locus, sgRNA location, homolog arm structure; E1-E3 represent *SOX9* exons (top panel). Post-integration genomic locus of the T2A-GFP-P2A-Neo reporter cassette at the *SOX9* locus (bottom panel).
b. Distribution of GFP fluorescence intensity of the LS180$^{\text{SOX9-GFP}}$ cells and LS180 parental cell line.
c. Percentage of GFP-positive cells in the cell line population of parental LS180 cell line and LS180$^{\textit{SOX9}\text{-GFP}}$ engineered line upon introduction of nontargeting control (NTC) and SOX9 shRNAs.
d. Design of the focused CRISPR-Cas9 screen of 76 sgRNAs and short hairpin RNA genetic screen of 161 shRNAs targeting 8 genes in the LS180$^{\text{SOX9-GFP}}$ endogenous differentiation reporter cell line.
e. Ranked $\log_2$ fold change plot of focused RNA interference screen and CRISPR-Cas9 screen comparing the top 2.5% to bottom 2.5% GFP positive sorted cell fractions of the LS180$^{\textit{SOX9}\text{-GFP}}$ genome-edited cell line. SOX9 shRNAs and sgRNAs are represented by blue dots. EGFP shRNAs and sgRNAs are represented by green dots.
f. Ranked $\log_2$ fold change plot of control CRISPR-Cas9 screen (76 sgRNAs) comparing between the top 2.5% GFP positive and the bottom 2.5% GFP positive sorted cell fractions of the LS180$^{\textit{SOX9}\text{-GFP}}$ genome-edited cell line sequenced after at day 3 and day 7 following library infection.
g. Volcano plot of the epigenetic CRISPR-Cas9 screen (542 sgRNAs targeting 78 genes) comparing 4 sorted cell fractions of the HT29$^{\textit{SOX9-mKate2}}$ genome-edited cell line sampled at day 7. Following library infection, HT29$^{\textit{SOX9}\text{-mKate2}}$ were assigned into 4 sorted fractions based on mKate2 fluorescence intensity. The *x* axis shows the *Z* score of gene-level FC (median $\log_2$fold change) for all sgRNAs targeting the same gene. The y axis shows the gene-level p values generated by MaGeCK MLE.
h. Beta score of each gene in the epigenetic CRISPR-Cas9 screen comparing between the top 25% mKate2 positive and the bottom 25% mKate2 positive sorted cell fractions of the HT29$^{\textit{SOX9}\text{-mKate2}}$ cell line from 3 technical replicates
i. Rank sum of each gene in the CRISPR-Cas9 screen targeting epigenetic regulators (78 genes) comparing the top 25% and the bottom 25% of mKate2 positive sorted cell fractions of the HT29$^{\textit{SOX9}\text{-mKate2}}$ cell line. The y-axis shows the rank sum of each sgRNA targeting each of the 78 genes in the epigenetic library shown on the x-axis (left panel). The rank sum is derived by summing the ordered rank of each sgRNA by fold change across the 3 technical replicates of the screen such that higher rank sum is associated with depletion and lower rank sum is associated with enrichment. Boxplots showing the distribution of sgRNA rank sums per gene are colored in green or red if there are at least 2 sgRNA targeting the same gene whose rank sums are above the enrichment threshold (green) or depletion threshold (red) respectively. Probability density plots showing the distribution of enriched (green), depleted (red), and all other sgRNA (grey) rank sums (right panel).

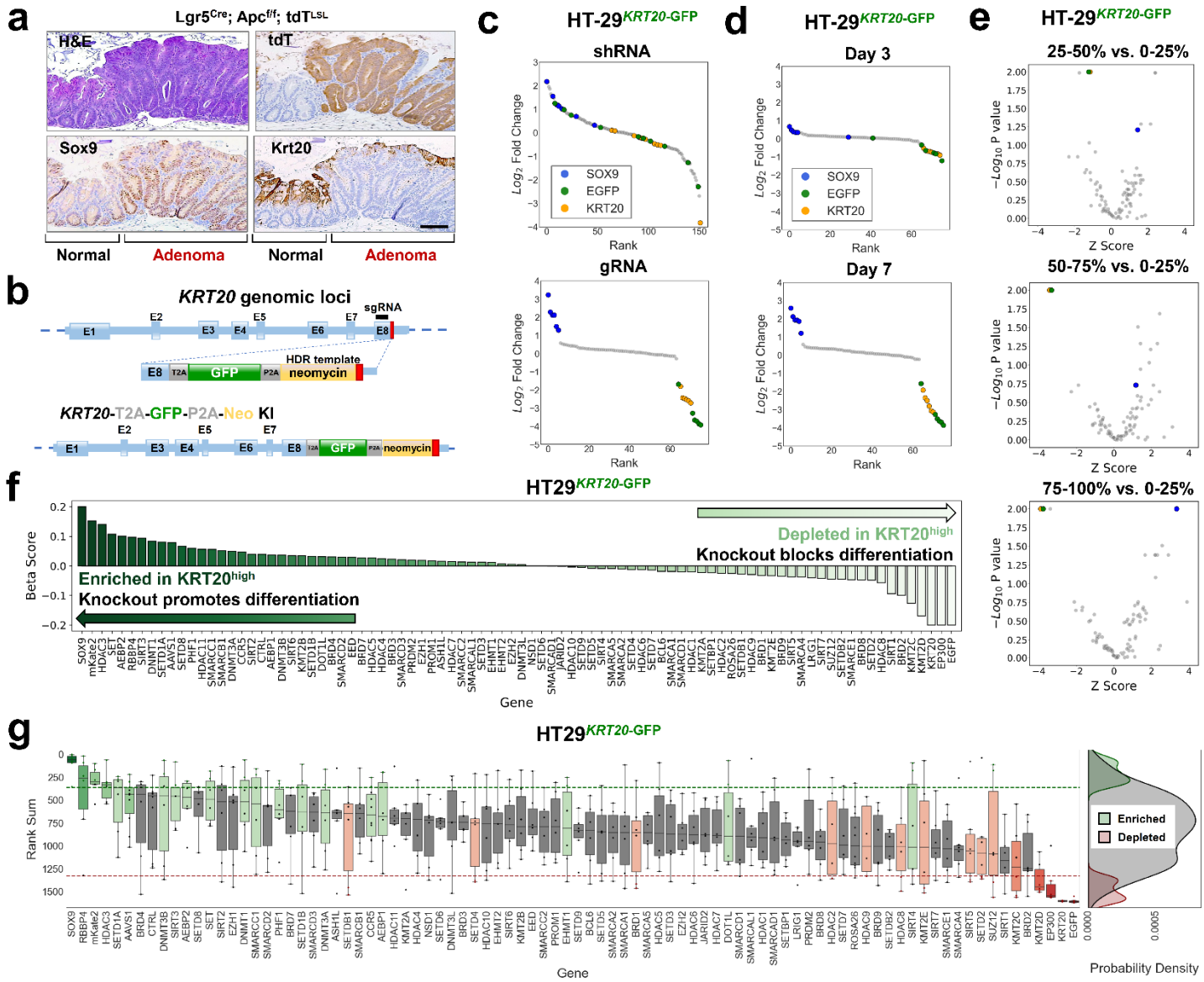

Figure 2. Endogenous differentiation reporter by genome editing *KRT20* locus

a. SOX9 and KRT20 immunohistochemistry in normal and Apc^{KO} mouse colon; scale bar = 100µM

b. Genomic locus of *KRT20* in which E1 through E8 represent exons; diagram of HDR template (cassette) containing the GFP fluorescent reporter and neomycin antibiotic selection marker (yellow); modified genomic *KRT20* locus after successful integration of the GFP fluorescent reporter cassette (bottom panel).

c. Ranked $\log_2$ fold change plot of a focused RNA interference screen (161 shRNAs) and CRISPR-Cas9 screen (76 sgRNAs) comparing between the top 2.5% GFP positive and the bottom 2.5% GFP positive sorted cell fractions of the LS180$^{KRT20\text{-}GFP}$ genome-edited cell line. KRT20 shRNAs and sgRNAs are represented by yellow dots. EGFP shRNAs and sgRNAs are represented by green dots. SOX9 shRNAs and sgRNAs are represented by blue dots.

d. Ranked $\log_2$ fold change plot of focused CRISPR-Cas9 screen (76 sgRNAs) comparing between the top 2.5% and the bottom 2.5% GFP positive sorted cell fractions of the LS180$^{KRT20\text{-}EGFP}$ cell line sampled at day 3 and day 7 post-library infection.

e. Volcano plot of the epigenetic CRISPR-Cas9 screen (542 sgRNAs targeting 78 genes) comparing between 4 sorted fractions of the HT29$^{KRT20\text{-}EGFP}$ genome-edited cell line sampled at day 7. Following library infection, HT29$^{KRT20\text{-}EGFP}$ were assigned into 4 sorted fractions based on GFP

fluorescence intensity. The *x* axis shows the *Z* score of gene-level FC (mean $\log_2$(fold change) for all sgRNAs targeting the same gene). The y axis shows the gene-level p values generated by MaGeCK Maximum Likelihood Estimation (MLE).

f. Beta score of each gene in the epigenetic CRISPR-Cas9 screen (78 genes) comparing between the top 25% GFP positive and the bottom 25% GFP positive sorted cell fractions of the HT29$^{KRT20\text{-}EGFP}$ cell line from 3 technical replicates.

g. Rank sum of each gene in CRISPR-Cas9 screen targeting epigenetic regulators (78 genes) comparing the top 25% and the bottom 25% GFP positive sorted cell fractions of the HT29$^{KRT20\text{-}EGFP}$ cell line. The y-axis shows the rank sum of each sgRNA targeting each of the 78 genes in the epigenetic CRISPR-Cas9 screen shown on the x-axis. Boxplots showing the distribution of sgRNA rank sums per gene are colored in green or red if there are at least 2 sgRNA targeting the same gene whose rank sums are above the top 15% of all sgRNA rank sum scores(green) or below the bottom 15 of all sgRNA rank sum scores (red) respectively. Probability density plots showing the distribution of enriched (green), depleted (red), and all other sgRNA (grey) rank sums (right panel).

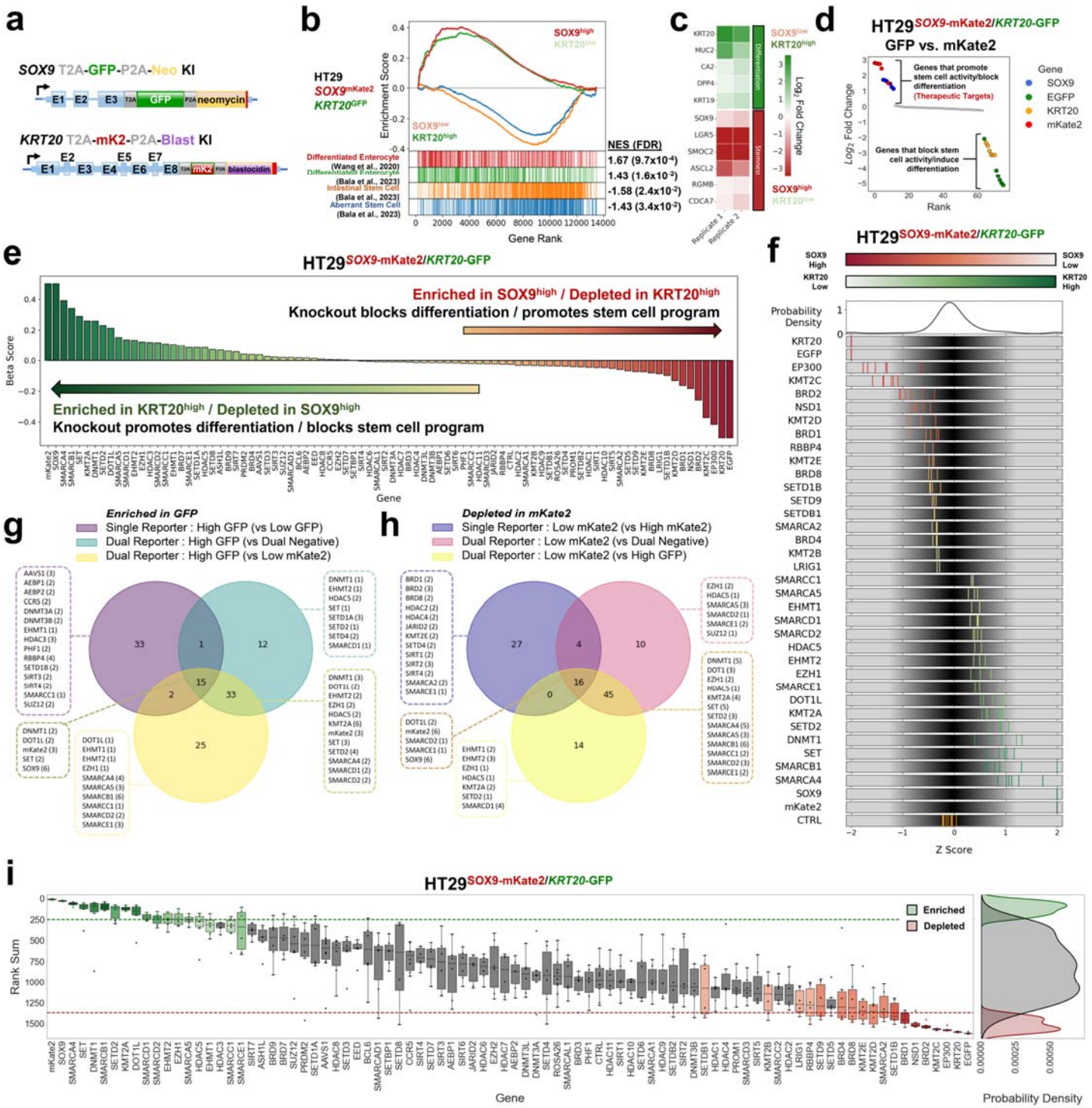

Figure 3. Epigenetic CRISPR screen using dual endogenous reporter system.

a. Schematic of engineered *SOX9* and *KRT20* loci in dual endogenous reporter lines following GFP and mKate2 HDR template integration, respectively.

b. Gene set enrichment analysis (GSEA) of bulk RNA-seq of mKate2^{low}/GFP^{high} fraction and mKate2^{high}/GFP^{low} FACS sorted fractions from HT-29$^{SOX9\text{-}mKate2/KRT20\text{-}GFP}$ dual reporter line. Two differentiation signatures are indicated in red and green, followed by a normal stem cell signature in orange, and an aberrant stem cell signature in blue. Normalized enrichment scores (NES) and false discovery rates (FDR) for each signature is listed to the right.

c. Heatmap showing select differentiation and stem cell genes from bulk RNA-seq of mKate2^{low}/GFP^{high} fraction and mKate2^{high}/GFP^{low} fractions from HT-29$^{SOX9\text{-}mKate2/KRT20\text{-}GFP}$ dual reporter line
d. Ranked $\log_2$ fold change plot of focused CRISPR-Cas9 screen (76 sgRNAs) comparing GFP^{high} and mKate2^{low} high sorted cell fractions using HT-29$^{SOX9\text{-}mKate2/KRT20\text{-}GFP}$ cells
e. Beta score of each gene in the epigenetic CRISPR-Cas9 screen (78 genes) comparing mKate2^{low}/GFP^{high} fraction and mKate2^{high}/GFP^{low} using HT-29$^{SOX9\text{-}mKate2/KRT20\text{-}GFP}$ cell line from 3 technical replicates.
f. Distribution of sgRNA Z-score of $\log_2$ fold change of the top and bottom hits from the epigenetic CRISPR-Cas9 screen. sgRNA are colored based on their enrichment (green) or depletion (red) in the mKate2^{low}/GFP^{high} fraction compared to the mKate2^{high}/GFP^{low} using the dual reporter system.
g. Overlap analysis of the number of shared and exclusive hits enriched in the 75-100% GFP fraction compared to 0-25% GFP fraction from the single differentiation program reporter system as well as enriched in mKate2^{low}/GFP^{high} compared to the mKate2^{high}/GFP^{low} fraction or enriched in mKate2^{low}/GFP^{high} compared to the mKate2^{low}/GFP^{low} fraction from the dual reporter system. Top-scoring hits were identified using the rank sum method applied to one replicate of each screen.
h. Overlap analysis of the number of shared and exclusive hits depleted in the 75-100% mKate2 fraction compared to 0-25% mKate2 fraction from the single stem cell program reporter system as well as depleted in mKate2^{high}/GFP^{low} compared to the mKate2^{low}/GFP^{low} fraction or depleted in mKate2^{high}/GFP^{low} compared to the mKate2^{low}/GFP^{high} fraction from the dual reporter system. Top-scoring hits were identified using the rank sum method applied to one replicate of each screen.
i. Rank sum score of the 542 sgRNA targeting 78 genes in the CRISPR-Cas9 screen targeting epigenetic regulators comparing the mKate2^{low}/GFP^{high} to the mKate2^{high}/GFP^{low} fractions from the HT-29$^{\text{SOX9-mKate2/}KRT20\text{-GFP}}$ dual reporter cell line. The y-axis shows the rank sum of each sgRNA targeting each of the 78 genes in the epigenetic library shown on the x-axis. Boxplots showing the distribution of sgRNA rank sums per gene are colored in green or red if there are at least 2 sgRNA targeting the same gene whose rank sums are above the top 15% of all sgRNA rank sum scores(green) or below the bottom 15 of all sgRNA rank sum scores (red) respectively. Probability density plots showing the distribution of enriched hits(green), depleted hits (red), and all other sgRNA (grey) rank sum scores (right panel).

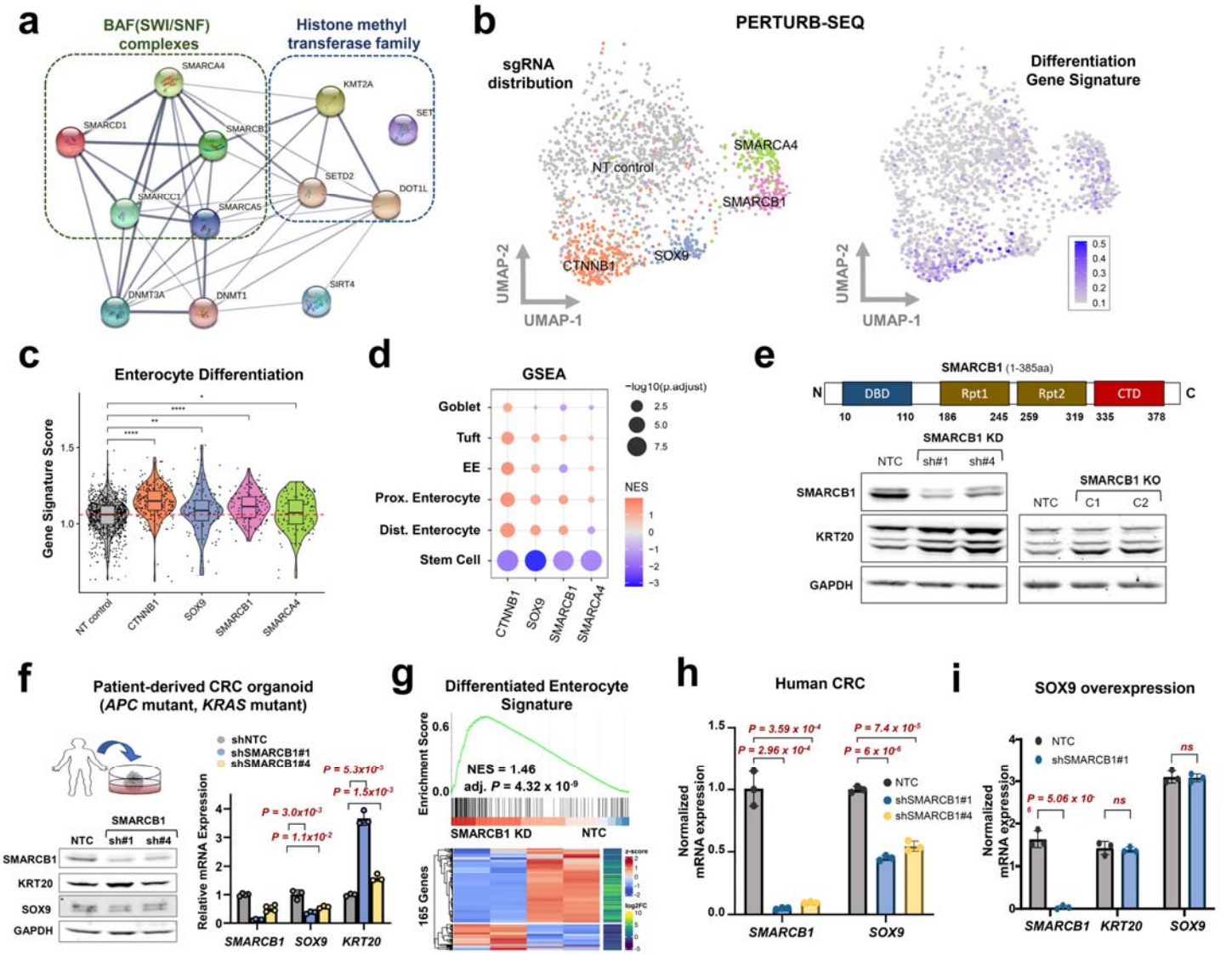

Figure 4. SMARCB1 restricts differentiation in models of CRC

a. The gene and protein regulatory network among the genetic perturbations that promoted differentiation and impeded aberrant stem-cell activity in the dual reporter using STRING.

b. UMAP representation of CRISPR screening of select genes coupled with scRNA-seq readout (Perturb-seq). sgRNA assignment of each cell (left). Representation of intestinal differentiation gene signature (right) [79].

c. Violin plot depicting the enterocyte gene expression signature in cells with NT control, CTNNB1, SOX9, SMARCB1, and SMARCA4 sgRNAs.

d. GSEA of single cell transcriptomes from cells with CTNNB1, SOX9, SMARCB1, and SMARCA4 sgRNAs relative to NT controls using distinct differentiation and one stem cell signature (Haber et al., 2017).

e. Immunoblot of SMARCB1, KRT20, and GAPDH in HT-29 SMARCB1 KD cell lines using 2 shRNAs and HT-29 SMARCB1 KO clones using CRISPR-Cas9; C1 has 32.9% and C2 has 72.6% editing.

f. Immunoblot of SMARCB1 and KRT20 (left) and mRNA expression (right) of *SMARCB1*, *SOX9*, and *KRT20* in the patient-derived CRC organoid transduced with 2 shRNAs against SMARCB1.

g. GSEA (top) and heatmap (bottom) of differentiation gene expression signature in bulk RNA-seq data from the patient-derived CRC organoid transduced with either NT Control or SMARCB1 KD shRNA.

h. mRNA expression of *SMARCB1* and *SOX9* in HT-29 SMARCB1 KD cell lines using 2 shRNAs.
i. mRNA expression of *SMARCB1, KRT20, SOX9* in HT-115 control and SMARCB1 KD lines engineered to conditionally overexpress SOX9.

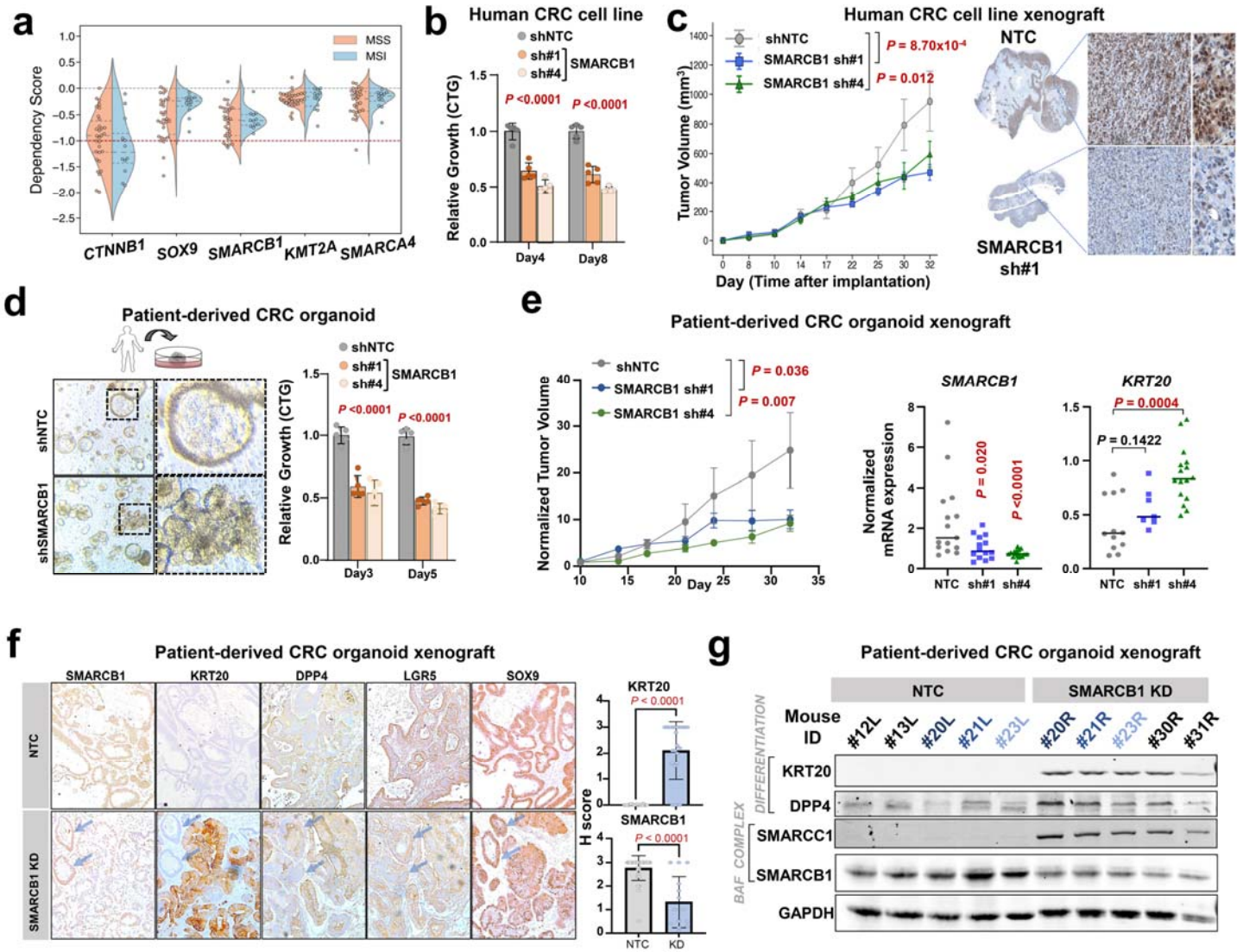

Figure 5. SMARCB1 is a CRC dependency and required for *in vivo* tumor growth

a. Distribution of dependency scores across microsatellite stable and microsatellite instable CRC cell lines in the CCLE for positive control gene dependencies in colorectal cancer (CTNNB1, SOX9) and novel candidate dependencies in colorectal cancer (SMARCB1, KMT2A, SMARCA4). Red line at a score –1.0 indicates the threshold below which genes are considered essential within the cell line.

b. Relative growth of the HT29 CRC cell line upon SMARCB1 knockdown with 2 shRNAs is reduced compared to the negative control non-targeting knockdown.

c. Tumor volume of the HT29 CRC cell line xenograft in mice models upon either SMARCB1 sh#1 knockdown (n=6) and SMARCB1 sh#4 knockdown (n=6) was reduced compared to the negative control non-targeting knockdown (n=6). Tumor volume data shown as mean□+□SEM. SMARCB1 IHC histopathology of xenografts at experiment endpoint (right) .

d. Morphology and relative growth of patient-derived CRC organoid upon SMARCB1 knockdown compared to the negative control non-targeting knockdown.

e. Relative growth of the patient-derived CRC xenograft in mice models upon either SMARCB1 sh#1 knockdown (n=6) and SMARCB1 sh#4 knockdown (n=8) was reduced compared to the negative control non-targeting knockdown (n=7). Relative growth data shown as mean□+□SEM. Normalized mRNA expression of *SMARCB1* and *KRT20* in patient-derived CRC xenografts at experimental endpoint.

f. Immunohistochemistry analysis of SMARCB1, KRT20, DPP4, LGR5, and SOX9 proteins in the patient-derived CRC organoid *in vivo* xenografts expressing SMARCB1 shRNAs or non-targeting controls. Blue arrows indicate two crypts that escaped SMARCB1 KD. Quantification of SMARCB1 and KRT20 shown in the right panel.
g. Immunoblot of differentiation proteins (KRT20, DPP4) and BAF complex proteins (SMARCB1, SMARCC1) in patient-derived CRC organoid xenografts.