Epigenetic dysregulation of the H3K14ac/H3K9me2 axis is associated with malignant transformation and prognosis in colon adenocarcinoma

Annalisa Roberti
Institute of Oncology of Asturias (IUOPA)
Juan Ramon Tejedor
Nanomaterials and Nanotechnology Research Centre (CINN-CSIC)
Claudia Martínez-Sánchez
Health Research Institute of Asturias (ISPA)
Virginia López
Nanomaterials and Nanotechnology Research Centre (CINN-CSIC)
Rocío G. Urdinguio
Nanomaterials and Nanotechnology Research Centre (CINN-CSIC)
Diego Quintana-Torres
Nanomaterials and Nanotechnology Research Centre (CINN-CSIC)
Jennifer M. Kefauver
Nanomaterials and Nanotechnology Research Centre (CINN-CSIC)
Darío Bagüés-Castro
Nanomaterials and Nanotechnology Research Centre (CINN-CSIC)
Cristina Mangas
Institute of Oncology of Asturias (IUOPA)
Ignacio Ortea
Nanomaterials and Nanotechnology Research Centre (CINN-CSIC)
Sergio Ciordia
Spanish National Centre for Biotechnology (CNB-CSIC)
Fernando Corrales
Spanish National Centre for Biotechnology (CNB-CSIC)
Luis Valledor
University of Oviedo
Mónica Meijón
University of Oviedo
María Jesús Cañal
University of Oviedo
Tamara Díaz-Vico

Central University Hospital of Asturias (HUCA)

Aida Suarez-Sánchez
Central University Hospital of Asturias (HUCA)

Oscar González-Bernardo
Central University Hospital of Asturias (HUCA)

Daniel Fernández-Martínez
Central University Hospital of Asturias (HUCA)

Luis Joaquín García-Flórez
Central University Hospital of Asturias (HUCA)

Agustín F. Fernández
Nanomaterials and Nanotechnology Research Centre (CINN-CSIC)

Mario F. Fraga
mffraga@cinn.es

Nanomaterials and Nanotechnology Research Centre (CINN-CSIC)

Research Article

Keywords: Colorectal cancer, Histone posttranslational modifications, Epigenomics, Mass spectrometry, Epigenetic drugs, Multiomics data integration

Posted Date: June 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9763213/v1

Additional Declarations: No competing interests reported.

Abstract

Background

Histone post-translational modifications (HPTMs) play central roles in chromatin regulation and cancer biology, yet their global and unbiased characterization in colon adenocarcinoma (COAD) remains limited. This gap has hindered a comprehensive understanding of the epigenetic mechanisms that shape tumor-specific transcriptional programs and influence clinical outcomes.

Methods

In this work we performed quantitative mass spectrometry–based profiling of HPTMs in paired tumor, peritumoral, and normal colon tissues from patients with COAD. These data were integrated with transcriptomic profiling and genome-wide CUT&RUN analyses to link chromatin modifications with gene expression and chromatin organization. Multi-omic integration was employed to identify coordinated histone marks crosstalk and candidate regulatory enzymes. Functional validation was carried out using pharmacological inhibition and CRISPR/Cas9-mediated knockout in colorectal cancer cell lines. To assess clinical relevance, machine learning approaches were applied to infer histone modification ratios in the TCGA-COAD cohort and to evaluate potential associations with patient outcomes. Tumor ecotypes were inferred using EcoTyper to contextualize epigenetic states within the tumor microenvironment.

Results

Comprehensive HPTM profiling revealed widespread tumor-specific epigenetic remodeling, whereas peritumoral tissues largely retained a normal HPTM landscape. Correlation analyses identified a robust inverse relationship between H3K14ac and H3K9me2, defining a previously underexplored epigenetic axis in COAD. Further integrative analyses implicated EHMT2 and HDAC7 as key regulators of this imbalance, and functional perturbation of these factors demonstrated that EHMT2 directly governs the H3K14ac-H3K9me2 switch, modulating chromatin state, transcriptional programs, and metabolic reprogramming in tumor cells. Genome-wide mapping supported a binary switch model within defined chromatin domains. In addition, machine learning–based inference revealed that a low H3K14ac/H3K9me2 ratio is associated with poor overall survival, particularly in non-irradiated patients, and correlates with more aggressive tumor ecotypes.

Conclusions

Our study provides a large-scale, unbiased proteomic characterization of HPTMs in COAD and identifies a clinically relevant H3K14ac/H3K9me2 epigenetic switch regulated by EHMT2. This axis integrates chromatin remodeling, transcriptional control, and metabolic reprogramming, stratifies patient prognosis, and highlights epigenetic vulnerabilities with potential therapeutic relevance in COAD. These findings reveal a new mechanism of tumor-specific epigenetic regulation and highlight opportunities for targeted epigenetic interventions

BACKGROUND

Colorectal cancer (CRC) accounts for ~ 10% of all cancers and remains the third leading cause of cancer-related mortality worldwide[1]. CRC is a multifactorial disease driven by the interplay of genetic, epigenetic, and environmental factors, which collectively contribute to its biological complexity and clinical heterogeneity. Omics-based studies have substantially advanced our understanding of CRC, revealing key molecular features and subclassifications, particularly in colorectal adenocarcinoma (COAD), the most prevalent form of CRC. At the genomic level, CRC can be broadly categorized into hypermutated and non-hypermutated subtypes. Hypermutated tumors frequently exhibit microsatellite instability (MSI), often resulting from MLH1 promoter hypermethylation or inactivating mutations in mismatch repair (MMR) genes or in POLE. In contrast, non-hypermutated tumors typically harbor recurrent alterations in canonical driver genes, including APC, TP53, SMAD4, PIK3CA, and KRAS, as well as mutations in ARID1A, SOX9, and FAM123B[2, 3]. Transcriptomic profiling has further refined this classification into Consensus Molecular Subtypes (CMS), which classify CRC into four groups with distinct biological characteristics and clinical implications (CMS1 – MSI-Immune; CMS2 – Canonical; CMS3 – Metabolic; and CMS4 – Mesenchymal)[4]. Adding another layer of molecular complexity, recent integrative proteogenomic analyses combining mass spectrometry–based proteomic data with genomic and transcriptomic profiles have elucidated post-transcriptional and post-translational regulatory mechanisms. These studies have uncovered distinct protein-expression signatures and signaling-related post-translational modifications associated with CRC progression, metastatic behavior, and therapeutic response[5].

Although the epigenetic landscape of CRC has been extensively characterized at the level of DNA methylation[6], the landscape of histone post-translational modifications (HPTMs) remains poorly characterized. Antibody-based approaches have identified several dysregulated HPTMs in CRC, including repressive marks such as H3K27me3 and H3K9me3[7, 8], which are often associated with EZH2 activity and transcriptional silencing, as well as activating marks such as H3K4me3, H3K36me3, H3K27ac, and H3K9ac, which are perturbed through deregulation of histone-modifying enzymes[9]. These alterations have been implicated in tumor progression, immune modulation, and therapeutic resistance. However, to date, no large-scale proteomic study has comprehensively mapped HPTMs in colon adenocarcinomas (COAD), even though similar approaches have successfully revealed epigenetic alterations in other cancer types[10]. To address this gap, we performed unbiased quantitative profiling of HPTMs using liquid chromatography–mass spectrometry (LC-MS)[11] coupled with the EpiProfile software[12]. This approach enabled the simultaneous detection of a wide array of HPTMs and histone variants in a single analysis, revealing novel modification patterns potentially involved in CRC pathogenesis. Notably, we identified a binary epigenetic switch defined by an inverse relationship between H3K14ac and H3K9me2. Integration with matched RNA-sequencing data implicated *EHMT2* and *HDAC7* as possible candidate genes mediating this crosstalk, which we validated through pharmacological and genetic perturbation. A machine learning–based inference model applied to the TCGA-COAD cohort showed that a low H3K14ac/H3K9me2 ratio correlates with poor prognosis, particularly in non-irradiated patients.

Moreover, digital cytometry via EcoTyper linked this epigenetic signature to aggressive tumor ecotypes and reduced overall survival. These findings uncover a novel, clinically relevant HPTM axis in CRC, supporting its potential as a prognostic biomarker and therapeutic target.

METHODS

Sample recruitment and ethics statement

A total of 24 matched tumor, peritumoral mucosa, and normal mucosa specimens were collected from patients diagnosed with stage II–III CRC, restricted in this study to colon adenocarcinoma, at the Hospital Universitario Central de Asturias (HUCA), Spain. All samples were obtained during surgical resection, immediately snap-frozen in liquid nitrogen, and stored at − 80°C until processing. None of the patients received neoadjuvant or adjuvant therapy prior to surgery. For each patient, three tissue types were collected: (i) tumor tissue from the primary lesion, sampled from two spatially distinct regions of the same tumor; (ii) peritumoral mucosa, defined as macroscopically normal mucosa located 2–5 cm from the tumor margin; and (iii) normal mucosa obtained from a distant, uninvolved segment of the colon. All specimens were histologically reviewed by a certified pathologist to confirm tissue identity and to verify the absence of tumor cells in peritumoral and normal mucosa samples. Clinical data for the TCGA-COAD cohort were obtained from the UCSC Xena Browser database[13] (Additional File 1: Table S1).

Cell lines and culture conditions

The human colon cancer cell lines DLD1 (RRID:CVCL_0248), HCT116 (RRID:CVCL_0291), SW480 (RRID:CVCL_0546), SW1116 (RRID:CVCL_0544), LoVo (RRID:CVCL_0399), RKO (RRID:CVCL_0504), Caco-2 (RRID:CVCL_0025), HT-29 (RRID:CVCL_0320), Colo205 (RRID:CVCL_0218) and human intestinal colonocytes CCD112 (RRID:CVCL_6382) were obtained from the American Type Culture Collection (ATCC). All cell lines were cultured in high-glucose Dulbecco's Modified Eagle Medium (DMEM) (Gibco, #41965039) supplemented with 10% heat-inactivated fetal bovine serum (FBS; Sigma-Aldrich, #F6178) and 1% penicillin–streptomycin (Gibco, #15240062). Cells were maintained in a humidified incubator at 37°C with 5% CO_2. Subconfluent cultures were passaged routinely, and cells used for experiments were between passages 3 and 6. Mycoplasma contamination was monitored monthly using standard PCR-based assays.

Cell viability assay

Cell viability was assessed using the CellTiter-Blue assay (Promega, #G8080) according to the manufacturer's instructions. Colorectal cancer cell lines were seeded in 96-well plates at a density of 1×10^4 cells per well and allowed to adhere for 24 h before treatment. Cells were then exposed to increasing concentrations of UNC0642 (1, 5, 10, and 25 μM; MedChemExpress, #HY-13980) or Apicidin

(MedChemExpress, #HY-N6735), along with a vehicle control (0.5% DMSO), for 48 h. Following treatment, CellTiter-Blue reagent was added to each well and incubated for 3 h at 37°C. Fluorescence was measured using an EnVision plate reader (PerkinElmer) at excitation/emission wavelengths of 560/590 nm. Cell viability was determined by normalizing fluorescence values to the Dimethyl Sulfoxide (DMSO) control. Proliferation assays in stable knockout lines (DLD1 EHMT2 KO, DLD1 HDAC7 KO, SW1116 EHMT2 KO, and SW1116 HDAC7 KO) were performed similarly. Cells were seeded in 96-well plates, and viability was measured at multiple time points. Fluorescence values were normalized to the baseline (time 0), and growth rates were compared with the corresponding mock control lines (DLD1 mock and SW1116 mock).

Histone Extraction

Histones were obtained from approximately 5 × 10^6 cultured cells or 25 mg of snap-frozen human tissue. For tumor samples, two tissue pieces obtained from distinct regions of the tumor were processed independently to assess reproducibility (Additional File 2: Figure S1). Nuclei were first isolated by resuspending samples in ice-cold nuclear extraction buffer (10 mM Tris-HCl, pH 7.4; 10 mM NaCl; 3 mM $MgCl_2$) supplemented with protease inhibitors (Roche, #12352204), followed by lysis with 0.2% NP-40. For tissue samples, mechanical homogenization was performed after initial mincing to facilitate nuclear release. Nuclei were collected by centrifugation at 1,000 × g for 5 min at 4°C and washed twice with detergent-free nuclear extraction buffer. Histones were extracted as previously described[14] with minor modifications. Nuclei were first treated with 1 volume of 0.4 M H_2SO_4, followed by 4 volumes of 0.2 M H_2SO_4 containing 1% β-mercaptoethanol. Samples were incubated with gentle agitation for 1 h at 4°C and centrifuged at 10,000 rpm for 10 min to remove acid-insoluble material. A second extraction with 0.2 M H_2SO_4 (1% β-mercaptoethanol) was performed under the same conditions. Supernatants from both extractions were pooled, and histones were precipitated with ice-cold 100% acetone. Precipitated proteins were collected by centrifugation at 10,000 rpm for 10 min, resuspended in distilled water, and stored at − 20°C. Protein concentrations were determined using the Qubit Protein Assay Kit (Thermo Fisher Scientific, #Q33212) according to the manufacturer's instructions.

Histone Propionylation and Digestion

Acid-extracted histones (~ 50 µg) were processed for mass spectrometry as previously described[11]. Briefly, histones were resuspended in 40 µL of 50 mM ammonium bicarbonate (NH_4HCO_3, pH 8.0) and mixed with 10 µL of freshly prepared derivatization solution (propionic anhydride:acetonitrile, 1:3 v/v; final reaction ratio 1:4 v/v). Ammonium hydroxide was added immediately to restore the pH to 8.0. Samples were incubated at 37°C for 15 min, dried under vacuum, and subjected to a second round of derivatization. Proteins were then digested overnight at 37°C with trypsin (Promega, #V5280) at an enzyme-to-substrate ratio of 1:20 in 50 mM NH_4HCO_3. Following digestion, a final derivatization step was performed to modify peptide N-termini.

Mass spectrometry and data analysis

Digested samples (500 ng equivalent) were loaded on Evotips (Evosep) following manufacturer's instructions, and were analyzed in a hybrid Q-TOF mass spectrometer (ZenoTOF 7600, Sciex) coupled to a Evosep One (Evosep) liquid chromatography system. The peptide digests were separated using the 30 samples per day Evosep program (44 min total run time) with water and acetonitrile, both with 0.1% formic acid, as solvent A and B, respectively. A 15 cm x 150 µm i.d. column (Performance, Evosep), operated at 40°C, was used for peptide separation. An Optiflow electrospray ion source (Sciex) with a low-micro electrode was used for peptide ionization, applying a voltage of 4200 V (positive mode). The mass spectrometer was operated using a top 100 DDA mode, with cycles of one TOF MS scan (300 to 1400 *m*/z, 150 ms accumulation time) followed by 100 TOF MS/MS scans of the top 100 ions as found in the previous survey TOF MS scan. For these MS/MS scans, ions with charge state 2 to 5 were fragmented with collision induced dissociation (CID) using dynamic collision energy, and were acquired from 100 to 1400 *m*/z for 20 ms, with Zeno pulsing activated. The following candidate ions were included in an inclusion list: 528.30, 570.84, 754.93, 761.94, 768.95, 557.98, 543.98, 548.65, 563.32, 549.31, and 553.98 *m*/z. Data were acquired with the SciexOS software v3.1.6.44 (Sciex). Automatic calibration at the TOF MS and MS/MS levels was performed after each sample with the X500 ESI Positive Calibration Solution (Sciex) using the calibrant delivery system of the mass spectrometer.

Differential peptide / HPTM analysis

Differentially modified peptides or PTMs between tumor, peritumoral, and normal mucosa samples, as well as between treated (pharmacological or genetic ablation) and corresponding control conditions, were identified using processed Epiprofile ratios. Peptides or PTMs with zero variance across all samples were excluded from downstream analyses. To ensure homoscedasticity, ratios were transformed to M-values using the *lumi* R/Bioconductor package (v2.56.0)[16]. Significant changes in HPTMs were assessed using the moderated t-test implemented in the *limma* package (v3.60.6)[17], incorporating patient pairing information (tumor vs. normal) in the model's design matrix. A p-value < 0.05 was used to define statistically significant alterations in the conditions described above.

RNA sequencing and data analysis.

Total RNA was extracted using the RNeasy Mini Kit (QIAGEN, #QIA74106) according to the manufacturer's instructions. RNA sequencing was performed by Genewiz (Azenta Life Sciences) using the NEBNext rRNA Depletion Kit and the NEBNext Ultra RNA Library Prep Kit for Illumina. Libraries were clustered on an Illumina NovaSeq platform and sequenced using a 2 × 150 bp paired-end (PE) configuration. We used *fastp* (v20.1)[18] for adapter removal, and the subsequent filtered reads were aligned to the human GRCh38 genome using *RSEM* (v1.3.3)[19] including the hisat2-hca PE mode. We estimated the counts per gene using the RSEM output and the tximport function of the R/Bioconductor package *tximport* (v1.32.0)[20]. We imputed the consensus molecular subtype CMS using the R package

CMScaller (v2.0.1)[21]. Low-abundance genes with less than 10 counts (per sample) or less than 200 counts (all samples) were discarded for downstream analyses. Differential gene expression analyses between paired tumor and control samples, or between KO vs control experiments were performed using the R/Bioconductor package *DESeq2* (v1.44.0)[22]. The resulting *P*-values were corrected using the Benjamini and Hochberg method, and we considered a threshold of adj. *P*-value < 10^{-3} to indicate statistical significance in regard to differentially expressed genes (DEGs). Subsequent downstream analyses were performed using the normalized gene expression matrix (VST approach) as described in Santamarina-Ojeda and colleagues[23]. Gene set enrichment analyses (GSEA) were performed with *GSEA* (v4.0.3)[24], using the vst normalized gene expression matrixes and the C2 and C5 MSigDB collections[25]. Classical gene ontology analyses were performed using the *GOrilla* tool[26] with the selected genes, and all the genes with detectable expression in the RNA-seq experiment as background. Further representation purposes were obtained using *Revigo*[27]. Survival data of COAD-TCGA patients was obtained from the XenaBrowser database [13] (Additional File 1 – Table S1). Survival estimates were generated using Kaplan–Meier plots, and differences between groups were assessed using the log-rank test.

Pairwise correlation analysis

Correlations between H3K14ac or H3K9me2 levels and gene expression were assessed using a robust correlation method, the percent bend correlation coefficient (pbcor), implemented in the R/CRAN *WRS2* package (v1.1.6)[28], using PTM ratios for these histone marks and the vst-normalized expression matrix derived from the RNA-seq data. Candidate epigenetic regulators were further prioritized using data from the EPIKOL CRISPR screening platform[29]. Correlations were considered significant when the absolute correlation coefficient between PTM and genes exceeded 0.3 and the p-value was < 0.05 for H3K9me2, H3K14ac, or both.

CRISPR/Cas9-mediated knockout of EHMT2 and HDAC7 in CRC cells line

Single-guide RNAs (sgRNAs) targeting human *EHMT2* and *HDAC7* were designed using the CRISPR-ERA tool[30]. Forward and reverse primers containing the sgRNA sequences targeting *EHMT2* and *HDAC7* (Additional File 3 – Table S2), along with flanking adaptor sequences, were annealed and cloned into BamHI/BsmBI-linearized pLenti-Guide-Puro vectors (Origene, GE100032) using the In-Fusion® HD Cloning Kit (Takara, #638911), according to the manufacturer's instructions. The resulting sgRNA-expressing pLenti-Guide-Puro constructs were transformed into *E. coli* DH5α via heat shock, and plasmids were purified with the QIAprep Spin Miniprep Kit (Qiagen, #27104). A dual-vector system was employed to deliver the sgRNAs and Cas9. Lentiviral particles were produced in Lenti-X™ 293T cells by co-transfecting psPAX2 (Addgene, #12260), pMD2.G (Addgene, #12259), and either lentiCRISPR v2-Blast (Addgene, #83480), scrambled control (Origene, #GE100032), or sgRNA-expressing pLenti-Guide-Puro vectors using Lipofectamine™ 3000 (Invitrogen, #L3000015). Briefly, for each transfection, 1 × 10^6 Lenti-

X™ 293T cells were seeded per 25 cm² flask and transfected with 5 µg psPAX2, 2 µg pMD2.G, and 6 µg of each transfer plasmid. Viral supernatants were harvested at 48 h and 72 h post-transfection, filtered (0.45 µm), and used to co-transduce DLD1 or SW1116 cells with the sgRNA-expressing pLenti-Guide-Puro constructs (targeting *EHMT2*, *HDAC7*, or scramble control), together with the lentiCRISPR v2-Blast vector. Stable CRISPR/Cas9 knockout cell lines were established using dual antibiotic selection with puromycin (2 µg/mL) and blasticidin (10 µg/mL) for two weeks. This resulted in the generation of the following stable cell lines: SW1116 *EHMT2* Knockout (KO), SW1116 *HDAC7* KO, DLD1 *EHMT2* KO, DLD1 *HDAC7* KO, as well as mock controls transduced with empty pLenti-Guide-Puro vectors: SW1116 mock and DLD1 mock. Successful gene knockout was confirmed by Western blot analysis.

Immunoblot

Cell pellets were lysed in RIPA buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS) supplemented with protease inhibitors (Roche, #11873580001). A total of 30 µg of protein per sample was separated by SDS-PAGE on an 8% gel and transferred via wet blotting to 0.45 µm PVDF membranes (Millipore, #IPVH00010). Membranes were blocked in 5% BSA for 1 hour at room temperature and incubated overnight at 4°C with the following primary antibodies: rabbit anti–EHMT2/G9A (1:1000, Abcam, #ab185050), rabbit anti–HDAC7 (1:1000, Cell Signaling Technology, #33418), and mouse anti–α-Tubulin (1:1000, Sigma-Aldrich, #T5168). The following day, membranes were washed and incubated with the appropriate HRP-conjugated secondary antibodies: goat anti-rabbit IgG H&L (Abcam, #ab6728) and goat anti-mouse IgG H&L (Jackson ImmunoResearch, #115-035-062). Signal detection was performed using an ECL detection kit (Millipore, # WBLUC0100), and images were acquired with the Odyssey Fc Imaging System.

Cut and Run

CUT&RUN assays were performed in SW1116 and DLD1 EHMT2 knockout cell lines, as well as in their respective mock control (SW1116 mock DLD1 mock). The assay was conducted using the CUTANA® ChIC/CUT&RUN Kit v4 (Epicypher, #14-1048), following the manufacturer's instructions. Briefly, approximately 500,000 cells per condition were harvested and incubated with activated ConA magnetic beads for 10 minutes at room temperature. The cell-bead complexes were then incubated with 0.5 µg of primary antibody targeting either H3K9me2 (Active Motif, #39239) or H3K14ac (Abcam, #ab52946). A negative control antibody (CUTANA™ Rabbit IgG, Epicypher, #13-0042k) and a positive control antibody (H3K27me3, SNAP-Certified™ for CUT&RUN, Epicypher, #13–0055) were also included in parallel reactions. The mixtures were incubated overnight on a nutator at 4°C. On the second day, pAG-MNase and calcium chloride were used to activate the cleavage of the target DNA. Chromatin digestion and DNA release were performed according to the manufacturer's protocol. DNA was purified using SPRI magnetic beads as specified in the kit instructions. Approximately 5–10 ng of CUT&RUN-enriched DNA per sample was used to generate sequencing libraries with the CUT&RUN Library Prep Kit (Epicypher, #14-1001), following the manufacturer's protocol. Library fragment size and quality were assessed using an Agilent TapeStation system. High-throughput sequencing was performed by CD Genomics on an Illumina NovaSeq 6000 platform, employing paired-end 150 bp reads (PE150). For each sample, a depth

of approximately 20 million reads was obtained to ensure adequate genome coverage and resolution of chromatin-bound regions.

Cut and run analysis

CUT&RUN data for the histone marks H3K9me2, H3K14ac, and H3K27me3 were processed using the nf-core/cutandrun pipeline (v3.2.2)[31] with the GRCh38 reference genome in SW1116 and DLD1 cells. Mitochondrial reads and reads mapping to ENCODE blacklisted regions were removed prior to downstream analyses. For visualization purposes, BigWig files were generated from BAM files using the bamCoverage function in *deepTools* (v3.5.6)[32] with a bin size of 500 bp, read extension enabled, a maximum fragment length of 2,000 bp, an effective genome size of 2,913,022,398, and normalization to reads per genomic content (RPGC). Signals from chromosomes X and Y were excluded. Log_2 differences between knockout and control samples were computed using the bigwigCompare function from deepTools with a bin size of 500 bp and a pseudocount of 1. Differential peak analysis for H3K9me2, H3K14ac, and H3K27me3 was performed using the ultra-fast implementation of SICER via *epic2* (v0.0.52)[33]. BEDPE files were generated from sorted BAM files corresponding to each histone mark and their respective IgG controls using the bamtobed function from *bedtools* (v2.31.1), following the epic2 workflow. Sorted BED files were then used as input for the epic2-df function with a bin size of 500 bp and allowing up to five gaps. Differential peaks were defined using a false discovery rate (FDR) < 0.001 and an absolute log_2 fold change > 1 between knockout and control conditions. In parallel, differential broad peaks for the same histone marks were identified using hidden Markov models implemented in the *HMMtBroadPeak* R package[34].

To assess genome-wide signal overlap between H3K9me2–H3K14ac, H3K9me2–H3K27me3, and H3K14ac–H3K27me3, the GRCh38 genome was segmented into 5-kb bins, and log_2 ratios (knockout versus control) were extracted for each histone mark using bigWigCompare from *deepTools*. An analogous analysis was conducted at the peak level using signal intensities assigned to each set of differentially enriched peaks identified above. As a control, random sets of non-significant peaks were generated using the shuffle function from *bedtools*, matching the number of significant peaks identified by epic2 for each condition. Histone mark tracks were visualized using the UCSC Genome Browser.

Inference of H3K14ac/H3K9me2 ratio in TCGA samples

Inference of H3K14ac/H3K9me2 ratio in TCGA samples was performed using machine learning (ML) models implemented in the R/CRAN package *caret* (v6.0.94). For each sample in our local cohort, the H3K14ac/H3K9me2 ratio was first calculated by dividing the corresponding PTM-level signals. Genes significantly associated with this continuous ratio were then identified using *DESeq2*, applying an adjusted p-value threshold of $< 10^{-3}$ to define differentially expressed genes (DEGs). A normalized gene expression matrix ($log_2(TPM + 1)$) containing these DEGs was used for model training. Four ML algorithms were evaluated for regression analysis: Random Forest, glmnet (lasso-based regularized regression), linear regression, and gradient boosting (gbm). Each model was trained using the normalized gene expression matrix and the H3K14ac/H3K9me2 ratio as the outcome variable,

employing repeated cross-validation (10 folds, 5 repeats). Model performance was assessed using multiple metrics, including RMSE, R^2, and MAE. The Random Forest model exhibited the best performance and was therefore selected for downstream inference in TCGA-COAD. Predicted H3K14ac/H3K9me2 ratio in TCGA samples were obtained using the predict function, applying the trained Random Forest model to the normalized gene expression matrix ($\log_2$(TPM + 1)) restricted to the DEGs of interest in the TCGA-COAD dataset.

Ecotyper analysis

Inference of molecular ecotypes was performed using *EcoTyper*, a machine learning framework designed for large-scale identification of cell type–specific transcriptional states and their co-association patterns from bulk RNA-seq data[35]. A normalized gene expression matrix containing TPM values from our local cohort was used as input for the EcoTyper recovery bulk workflow, specifying “Carcinoma” as the discovery dataset. EcoTyper results for the TCGA-COAD cohort were obtained from the original EcoTyper publication[35]. For each State and Ecotype across all samples analyzed, robust correlations were computed using the pbcor function from the R/CRAN package *WRS2*, as described above.

RESULTS

Global HPTM remodeling in colon adenocarcinoma revealed by LC–MS profiling

To define COAD-specific changes in the global histone post-translational modification (HPTM) landscape, we analyzed paired fresh colorectal tumor, peritumoral and matched normal mucosa tissues obtained from 24 patients recruited at the Hospital Universitario Central de Asturias (Fig. 1A) using comprehensive LC–MS/MS profiling with the EpiProfile workflow. [11, 12](Additional File 4 – Table S3). We identified 12 HPTMs that were significantly altered between normal mucosa and tumor tissue, namely H3K14ac, H3K9me1, H3K9me2, H3K9me3, H3K27me3, H4K20me3, H2B1F, H2AVK7ac, H1-3, H1-4, H2B1C, and H2B1D (Additional File 5 – Table S4). In contrast, only a single HPTM (H2A19ac) was significantly different between peritumoral tissue and its matched control (Fig. 1B), highlighting a pronounced epigenomic remodeling specific to tumor samples in COAD. These tumor-associated HPTMs robustly segregated colon cancer cell lines and tumor samples from peritumoral and normal mucosa (Figs. 1C and 1D). In addition, this analysis revealed a significant anticorrelation between H3K14ac and H3K9me2 levels, detected both at the global PTM level (Figs. 1E and 1F) and at the individual peptide level, specifically within histone H3 peptides spanning residues 9–17, and in peptides carrying either K9me2 or K14ac as isolated modifications (Additional File 6 – Figure S2). In COAD samples, H3K9me2 increased on K14-unmodified peptides, whereas H3K14ac decreased on K9-unmodified peptides. Peptides carrying both K9me2 and K14ac did not exhibit this reciprocal pattern, indicating that the observed shift occurs primarily between two mutually exclusive single-modified states. The consistency of this pattern prompted us to further investigate the functional implications of the H3K14ac and H3K9me2 modifications in COAD.

Identification of chromatin regulators driving H3K14ac - H3K9me2 imbalance in colon adenocarcinoma

To identify potential regulators mediating the observed alterations in H3K14ac and H3K9me2 levels, we performed transcriptomic RNA-seq analyses on the same set of paired tumor and control samples profiled by LC–MS/MS (Additional File 7 – Table S5). As expected, tumor and control samples exhibited substantial differences in gene expression, with the 1,000 most variable genes clearly segregating controls from tumors (Fig. 2A). Differential expression analysis identified 1,385 downregulated and 1,857 upregulated genes in COAD (Fig. 2B), forming two distinct clusters corresponding to control and tumor samples, respectively (Fig. 2C). Validation using a publicly available dataset (TCGA-COAD) (Additional File 8 – Table S6) revealed strong concordance with our results, including 1,025 commonly downregulated and 1,419 commonly upregulated genes, demonstrating a strong correlation of fold changes between tumor and control tissues across both cohorts. (Additional File 9 - Figure S3, Additional File 10 – Table S7).

To infer potential links between gene expression and the global alterations in H3K14ac and H3K9me2, we performed a robust pairwise correlation analysis examining all gene–HPTM pairs for these marks (Fig. 2D). We focused on genes that were significantly correlated or anticorrelated with both histone marks (highlighted in red) and those with known epigenetic functions according to the chromatin-focused CRISPR/Cas9 screening platform EPIKOL[29]. Within this subset, we prioritized genes exhibiting significant gene expression changes in both our local cohort and the TCGA-COAD dataset (Fig. 2E, Additional File 11 – Table S8). This analysis highlighted the chromatin regulators EHMT2 (G9a) and HDAC7 as top candidates, based on the strength of their correlation with the respective histone marks (Fig. 2F) and the reproducibility of their expression changes across cohorts (Additional File 9 – Figure S3F). EHMT2 was overexpressed in tumor samples and showed a direct correlation with H3K9 methylation, while HDAC7 was also upregulated in tumors and inversely correlated with H3K14 acetylation, suggesting that these enzymes may directly drive the tumor-specific remodeling of these two histone post-translational modifications.

Pharmacological inhibition of EHMT2 or HDAC7 modulates the H3K14ac/H3K9me2 ratio in colorectal cancer cell lines

To assess the functional contribution of EHMT2 and HDAC7 in COAD, we evaluated cell viability in a panel of 10 CRC cell lines treated with either the histone deacetylase (HDAC) inhibitor Apicidin or the EHMT2 (G9a) specific inhibitor UNC0642. Both compounds induced a dose-dependent decrease in cell viability across multiple cell lines (Fig. 3A). We next performed an Epiprofile-based analysis in two representative cell lines (SW1116 and DLD1), which revealed that both inhibitors altered the global HPTM profile of the H3 peptide encompassing residues 9–17 (H3_9_17), containing the H3K14ac and H3K9me2 marks (Fig. 3B, Additional File 12 – Table S9). Treatment with UNC0642 consistently reduced H3K9me2 levels and increased H3K14ac levels in a dose-dependent manner (Fig. 3C), predominantly affecting peptides with individual modification states, including K9me2, K14ac, and unmodified forms,

indicating that pharmacological inhibition of EHMT2 selectively modulates the anticorrelation between these two histone marks in CRC cells (Additional File 13 – Figure S4). In contrast, treatment with Apicidin also reduced H3K9me2 and increased H3K14ac in a dose-dependent manner, but exerted a broader effect by enhancing H3K14 acetylation across both unmodified and K9-methylated histone contexts, consistent with a redistribution of combinatorial histone modification states and a global hyperacetylation phenotype (Additional File 13 – Figure S4). These results indicate that pharmacological inhibition of EHMT2 effectively modulates the anticorrelation observed between these two histone marks while inhibition of HDAC7 more broadly increased H3K14ac levels in CRC cells.

Genetic ablation of EHMT2 impairs the H3K14ac/H3K9me2 ratio in colorectal cancer cell lines

As epigenetic drugs may exert unintended off-target effects, we sought to validate our findings using CRISPR/Cas9-mediated knockout (KO) of EHMT2 or HDAC7. Although both genes were efficiently depleted under our experimental conditions (Additional File 14 – Figure S5A), loss of HDAC7 produced no detectable effects on cell viability (Additional File 14 – Figure S5B) or on the HPTM landscape in either DLD1 or SW1116 cells, as shown by Epiprofile analysis (Additional File 13 – Figure S4D, Additional File 14 – Figures S5C-S5D, Additional File 15 – Table S10). In contrast, EHMT2 KO markedly reduced cell viability in SW1116 cells and led to a robust alteration of the H3K14ac/H3K9me2 axis. Specifically, EHMT2 KO caused a significant decrease in H3K9me2 levels in both cell lines, accompanied by a selective increase in H3K14ac and in the unmodified H3_9_17 peptide (Figs. 4A–B, Additional File 13 – Figure S4C, Additional File 15 – Table S10), while other PTM combinations remained largely unaffected. These results confirm the direct contribution of EHMT2 to the regulation of the H3K14ac/H3K9me2 imbalance observed in CRC. To investigate potential pathways regulated by EHMT2, we performed differential gene expression profiling in EHMT2 KO DLD1 and SW1116 cells relative to their respective controls. In SW1116 cells, EHMT2 KO led to 1,046 significantly downregulated and 1,871 significantly upregulated genes, whereas in DLD1 cells, 2,078 transcripts were downregulated and 1,938 upregulated (Fig. 4C, Additional File 16 – Table S11). Given that EHMT2 is overexpressed in CRC and may drive tumor-specific transcriptional programs, we sought to determine which genes are directly responsive to its modulation. We focused on transcripts exhibiting consistent directional changes in both EHMT2 KO cell lines, and displaying opposite trends in tumor versus normal mucosa in our local cohort and TCGA-COAD dataset. This analysis identified 89 genes that were upregulated upon EHMT2 depletion but downregulated in CRC, and 62 genes downregulated upon EHMT2 KO but upregulated in CRC (Fig. 4D). The first group of genes was enriched for pathways involved in metabolic rewiring, including lipid catabolism and mitochondrial function (Fig. 4E; Additional File 17 – Figures S6B–S6C). In contrast, the second group showed significant enrichment for cell-cycle regulation and DNA replication pathways (Fig. 4F; Additional File 17 – Figures S6A–S6C). Together, these findings indicate that EHMT2-mediated epigenetic remodeling reprograms tumor cell biology toward a more proliferative and replication-competent state.

EHMT2 loss drives genome-wide redistribution of H3K9me2 and H3K14ac signals

Considering the global impact observed in EHMT2-knockout cells, we performed a more comprehensive analysis to evaluate the extent of the genome-wide correlation between H3K9me2 and H3K14ac. To this end, we conducted CUT&RUN profiling of H3K14ac and H3K9me2 in SW1116 and DLD1 cells with or without functional EHMT2 (see Methods). After obtaining CUT&RUN signals and peak calls, we segmented the genome into 5,000 bp bins (global signal quantification, Fig. 5A) to quantify changes in fold-signal correlations between these histone marks in EHMT2-knockout versus wild-type cells. We observed that EHMT2 knockout resulted in a global reduction of H3K9me2 signal, accompanied by a compensatory global increase in H3K14ac signal, particularly in DLD1 cells (Fig. 5A). These signal distributions were specific, as profiling of other histone marks, such as H3K27me3, produced distinct correlation patterns. We next examined these correlations at the level of significantly detected peaks across conditions. As expected, loss of EHMT2 caused a genome-wide depletion of H3K9me2 peaks and a substantial increase in H3K14ac peaks, as detected by multiple computational algorithms (EPIC2, HMMBroadPeak; Fig. 5B, Additional File 18 – Table S12, Additional File 19 – Table S13). The global shifts between H3K14ac and H3K9me2 signals within these peaks recapitulated the genome-bin analysis, whereas randomly selected genomic regions showed no consistent, distinct pattern following EHMT2 knockout (Additional File 20 – Figure S7). Together, these results highlight that the global increase in H3K14ac upon EHMT2 loss occurs at the expense of H3K9me2 across the genome. To recapitulate these findings, we focus on two genes with potential clinical relevance in COAD. We highlighted the mitochondrial genes, *ATP5F1B* and *ATP5MC3*, whose genomic regions display reduced H3K9me2 and concomitant gains in H3K14ac in EHMT2-knockout cells (Fig. 5C). These epigenomic changes lead to restored gene expression in the knockout cells. Consistently, in COAD, where EHMT2 is overexpressed and associated with the H3K14ac/H3K9me2 axis, these genes are downregulated in tumors versus normal tissue in both our local cohort and the TCGA-COAD dataset (Fig. 5D). Importantly, higher expression of both mitochondrial genes is associated with improved prognosis in COAD patients (Fig. 5E), suggesting a potential prognostic relevance for global H3K14ac and H3K9me2 levels in colorectal cancer.

Prognostic value of H3K14ac/H3K9me2 levels in colon adenocarcinoma

Based on the evidence described above, we hypothesized that the H3K14ac/H3K9me2 ratio may hold prognostic value in COAD. To evaluate this, we leveraged the independent TCGA-COAD cohort. As TCGA data do not include direct measurements of histone PTM abundance, we inferred the H3K14ac/H3K9me2 ratio using an indirect machine-learning-based approach (Fig. 6A). Given the strong concordance in gene expression patterns between our local cohort and TCGA-COAD, we derived a gene-expression signature associated with the H3K14ac/H3K9me2 ratio in our dataset, identifying 76 negatively and 51 positively correlated genes (Fig. 6B). Using these genes and the corresponding

histone-ratio values in the local cohort, we trained several machine-learning models to predict the H3K14ac/H3K9me2 ratio; selecting Random Forest for downstream analyses due to its superior performance across multiple evaluation metrics (Additional File 21 – Figure S8). When applied to TCGA-COAD transcriptomic data, the model inferred H3K14ac/H3K9me2 ratio was consistently reduced in tumor samples compared with normal tissue (Fig. 6C). This result indicates that the inferred signature not only discriminates tumor from control samples but also captures variation in the relative balance of these two histone marks across CRC tumors.

To explore the biological pathways associated with this epigenetic axis, we performed gene set enrichment analyses (GSEA) using all genes correlated with the inferred H3K14ac/H3K9me2 ratio in both cohorts. We identified a robust and reproducible enrichment for pathways involved in metabolic regulation, including multiple canonical pathways (MSigDB C2; Figs. 6D-E) and gene-ontology terms related to metabolic processes (MSigDB C5; Additional File 22 – Figures S9A-S9B). To further assess the prognostic relevance of the inferred H3K14ac/H3K9me2 ratio in the TCGA-COAD cohort, we performed survival analyses stratifying patients according to the median value of this histone score. Notably, patients with higher inferred H3K14ac/H3K9me2 ratio exhibited significantly improved overall survival compared with those with lower ratio, an association that was particularly evident in tumors not treated with radiotherapy (Fig. 6F, Additional File 23 – Table S14). To evaluate whether this prognostic signal could be attributed to the expression of the regulatory enzymes themselves, we examined survival as a function of *EHMT2* or *HDAC7* expression. While elevated *EHMT2* expression showed a trend toward poorer survival, this association became statistically significant specifically in non-irradiated patients (Additional File 22 – Figure S9C, Additional File 23 – Table S14,), supporting the notion that *EHMT2* contributes to the imbalance of this epigenetic axis in CRC.

Finally, we assessed the prognostic relevance of the inferred H3K14ac/H3K9me2 ratio in the context of tumor ecotypes using the EcoTyper framework, applied to both our local cohort and TCGA-COAD samples. Among the significant ecotypes shared across cohorts, we observed a negative correlation between the histone ratio and the lymphocyte-deficient CE2 and age-associated CE8 ecotypes, and a positive correlation with more benign ecotypes such as CE6 and CE10, which are enriched in normal tissue or characterized by higher B-cell content (Additional File 24 – Figure S10). Together, these findings indicate that increased H3K14ac/H3K9me2 ratio is associated with more benign tumor ecotypes, consistent with the improved overall survival observed in these patients.

DISCUSSION

Although histone post-translational modifications play critical roles in colorectal cancer, their comprehensive and unbiased characterization remains limited, leaving the epigenetic landscape only partially defined. In this study, we performed high-resolution mass spectrometry–based profiling of HPTMs in tumor, peritumoral, and normal colon mucosa, enabling a systematic assessment of global HPTM patterns and potential crosstalk between distinct histone marks. This analysis revealed multiple statistically significant alterations in tumor samples compared with matched controls tissues,

encompassing both previously reported modifications (including H4K20me3, H3K9me3, and H3K27me2) [36–39] and previously uncharacterized HPTMs, thereby expanding the known epigenetic complexity of COAD. Although peritumoral regions are often considered early sites of molecular reprogramming, we observed minimal differences between peritumoral and normal tissues at the HPTM level, suggesting that the peritumoral compartment largely retains a normal histone modification profile. Nevertheless, subtle alterations cannot be excluded, particularly at other molecular layers or in larger cohorts. Importantly, tumor-specific HPTM patterns robustly distinguished malignant tissue from both peritumoral and control samples in unsupervised clustering analyses, underscoring the specificity of these epigenetic alterations to cancer cells. Notably, correlation analyses uncovered potential functional interactions between histone marks, most prominently a consistent inverse relationship between H3K14ac and H3K9me2.

Bivalent modifications involving H3K9 methylation (me1/2/3) and H3K14 acetylation have been previously described[40], although most studies have historically focused on the H3K9me3/H3K14ac combination, due to the well-established role of H3K9me3 as a canonical mark of constitutive heterochromatin and transcriptional repression[41]. The H3K9me3/H3K14ac state is generally associated with transcriptional silencing; in mouse models, the coexistence of these marks has been proposed to establish a poised, transcriptionally repressed chromatin state, with loss of one or both modifications triggering transcriptional reprogramming[42]. Importantly, these marks are not strictly antagonistic. H3K14ac is required for SETDB1-mediated trimethylation of H3K9, which is subsequently followed by HDAC-dependent removal of H3K14ac to establish stable heterochromatin [43, 44]. This dynamic coordination suggests that while the presence of H3K14ac is necessary to license trimethylation at H3K9, it simultaneously primes the acetyl group for removal, highlighting a tightly regulated chromatin transition. Beyond this well-characterized interplay, the regulatory significance of H3K9me2 and its relationship with H3K14ac remain far less extensively characterized than those of the trimethylated state. In our COAD samples, we observed an increase in H3K9me2 accompanied by a concomitant reduction in H3K14ac, suggesting a shift toward a globally more repressive chromatin environment, which may constrain canonical gene activation while enabling tumor-specific transcriptional programs. H3K14ac and H3K9me2 exhibited a complementary pattern, with regions enriched in one mark generally showing lower relative levels of the other. In fact, peptide-resolved mass spectrometry revealed that this pattern results from a redistribution between H3K14ac unmodified/H3K9me2 and H3K9 unmodified/H3K14ac peptides, whereas the double-modified H3K14ac/H3K9me2 fraction remained largely unchanged. Notably, this pattern mirrors observations in induced pluripotent stem cells (iPSCs) and mouse embryonic fibroblasts (MEFs), where H3K14ac levels were markedly higher in iPSCs than in MEFs specifically when H3K9 was unmodified or acetylated on the same histone tail, further supporting a direct influence of H3K9 modifications on H3K14 acetylation[45].

The inverse relationship between H3K14ac and H3K9me2 observed in our colon adenocarcinoma models suggests the existence of a previously underexplored regulatory axis that may contribute to the establishment of tumor-specific transcriptional programs, underscoring the importance of

systematically dissecting the functional crosstalk between these modifications. Our integrative multiomic analysis combining RNA sequencing and mass spectrometry data identified EHMT2 and HDAC7 as candidate regulators underlying the H3K14ac/H3K9me2 imbalance in CRC, both of which were significantly upregulated in tumor samples. Consistent with previous reports [46, 47], EHMT2 expression and its catalytic product H3K9me2 were elevated in CRC; however, our data additionally reveal a novel inverse association between EHMT2 expression and H3K14ac levels, suggesting that EHMT2 may indirectly influence acetylation dynamics at neighboring residues. Gene expression analyses in the TCGA-COAD cohort confirmed the robustness of these transcriptional patterns and the conserved upregulation of both EHMT2 and HDAC7, reinforcing the generalizability of our findings and highlighting these enzymes as potential epigenetic therapeutic targets in COAD.

Treatment with the EHMT2-specific inhibitor UNC0642 selectively reduced H3K9me1 and H3K9me2 levels, while H3K9me3 remained largely unaffected, and promoted H3K14 acetylation predominantly on peptides harboring unmodified K9. In contrast, peptides retaining K9 methylation were minimally affected. These findings support our hypothesis and highlight a preferential crosstalk between H3K14ac/H3K9me2. Genetic ablation of EHMT2 recapitulated this pattern, supporting that EHMT2 directly contributes to the balance between H3K9 methylation and H3K14 acetylation, with loss of methylation being sufficient to shift chromatin toward a more acetylated state. Consistent with our observations, independent studies have reported that knockout of EHMT2[48] or SETDB1[43] not only reduces H3K9me1/2 or H3K9me3, respectively, but also markedly increases H3K14ac. Together, these data support a dynamic and reciprocal crosstalk between H3K9 methylation and H3K14 acetylation, in which the modification state of one residue directly influences the other, thereby shaping chromatin states.

By contrast, broad HDAC inhibition globally increased H3K14ac across peptides with both unmodified and methylated H3K9, suggesting a redistribution of acetylation rather than residue-specific modulation. This pattern likely reflects the non-selective activity of pan-HDAC inhibitors, which promotes acetylation independently of H3K9 methylation status, or an intrinsic flexibility in the methylation-acetylation interplay that permits H3K14ac even in the presence of H3K9 methylation. Of note, HDAC7 knockout did not alter global acetylation levels or the H3K14ac/H3K9me2 ratio in our CRC models, indicating a lack of intrinsic deacetylase activity in this context. Despite this, HDAC7 was consistently upregulated in CRC, suggesting a non-enzymatic role in chromatin regulation. This observation aligns with previous reports demonstrating that class IIa HDACs, including HDAC7, are catalytically inactive[49, 50] and function primarily as scaffolding proteins that recruit transcriptional regulators[51]. In this context, HDAC7 may act as a chromatin-associated factor that stabilizes EHMT2-dependent repressive states rather than serving as a classical histone deacetylase. Additionally, functional redundancy among HDAC family members may compensate for HDAC7 loss[52], thereby maintaining the altered H3K14ac/H3K9me2 balance observed in CRC cells. Our genome-wide mapping of H3K14ac and H3K9me2 reinforces the hypothesis that these marks function as a binary epigenetic switch within defined chromatin regions. Consistent with our mass spectrometry data and previous reports[53], H3K9me2 signal was globally reduced in

EHMT2 knockout cells, and this reduction coincided with a marked increase in H3K14ac at the same genomic regions, supporting a bivalent switching model in which modulation of one mark directly influences the other. Importantly, unlike H3K9me3, which marks constitutive pericentromeric heterochromatin[54, 55] and forms compact clusters at both the nuclear periphery and the nuclear compartment, H3K9me2 defines broader and more plastic repressive domains known as Large Organized Chromatin H3K9 modifications (LOCKs)[56]. These domains are predominantly located at the nuclear periphery and frequently overlap with lamina-associated domains (LADs) and inactive B compartments[57]. Although our dataset does not directly resolve higher-order chromatin compartmentalization, Smith et al. [55]. demonstrated that H3K9me2 levels within LADs inversely correlate with H3K14ac, highlighting a compartment-specific antagonism between these marks. In contrast, H3K9me2-marked regions outside LADs exhibit increased chromatin accessibility and enriched acetylation, suggesting that the functional consequences of H3K9me2 deposition are highly context dependent. Future studies will be required to delineate how these compartment-specific effects contribute to transcriptional regulation in COAD.

To examine the context-specific role of EHMT2 in COAD, we integrated transcriptomic data from EHMT2 knockout CRC cell lines with patient-derived expression profiles from our local cohort and the TCGA-COAD dataset. This analysis indicates that EHMT2 is associated with transcriptional programs linked to proliferation and metabolic adaptation in CRC, promoting DNA replication and repair while concurrently suppressing oxidative metabolism and lipid catabolism in favor of glycolytic pathways. Such metabolic reprogramming is consistent with a Warburg-like phenotype that supports tumor growth under hypoxic conditions[58]. Although H3K9me2 primarily establishes broad repressive chromatin domains rather than locus-specific silencing, the integration of genome-wide CUT&RUN and RNA-seq datasets enabled the identification of candidate genes within these megabase-scale regions that are directly influenced by H3K14ac-H3K9me2 crosstalk. Among these candidates, ATP5F1B, previously associated with increased metastatic potential and poor prognosis across multiple cancer types[59–61], was highly expressed in CRC tumors and correlated with reduced overall survival in the TCGA-COAD cohort, supporting its potential utility as a prognostic biomarker.

Building on these mechanistic insights, we assessed the prognostic relevance of the H3K14ac/H3K9me2 axis. Using a machine learning approach, we inferred the H3K14ac/H3K9me2 ratio in the TCGA-COAD cohort and observed that, consistent with our local cohort, tumors exhibited a lower ratio than normal tissue, reflecting increased H3K9me2 and reduced H3K14ac levels. Importantly, a low H3K14ac/H3K9me2 ratio was associated with poorer overall survival, particularly in patients who did not receive radiotherapy, whereas this prognostic signal was attenuated in radiotherapy-treated patients. These findings suggest that the H3K14ac/H3K9me2 ratio captures chromatin states that stratify patient prognosis in the absence of radiotherapy, while treatment may mitigate the clinical impact of these epigenetic vulnerabilities. Collectively, our data support a model in which H3K14ac and H3K9me2 act as a bivalent epigenetic switch controlling transcriptional programs relevant to CRC biology, and suggest

that chromatin states defined by this axis may serve as prognostic markers while highlighting potential vulnerabilities for targeted therapeutic intervention.

CONCLUSIONS

This study provides a comprehensive characterization of histone post-translational modifications in colorectal cancer, uncovering a previously underexplored regulatory axis between H3K14ac and H3K9me2. Our results support a model in which EHMT2 contributes to the establishment of this bivalent chromatin switch, influencing transcriptional programs associated with proliferation, metabolism, and tumor-specific gene expression. Notably, the H3K14ac/H3K9me2 ratio reflects distinct chromatin states that are associated with patient prognosis, particularly in the absence of radiotherapy, suggesting its potential value as a prognostic biomarker and an indicator of epigenetic vulnerability. Together, these findings provide a mechanistic framework linking epigenetic regulation to COAD progression and highlight the H3K14ac/H3K9me2 axis as a potential targetable node for future epigenetic strategies in precision oncology.

Abbreviations

ATCC - American Type Culture Collection

AUC - Area Under the Curve

CMS - Consensus Molecular Subtypes

COAD - Colon Adenocarcinoma

CRC - Colorectal cancer

DEG - Differentially Expressed Genes

DMEM - Dulbecco's Modified Eagle Medium

DMSO - Dimethyl Sulfoxide

EHMT2 - Euchromatic Histone-Lysine N-Methyltransferase 2

FBS - Fetal Bovine Serum

GSEA - Gene Set Enrichment Analyses

HDAC - Histone Deacetylase

HPTMs - Histone Post-Translational Modifications

HUCA - Hospital Universitario Central de Asturias

IPSC - Induced Pluripotent Stem Cells

KO - Knock-out

LADs - Lamina-Associated Domains

LC-MS - Liquid Chromatography–Mass Spectrometry

LOCKs - Large Organized Chromatin H3K9 modifications

MEF - Mouse Embryonic Fibroblasts

MMR - MisMatch Repair

MSI - MicroSatellite Instability

PE - Paired-end

PTM - Post-Translational Modifications

RSEM - RNA-Seq by Expectation-Maximization

sgRNAs - Single-guide RNAs

TCGA - The Cancer Genome Atlas

VST - Variance Stabilizing Transformation

Declarations

Ethics approval and consent to participate

All procedures involving human samples were approved by the Research Ethics Committee of the Principality of Asturias (code: 2021.421). Written informed consent was obtained from all participants, and all procedures adhered to institutional guidelines and the principles of the Declaration of Helsinki.

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests

Funding

This work has received funding from the Health Research Institute Carlos III (Plan Nacional de I+D+I) co-funding FEDER (PI21/01067 and PI24/00641 to M.F.F and A.F.F), the Spanish Association Against Cancer (AECC, PRYGN235109FERN to M.F.F.), the Asturias Government (PCTI) (IDE/2021/000077 and IDE/2024/000744 to M.F.F.), the European Commission-NextGenerationEU, through CSIC's Global Health Platform (PTI Salud Global), the European Commission-Horizon 2020 (EprObes and AC24/00163 to M.F.F.), the Spanish Ministry for Digital Transformation and Public Function (MMT24-FAGOTECA-01 to J.R.T.), Asociación Galbán (2023-165-GALBAN-TEVAJ to J.R.T.) and the Spanish Ministry of Science and Innovation through the Recovery, Transformation and Resilience Plan (GL2021-03-39 and GL2021-03-040 to M.F.F.). J.R.T. and V.L. are supported by a Ramón y Cajal contract from the Spanish Ministry of Science and Innovation (RYC2021-031799-I and RYC2022-035313-I respectively). A.R. and C.M. are supported by the IUOPA-Obra Social Cajastur-Liberbank.

Authors' contributions

A.R., J.R.T., A.F.F., and M.F.F. conceived the study and designed the experiments. J.R.T., T.D.V., A.S.S., O.G.B., D.F.M., and L.J.G.F. collected and processed patient samples. A.R., V.L., S.C., F.C., and I.O. established and implemented the EpiProfile protocol at the ISPA and CNB Proteomics facilities. A.R., J.R.T., V.L., R.G.U., C.M., D.B.G., and C.M.S. performed the experiments. J.R.T., J.K., M.M., M.J.C., and L.V. contributed to the implementation and optimization of the EpiProfile analysis platform. J.R.T. and D.Q.T. conducted the computational analyses and data interpretation. A.R., J.R.T., A.F.F., and M.F.F. wrote the manuscript. All authors reviewed and approved the final manuscript.

Acknowledgements

We would like to thank the members of IUOPA, ISPA and CINN, as well as former members of the Cancer Epigenetics Laboratory for their positive feedback and helpful discussions.

Authors' information

Annalisa Roberti and Juan Ramon Tejedor are joint first authors.

Agustin F. Fernandez and Mario F. Fraga are joint senior authors.

DATA AVAILABILITY

Raw RNA sequencing FASTQ files derived from patient's samples have been deposited in the European Genome-Phenome Archive under the accession number (EGAS50000001831). Raw RNA sequencing FASTQ files from DLD1 or SW1116 cells subjected to EHMT2 knockout have been uploaded to the ENA repository (PRJEB112401). Raw ChIP-seq files of H3K9me2, H3K14ac and H3K27me3 histone marks from DLD1 or SW1116 cells subjected to EHMT2 knockout have been uploaded to the ENA repository (PRJEB112868). Raw proteomic wiff files corresponding to the patient or colon cancer cell lines data have been deposited in the PRIDE repository (PDX077815 and PDX078016, respectively).

References

1. Morgan E, Arnold M, Gini A, Lorenzoni V, Cabasag CJ, Laversanne M, et al. Global burden of colorectal cancer in 2020 and 2040: incidence and mortality estimates from GLOBOCAN. Gut. 2023;72(2):338-44. doi:10.1136/gutjnl-2022-327736 PubMed PMID: 36604116.
2. Muzny DM, Bainbridge MN, Chang K, Dinh HH, Drummond JA, Fowler G, et al. Comprehensive molecular characterization of human colon and rectal cancer. Nature. 2012;487(7407):330-7. doi:10.1038/nature11252
3. Cornish AJ, Gruber AJ, Kinnersley B, Chubb D, Frangou A, Caravagna G, et al. The genomic landscape of 2,023 colorectal cancers. Nature. 2024;633(8028):127-36. doi:10.1038/s41586-024-07747-9 PubMed PMID: 39112709; PubMed Central PMCID: PMC11374690.
4. Guinney J, Dienstmann R, Wang X, de Reyniès A, Schlicker A, Soneson C, et al. The consensus molecular subtypes of colorectal cancer. Nat Med. 2015;21(11):1350-6. doi:10.1038/nm.3967 PubMed PMID: 26457759; PubMed Central PMCID: PMC4636487.
5. Vasaikar S, Huang C, Wang X, Petyuk VA, Savage SR, Wen B, et al. Proteogenomic Analysis of Human Colon Cancer Reveals New Therapeutic Opportunities. Cell. 2019;177(4):1035-1049.e19. doi:10.1016/j.cell.2019.03.030 PubMed PMID: 31031003; PubMed Central PMCID: PMC6768830.
6. Hinoue T, Weisenberger DJ, Lange CPE, Shen H, Byun HM, Van Den Berg D, et al. Genome-scale analysis of aberrant DNA methylation in colorectal cancer. Genome Res. 2012;22(2):271-82. doi:10.1101/gr.117523.110 PubMed PMID: 21659424; PubMed Central PMCID: PMC3266034.
7. Hahn MA, Li AX, Wu X, Yang R, Drew DA, Rosenberg DW, et al. Loss of the polycomb mark from bivalent promoters leads to activation of cancer-promoting genes in colorectal tumors. Cancer Res. 2014;74(13):3617-29. doi:10.1158/0008-5472.CAN-13-3147 PubMed PMID: 24786786; PubMed Central PMCID: PMC4079738.
8. Li Q, Geng S, Luo H, Wang W, Mo YQ, Luo Q, et al. Signaling pathways involved in colorectal cancer: pathogenesis and targeted therapy. Signal Transduct Target Ther. 2024;9(1):266. doi:10.1038/s41392-024-01953-7 PubMed PMID: 39370455; PubMed Central PMCID: PMC11456611.
9. Qin J, Wen B, Liang Y, Yu W, Li H. Histone Modifications and their Role in Colorectal Cancer (Review). Pathol Oncol Res. 2020;26(4):2023-33. doi:10.1007/s12253-019-00663-8 PubMed PMID: 31055775; PubMed Central PMCID: PMC7471167.
10. Noberini R, Robusti G, Vai A, Savoia EO, Jodice MG, Bertalot G, et al. A histone-centric multi-omics study shows that increased H3K4 methylation sustains triple-negative breast cancer phenotypes. Nat Commun. 2025;16(1):8716. doi:10.1038/s41467-025-63745-z PubMed PMID: 41028718; PubMed Central PMCID: PMC12484688.
11. Sidoli S, Bhanu NV, Karch KR, Wang X, Garcia BA. Complete Workflow for Analysis of Histone Post-translational Modifications Using Bottom-up Mass Spectrometry: From Histone Extraction to Data Analysis. J Vis Exp. 2016;(111):54112. doi:10.3791/54112 PubMed PMID: 27286567; PubMed Central PMCID: PMC4927705.

12. Yuan ZF, Sidoli S, Marchione DM, Simithy J, Janssen KA, Szurgot MR, et al. EpiProfile 2.0: A Computational Platform for Processing Epi-Proteomics Mass Spectrometry Data. J Proteome Res. 2018;17(7):2533-41. doi:10.1021/acs.jproteome.8b00133 PubMed PMID: 29790754; PubMed Central PMCID: PMC6387837.
13. Tang Z, Kang B, Li C, Chen T, Zhang Z. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. Nucleic Acids Res. 2019;47(W1):W556-60. doi:10.1093/nar/gkz430 PubMed PMID: 31114875; PubMed Central PMCID: PMC6602440.
14. Sarg B, Koutzamani E, Helliger W, Rundquist I, Lindner HH. Postsynthetic trimethylation of histone H4 at lysine 20 in mammalian tissues is associated with aging. J Biol Chem. 2002;277(42):39195-201. doi:10.1074/jbc.M205166200 PubMed PMID: 12154089.
15. Hicks SC, Okrah K, Paulson JN, Quackenbush J, Irizarry RA, Bravo HC. Smooth quantile normalization. Biostatistics. 2018;19(2):185-98. doi:10.1093/biostatistics/kxx028 PubMed PMID: 29036413; PubMed Central PMCID: PMC5862355.
16. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. Bioinformatics. 2008;24(13):1547-8. doi:10.1093/bioinformatics/btn224 PubMed PMID: 18467348.
17. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47. doi:10.1093/nar/gkv007 PubMed PMID: 25605792; PubMed Central PMCID: PMC4402510.
18. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34(17):i884-90. doi:10.1093/bioinformatics/bty560 PubMed PMID: 30423086; PubMed Central PMCID: PMC6129281.
19. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12:323. doi:10.1186/1471-2105-12-323 PubMed PMID: 21816040; PubMed Central PMCID: PMC3163565.
20. Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Res. 2015;4:1521. doi:10.12688/f1000research.7563.2 PubMed PMID: 26925227; PubMed Central PMCID: PMC4712774.
21. Eide PW, Bruun J, Lothe RA, Sveen A. CMScaller: an R package for consensus molecular subtyping of colorectal cancer pre-clinical models. Sci Rep. 2017;7(1):16618. doi:10.1038/s41598-017-16747-x PubMed PMID: 29192179; PubMed Central PMCID: PMC5709354.
22. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550. doi:10.1186/s13059-014-0550-8 PubMed PMID: 25516281; PubMed Central PMCID: PMC4302049.
23. Santamarina-Ojeda P, Tejedor JR, Pérez RF, López V, Roberti A, Mangas C, et al. Multi-omic integration of DNA methylation and gene expression data reveals molecular vulnerabilities in glioblastoma. Mol Oncol. 2023;17(9):1726-43. doi:10.1002/1878-0261.13479 PubMed PMID: 37357610; PubMed Central PMCID: PMC10483606.

24. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102(43):15545-50. doi:10.1073/pnas.0506580102 PubMed PMID: 16199517; PubMed Central PMCID: PMC1239896.
25. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 2015;1(6):417-25. doi:10.1016/j.cels.2015.12.004 PubMed PMID: 26771021; PubMed Central PMCID: PMC4707969.
26. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics. 2009;10:48. doi:10.1186/1471-2105-10-48 PubMed PMID: 19192299; PubMed Central PMCID: PMC2644678.
27. Supek F, Bošnjak M, Škunca N, Šmuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS One. 2011;6(7):e21800. doi:10.1371/journal.pone.0021800 PubMed PMID: 21789182; PubMed Central PMCID: PMC3138752.
28. Mair P, Wilcox R. Robust statistical methods in R using the WRS2 package. Behav Res Methods. 2020;52(2):464-88. doi:10.3758/s13428-019-01246-w PubMed PMID: 31152384.
29. Yedier-Bayram O, Gokbayrak B, Kayabolen A, Aksu AC, Cavga AD, Cingöz A, et al. EPIKOL, a chromatin-focused CRISPR/Cas9-based screening platform, to identify cancer-specific epigenetic vulnerabilities. Cell Death Dis. 2022;13(8):710. doi:10.1038/s41419-022-05146-4 PubMed PMID: 35973998; PubMed Central PMCID: PMC9381743.
30. Liu H, Wei Z, Dominguez A, Li Y, Wang X, Qi LS. CRISPR-ERA: a comprehensive design tool for CRISPR-mediated gene editing, repression and activation. Bioinformatics. 2015;31(22):3676-8. doi:10.1093/bioinformatics/btv423
31. Ewels PA, Peltzer A, Fillinger S, Patel H, Alneberg J, Wilm A, et al. The nf-core framework for community-curated bioinformatics pipelines. Nat Biotechnol. 2020;38(3):276-8. doi:10.1038/s41587-020-0439-x PubMed PMID: 32055031.
32. Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. 2014;42(Web Server issue):W187-191. doi:10.1093/nar/gku365 PubMed PMID: 24799436; PubMed Central PMCID: PMC4086134.
33. Stovner EB, Sætrom P. epic2 efficiently finds diffuse domains in ChIP-seq data. Bioinformatics. 2019;35(21):4392-3. doi:10.1093/bioinformatics/btz232 PubMed PMID: 30923821.
34. Leemans C, van der Zwalm MCH, Brueckner L, Comoglio F, van Schaik T, Pagie L, et al. Promoter-Intrinsic and Local Chromatin Features Determine Gene Repression in LADs. Cell. 2019;177(4):852-864.e14. doi:10.1016/j.cell.2019.03.009 PubMed PMID: 30982597; PubMed Central PMCID: PMC6506275.
35. Luca BA, Steen CB, Matusiak M, Azizi A, Varma S, Zhu C, et al. Atlas of clinically distinct cell states and ecosystems across human solid tumors. Cell. 2021;184(21):5482-5496.e28. doi:10.1016/j.cell.2021.09.014 PubMed PMID: 34597583; PubMed Central PMCID: PMC8526411.

36. Jung G, Hernández-Illán E, Moreira L, Balaguer F, Goel A. Epigenetics of colorectal cancer: biomarker and therapeutic potential. Nat Rev Gastroenterol Hepatol. 2020;17(2):111-30. doi:10.1038/s41575-019-0230-y PubMed PMID: 31900466; PubMed Central PMCID: PMC7228650.
37. Fraga MF, Ballestar E, Villar-Garea A, Boix-Chornet M, Espada J, Schotta G, et al. Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. Nat Genet. 2005;37(4):391-400. doi:10.1038/ng1531 PubMed PMID: 15765097.
38. Benard A, Goossens-Beumer IJ, van Hoesel AQ, de Graaf W, Horati H, Putter H, et al. Histone trimethylation at H3K4, H3K9 and H4K20 correlates with patient survival and tumor recurrence in early-stage colon cancer. BMC Cancer. 2014;14:531. doi:10.1186/1471-2407-14-531 PubMed PMID: 25047223; PubMed Central PMCID: PMC4223547.
39. Wang CG, Ye YJ, Yuan J, Liu FF, Zhang H, Wang S. EZH2 and STAT6 expression profiles are correlated with colorectal cancer stage and prognosis. World J Gastroenterol. 2010;16(19):2421-7. doi:10.3748/wjg.v16.i19.2421 PubMed PMID: 20480530; PubMed Central PMCID: PMC2874149.
40. Young NL, DiMaggio PA, Plazas-Mayorca MD, Baliban RC, Floudas CA, Garcia BA. High throughput characterization of combinatorial histone codes. Mol Cell Proteomics. 2009;8(10):2266-84. doi:10.1074/mcp.M900238-MCP200 PubMed PMID: 19654425; PubMed Central PMCID: PMC2758755.
41. Nicetto D, Zaret KS. Role of H3K9me3 heterochromatin in cell identity establishment and maintenance. Current Opinion in Genetics & Development. 2019;Genome Architecture and Expression55:1-10. doi:10.1016/j.gde.2019.04.013
42. Price AJ, Manjegowda MC, Kain J, Anandh S, Bochkis IM. Hdac3, Setdb1, and Kap1 mark H3K9me3/H3K14ac bivalent regions in young and aged liver. Aging Cell. 2020;19(2):e13092. doi:10.1111/acel.13092
43. Jurkowska RZ, Qin S, Kungulovski G, Tempel W, Liu Y, Bashtrykov P, et al. H3K14ac is linked to methylation of H3K9 by the triple Tudor domain of SETDB1. Nat Commun. 2017;8(1):2057. doi:10.1038/s41467-017-02259-9
44. Chandrasekaran TT, Choudalakis M, Bröhm A, Weirich S, Kouroukli AG, Ammerpohl O, et al. SETDB1 activity is globally directed by H3K14 acetylation via its Triple Tudor Domain. Nucleic Acids Res. 2024;52(22):13690-705. doi:10.1093/nar/gkae1053 PubMed PMID: 39540436; PubMed Central PMCID: PMC11662677.
45. Sridharan R, Gonzales-Cope M, Chronis C, Bonora G, McKee R, Huang C, et al. Proteomic and genomic approaches reveal critical functions of H3K9 methylation and Heterochromatin Protein-1γ in reprogramming to pluripotency. Nat Cell Biol. 2013;15(7):872-82. doi:10.1038/ncb2768 PubMed PMID: 23748610; PubMed Central PMCID: PMC3733997.
46. Bergin CJ, Zouggar A, Haebe JR, Masibag AN, Desrochers FM, Reilley SY, et al. G9a controls pluripotent-like identity and tumor-initiating function in human colorectal cancer. Oncogene. 2021;40(6):1191-202. doi:10.1038/s41388-020-01591-7

47. Wang H, Cui L, Li D, Fan M, Liu Z, Liu C, et al. Overexpression of PSAT1 regulated by G9A sustains cell proliferation in colorectal cancer. Signal Transduct Target Ther. 2020;5:47. doi:10.1038/s41392-020-0147-5 PubMed PMID: 32300099; PubMed Central PMCID: PMC7162942.
48. Plazas-Mayorca MD, Bloom JS, Zeissler U, Leroy G, Young NL, DiMaggio PA, et al. Quantitative proteomics reveals direct and indirect alterations in the histone code following methyltransferase knockdown. Mol Biosyst. 2010;6(9):1719-29. doi:10.1039/c003307c PubMed PMID: 20577673; PubMed Central PMCID: PMC3321735.
49. Jones P, Altamura S, De Francesco R, Gallinari P, Lahm A, Neddermann P, et al. Probing the elusive catalytic activity of vertebrate class IIa histone deacetylases. Bioorg Med Chem Lett. 2008;18(6):1814-9. doi:10.1016/j.bmcl.2008.02.025 PubMed PMID: 18308563.
50. Clocchiatti A, Florean C, Brancolini C. Class IIa HDACs: from important roles in differentiation to possible implications in tumourigenesis. J Cell Mol Med. 2011;15(9):1833-46. doi:10.1111/j.1582-4934.2011.01321.x PubMed PMID: 21435179; PubMed Central PMCID: PMC3918040.
51. Zhang Y, Andrade R, Hanna AA, Pflum MKH. Evidence that HDAC7 acts as an epigenetic «reader» of AR acetylation through NCoR-HDAC3 dissociation. Cell Chem Biol. 2022;29(7):1162-1173.e5. doi:10.1016/j.chembiol.2022.05.008 PubMed PMID: 35709754; PubMed Central PMCID: PMC9450512.
52. Seto E, Yoshida M. Erasers of Histone Acetylation: The Histone Deacetylase Enzymes. Cold Spring Harb Perspect Biol. 2014;6(4):a018713. doi:10.1101/cshperspect.a018713 PubMed PMID: 24691964; PubMed Central PMCID: PMC3970420.
53. Yan Z, Ji L, Huo X, Wang Q, Zhang Y, Wen B. G9a/GLP-Sensitivity of H3K9me2 Demarcates Two Types of Genomic Compartments. genom proteom bioinform. 2020;18(4):359-70. doi:10.1016/j.gpb.2020.08.001
54. Stutzman AV, Hill CA, Armstrong RL, Gohil R, Duronio RJ, Dowen JM, et al. Heterochromatic 3D genome organization is directed by HP1a- and H3K9-dependent and independent mechanisms. Molecular Cell. 2024;84(11):2017-2035.e6. doi:10.1016/j.molcel.2024.05.002
55. Smith CL, Poleshko A, Epstein JA. The nuclear periphery is a scaffold for tissue-specific enhancers. Nucleic Acids Res. 2021;49(11):6181-95. doi:10.1093/nar/gkab392 PubMed PMID: 34023908; PubMed Central PMCID: PMC8216274.
56. Wen B, Wu H, Shinkai Y, Irizarry RA, Feinberg AP. Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. Nat Genet. 2009;41(2):246-50. doi:10.1038/ng.297 PubMed PMID: 19151716; PubMed Central PMCID: PMC2632725.
57. van Steensel B, Belmont AS. Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. Cell. 2017;169(5):780-91. doi:10.1016/j.cell.2017.04.022 PubMed PMID: 28525751; PubMed Central PMCID: PMC5532494.
58. Faubert B, Solmonson A, DeBerardinis RJ. Metabolic reprogramming and cancer progression. Science. 2020;368(6487):eaaw5473. doi:10.1126/science.aaw5473 PubMed PMID: 32273439; PubMed Central PMCID: PMC7227780.

59. Wang X, Chang X, He C, Fan Z, Yu Z, Yu B, et al. ATP5B promotes the metastasis and growth of gastric cancer by activating the FAK/AKT/MMP2 pathway. FASEB J. 2021;35(4):e20649. doi:10.1096/fj.202000608R PubMed PMID: 33715234.
60. Sharma RK, Chafik A, Bertolin G. Aurora kinase A/AURKA functionally interacts with the mitochondrial ATP synthase to regulate energy metabolism and cell death. Cell Death Discov. 2023;9(1):203. doi:10.1038/s41420-023-01501-2
61. Lu Z jun, Song Q fang, Jiang S sa, Song Q, Wang W, Zhang G hua, et al. Identification of ATP synthase beta subunit (ATPB) on the cell surface as a non-small cell lung cancer (NSCLC) associated antigen. BMC Cancer. 2009;9:16. doi:10.1186/1471-2407-9-16 PubMed PMID: 19144153; PubMed Central PMCID: PMC2654462.

Figures

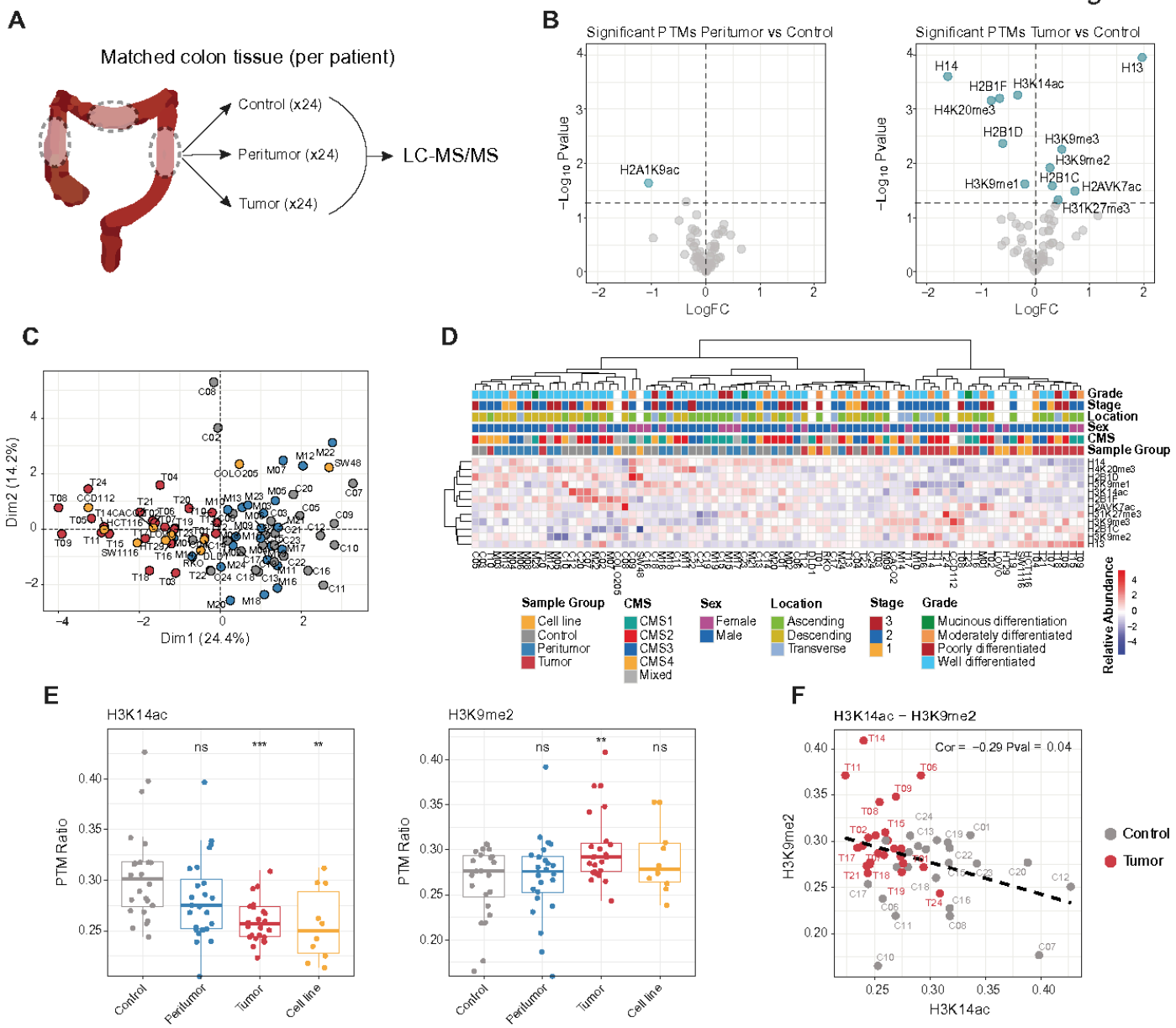

Figure 1

Identification of global histone posttranslational modifications in colon adenocarcinoma. A) *Schematic overview of patient cohort and sample types included in the study.* ***B)****Scatter plot showing the fold change and the statistical significance of the HPTMs identified in the context of peritumor (left) or tumor (right) versus control samples. Significantly altered HPTMs (p.adj < 0.05) are shown in blue.* ***C)*** *PCA based on the significant HPTMs between tumor and control samples from the local cohort, including matched peritumor tissue and 10 CRC cell lines. The percentage of variance explained by the first two principal components, is indicated in the corresponding axes.* ***D)****Heatmap representing the relative ratio of the significant HPTMs included in the above comparisons. The color scale represents the relative abundance of the indicated HPTMs. Relevant clinical information for each sample is indicated in the legend.* ***E)*** *Boxplot showing the*

*observed ratio of the indicated HPTMs (H3K14ac and H3K9me2) across the samples included in this study. Statistical significance between each condition and its corresponding normal colon mucosa was assessed using limma analysis (*p < 0.05; **p < 0.01; ***p < 0.001; ns, not significant).* ***F)*** *Scatter plot illustrating the robust correlation score between the histone marks H3K14ac and H3K9me2 in the context of tumor and*

control samples.

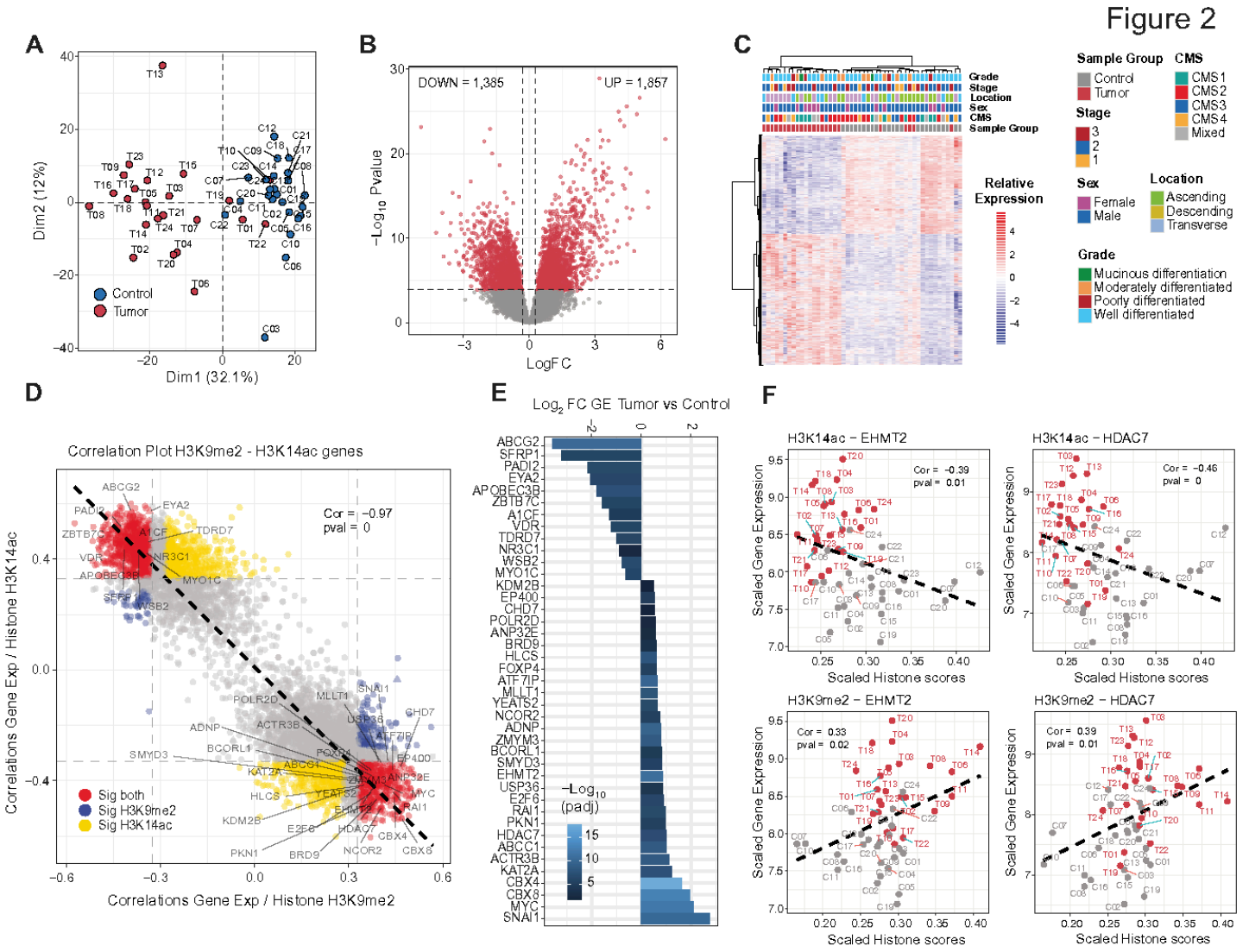

Figure 2

Identification of candidate genes involved in H3K14ac / H3K9me2 imbalance in colon adenocarcinoma. A) *PCA generated using the 1,000 most variable genes expressed in a cohort of 24 paired control-tumor samples from the local COAD dataset. The percentage of variance explained for the first two dimensions is indicated in the corresponding axes.* ***B)*** *Scatterplot showing the number of up- or downregulated genes identified between paired control and tumor samples from the local COAD cohort. The total number of significant up- and downregulated genes is indicated (p.adj < 0.001).* ***C)*** *Heatmap illustrating the*

hierarchical clustering of samples using the differentially expressed genes (DEGs) identified in the local COAD dataset. Color scale represents the relative expression of the genes displayed. Additional clinical information is indicated in the legend. ***D)*** *Scatter plot illustrating pairwise robust correlation scores between gene expression and H3K9me2 (x-axis) or H3K14ac (y-axis) levels for all the significant DEGs identified in previous analyses. Colored dots indicate genes with significant correlations ($p < 0.05$) with both H3K9me2 and H3K14ac (red), only H3K9me2 (blue), or only H3K14ac (yellow). The name of significant, epigenetic-related genes is indicated.* ***E)****Barplot representing the log2 fold changes in gene expression between paired tumor and control tissue of the epigenetic DEGs highly correlated with H3K9me2 and H3K14ac levels in the local COAD cohort. Color indicates the statistical significance (adj.p.value) obtained from DeSeq2 comparisons.* ***F)*** *Scatter plots illustrating the robust correlation score between the indicated histone marks (H3K14ac and H3K9me2, x-axis) and EHMT2 or HDAC7 gene expression levels (y-axis) in the context of tumor and control*
samples from the local COAD cohort.

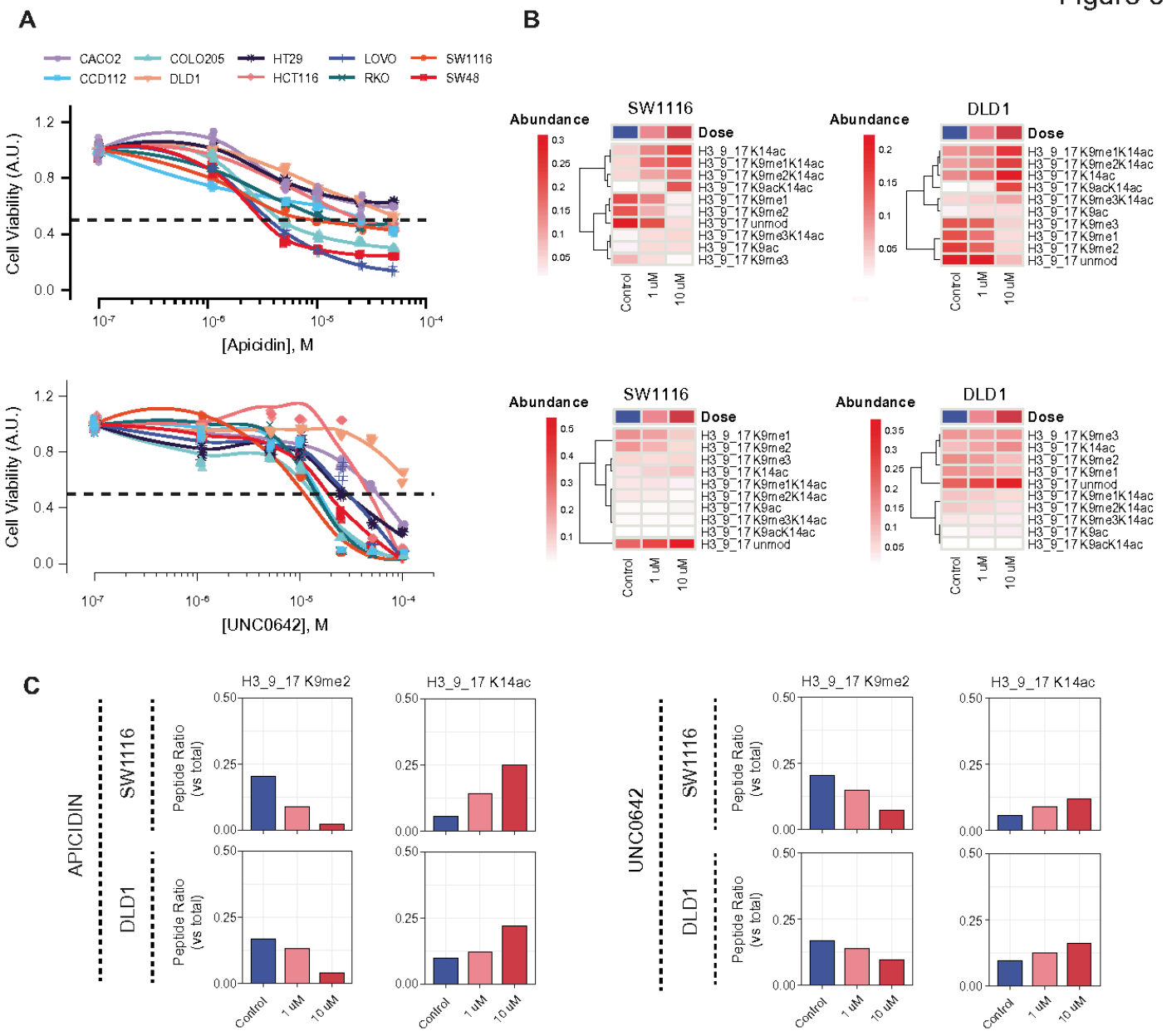

Figure 3

Pharmacological modulation of H3K9me2 and H3K14ac affects the viability of colon adenocarcinoma cell lines. A) *Line plots illustrating the reduced cell viability in 10 colon adenocarcinoma cell lines after 72 h incubation with increasing doses of the HDAC inhibitor Apicidin (top) or the EHMT2 inhibitor UNC0642 (bottom).* ***B)*** *Heatmaps showing relative abundance ratios of the indicated H3_9_17 peptide modifications following 72 h treatment with the indicated concentrations of Apicidin (top) or UNC0642 (bottom) in SW1116 (left) and DLD1 (right) cell lines.* ***C)*** *Bar plot showing relative abundance ratios of H3_9_17 K9me2 and H3_9_17 K14ac peptides in the experiments described above. A dose-dependent effect is observed for both treatment conditions in the cell lines tested.*

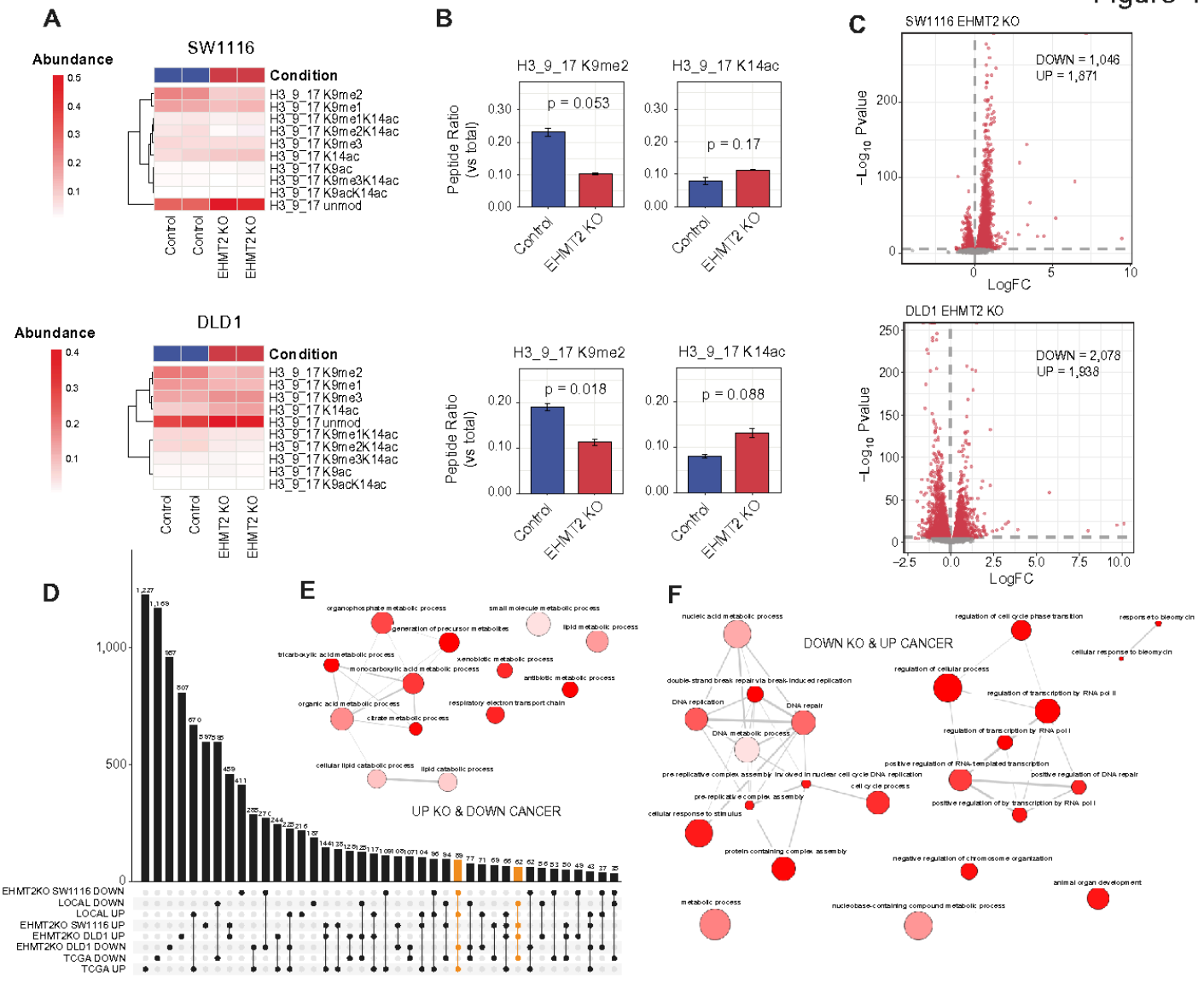

Figure 4

Genetic ablation of EHMT2 modulates H3K14ac / H3K9me2 ratio in SW1116 and DLD1 colon adenocarcinoma cell lines. A) *Heatmaps showing relative abundance ratios of the indicated peptide modifications included in the peptide H3_9_17 following CRISPR/Cas9-mediated EHMT2 knockout in SW1116 (top) and DLD1 (bottom) cells.* ***B)****Bar plot showing relative abundance ratios of H3_9_17 K9me2 and H3_9_17 K14ac peptides. Data represent mean ± standard deviation of biological replicates.* ***C)*** *Scatter plots showing the number of up- and downregulated genes following EHMT2 knockout in SW1116 (top) and DLD1 (bottom) cells (n = 4 biological replicates per condition). The total number of significant up- and downregulated genes is indicated (p.adj < 0.0001).* ***D)*** *UpSetR plot showing the total number of DEGs (left) and their potential overlaps –main plot– between control and tumor samples from the local COAD cohort, the TCGA COAD cohort and the genes altered upon EHMT2 KO in SW1116 and DLD1 cell lines. Genes*

exhibiting opposite expression trends between patient samples and knockout models are highlighted in yellow. ***E)*** *Gene ontology network analysis derived from genes upregulated upon EHMT2 knockout and downregulated in colon cancer patients relative to paired normal tissue.* ***F)*** *Same as E, but for genes*

downregulated upon EHMT2 knockout and upregulated in colon cancer patients.

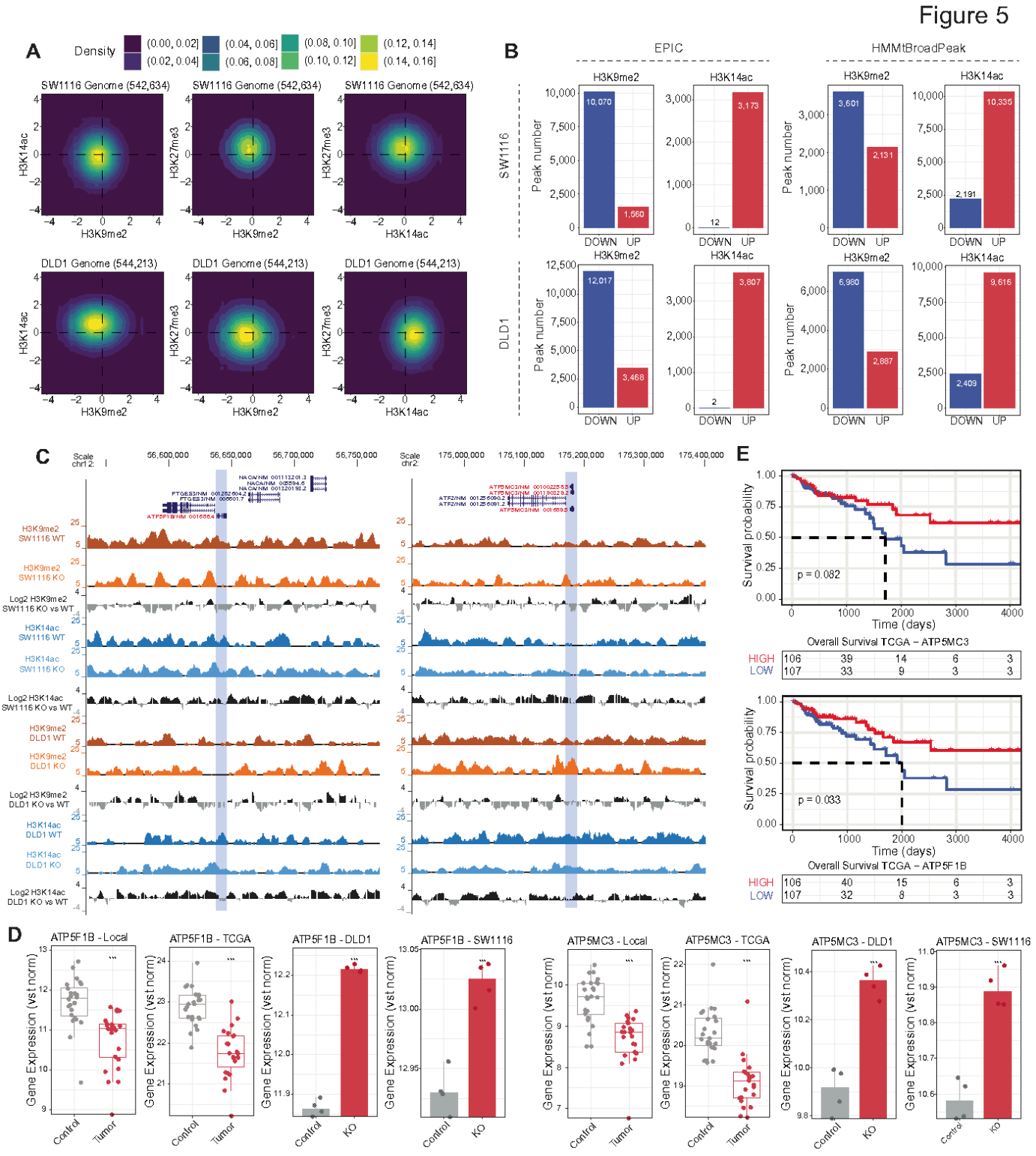

Figure 5

Correlation between H3K9me2 and H3K14ac levels in colon cancer cell lines following genetic modulation of EHMT2. A) *Density plots indicating the genome-wide, pairwise relationships between H3K9me2-H3K14ac (left), H3K9me2-H3K27me3 (middle) or H3K14ac-H3K27me3 (right) signals in the context of SW1116 (Top) or DLD1 cells (Bottom). Each axis represents the log2 fold change of the CUT&RUN signal per genomic bin observed between EHMT2 KO and control conditions.* ***B)*** *Barplot illustrating the number of significant up or downregulated peaks between EHMT2 KO and control conditions in the context of SW1116 (Top) or DLD1 (Bottom) cell lines. Two peak-calling algorithms (EPIC – left; HMMtBroadPeak – right) were used.* ***C)*** *UCSC Genome Browser tracks representing the CUT&RUN profiles of H3K9me2 or H3K14ac in the tested cell lines at the ATP5MC3 (left) and ATP5F1B (right) loci. The log2 fold change between EHMT2 KO and control conditions is indicated for each experiment.* ***D)****Boxplots and barplots depicting expression levels of ATP5F1B and ATP5MC3 in patient samples (local and TCGA cohorts) and in SW1116 and DLD1 cells upon EHMT2 knockout. Statistical significance was calculated using limma analysis (** = p < 0.01; *** = p < 0.001).* ***E)*** *Kaplan-Meier curves showing overall survival for patients with low or high expression of ATP5MC3 (top) or ATP5F1B (bottom) in the TCGA cohort. P-value refers to differences in event rates between Kaplan-Meier curves and was calculated with the log-rank test.*

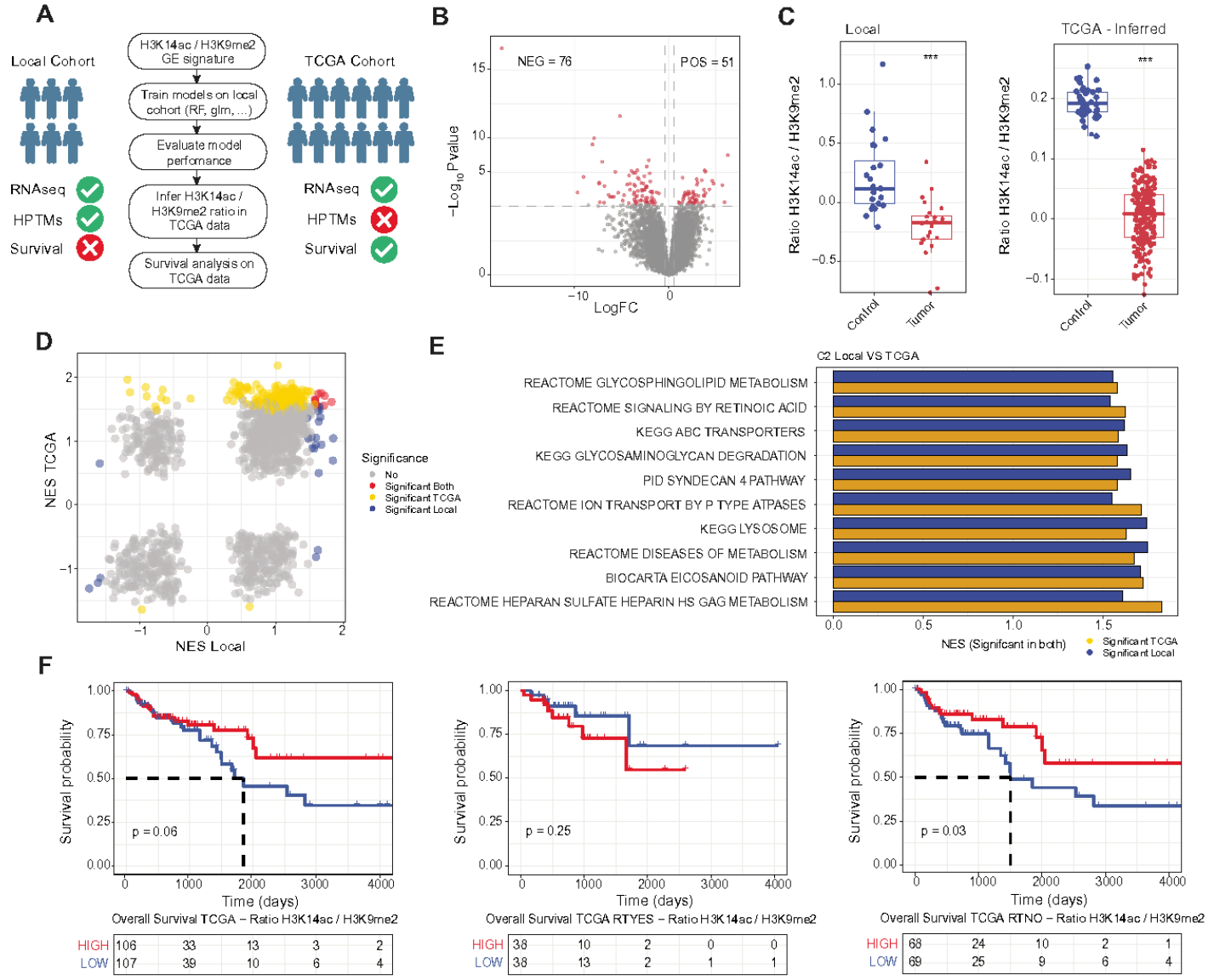

Figure 6

Prognostic potential of the H3K14ac / H3K9me2 signature in colon cancer patients. A)*Schematic depicting the protocol used to infer H3K14ac / H3K9me2 ratio in TCGA COAD samples based on data from our local cohort.* ***B)*** *Scatterplot showing the number of positive or negative correlated gene expression changes with the H3K14ac / H3K9me2 ratio observed in control and tumor samples from the local COAD cohort. Significant up- and downregulated genes are indicated (p.adj < 0.001). The total number of significant up and downregulated genes is indicated (p.adj < 0.001).* ***C)*** *Boxplots illustrating the H3K14ac / H3K9me2 ratio signals in the local cohort (left) or the inferred values in the TCGA COAD cohort (right). Statistical significance was calculated using the Welch's t-test (*** = p < 0.001).* ***D)****Scatterplot showing the GSEA normalized enrichment scores (NES) in canonical pathways (C2) for the patient's data associated to the H3K14ac / H3K9me2 ratio (treated as continuous variable) in the context of the local cohort (x-axis) or*

the TCGA cohort (y-axis). Colored dots indicate significance in both cohorts (red), local only (blue), or TCGA only (yellow) (p.val < 0.05). ***E)****Barplot showing NES scores for significantly enriched pathways in both cohorts (MSigDB C2 – canonical pathways).* ***F)*** *Kaplan-Meier curves illustrating overall survival of patients with low or high inferred H3K14ac / H3K9me2 ratio in TCGA cohort. The left panel includes all the samples selected for these comparisons; the middle panel includes only samples that received radiotherapy treatment and the right panel includes only samples that were not treated with radiotherapy. P-value refers to differences in event rates between Kaplan-Meier curves and was calculated with the log-rank test.*

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- FigureS1.pdf
- FigureS2.pdf
- FigureS3.pdf
- FigureS4.pdf
- FigureS5.pdf
- FigureS6.pdf
- FigureS7.pdf
- FigureS8.pdf
- FigureS9.pdf
- FigureS10.pdf
- TableS1.xlsx
- TableS2.xlsx
- TableS3.xlsx
- TableS4.xlsx
- TableS5.xlsx
- TableS6.xlsx
- TableS7.xlsx
- TableS8.xlsx
- TableS9.xlsx
- TableS10.xlsx
- TableS11.xlsx
- TableS12.xlsx
- TableS13.xlsx
- TableS14.xlsx

- SupplementaryInformation.docx