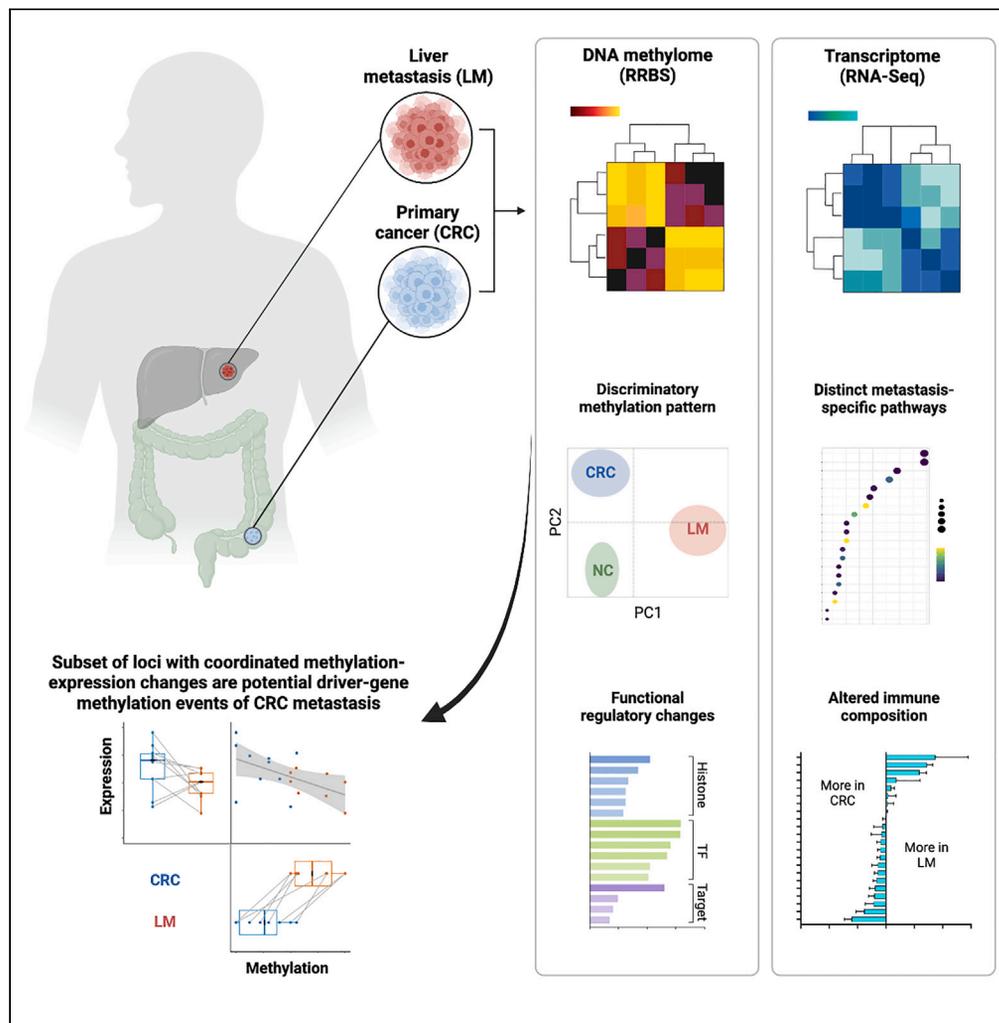


Article

# An epigenetic signature of advanced colorectal cancer metastasis



Euan J. Rodger,  
Gregory Gimenez,  
Priyadarshana  
Ajithkumar, ...,  
Michael R. Eccles,  
Rachel V. Purcell,  
Aniruddha  
Chatterjee

euan.rodger@otago.ac.nz  
(E.J.R.)  
rachel.purcell@otago.ac.nz  
(R.V.P.)  
aniruddha.chatterjee@otago.  
ac.nz (A.C.)

**Highlights**

Epigenetic signature  
discriminates non-  
malignant, primary CRC  
and liver metastasis

Methylomes provide  
better separation of  
advanced CRC metastasis  
than transcriptomes

Altered CRC metastasis  
methylation associated  
with functional regulatory  
changes

Potential driver gene-  
methylation events of CRC  
metastasis identified

Rodger et al., iScience 26,  
106986  
June 16, 2023 © 2023 The  
Authors.  
[https://doi.org/10.1016/  
j.isci.2023.106986](https://doi.org/10.1016/j.isci.2023.106986)



## Article

## An epigenetic signature of advanced colorectal cancer metastasis

Euan J. Rodger,<sup>1,10,\*</sup> Gregory Gimenez,<sup>1</sup> Priyadarshana Ajithkumar,<sup>1</sup> Peter A. Stockwell,<sup>1</sup> Suzan Almomani,<sup>1</sup> Sarah A. Bowden,<sup>1</sup> Anna L. Leichter,<sup>1</sup> Antonio Ahn,<sup>2,3</sup> Sharon Pattison,<sup>4</sup> John L. McCall,<sup>5</sup> Sebastian Schmeier,<sup>6</sup> Frank A. Frizelle,<sup>7</sup> Michael R. Eccles,<sup>1,9</sup> Rachel V. Purcell,<sup>7,9,11,\*</sup> and Aniruddha Chatterjee<sup>1,8,9,12,13,\*</sup>

## SUMMARY

**Colorectal cancer (CRC) is a leading cause of morbidity and mortality worldwide. The majority of CRC deaths are caused by tumor metastasis, even following treatment. There is strong evidence for epigenetic changes, such as DNA methylation, accompanying CRC metastasis and poorer patient survival. Earlier detection and a better understanding of molecular drivers for CRC metastasis are of critical clinical importance. Here, we identify a signature of advanced CRC metastasis by performing whole genome-scale DNA methylation and full transcriptome analyses of paired primary cancers and liver metastases from CRC patients. We observed striking methylation differences between primary and metastatic pairs. A subset of loci showed coordinated methylation-expression changes, suggesting these are potentially epigenetic drivers that control the expression of critical genes in the metastatic cascade. The identification of CRC epigenomic markers of metastasis has the potential to enable better outcome prediction and lead to the discovery of new therapeutic targets.**

## INTRODUCTION

Colorectal cancer (CRC) is a leading cause of morbidity and mortality worldwide. Globally, more than 1.4 million individuals develop CRC and 0.6 million people die from CRC every year.<sup>1,2</sup> Alarming, the global incidence of CRC has consistently risen at an annual rate of 3.2% over the last two decades and is predicted to continue to increase in the future.<sup>3</sup> Furthermore, there is a particular trend of increasing rates of CRC in younger people (<50 years of age).<sup>4</sup> CRC-related death occurs mainly because of metastasis of primary CRC (>90% cases), even after treatment,<sup>5</sup> and patients with distant metastases have a 5-year survival rate of only ~12%.<sup>6</sup> The liver is the most common organ of distant metastasis in CRC – during the course of the disease ~50% of CRC patients will develop liver metastasis and in ~30% of patients with metastatic CRC, the liver is the only organ with metastases.<sup>7,8</sup> Although significant efforts have been made to improve outcomes, CRC patients with distant metastasis who receive chemotherapy (such as fluoropyrimidines and oxaliplatin chemotherapy), eventually develop resistance.<sup>9</sup> Earlier diagnosis or detection of metastatic potential would allow for earlier intervention and possibly improved prognosis. For example, adjuvant chemotherapy could be of significant benefit for early-stage patients with an increased risk of metastasis.<sup>10</sup> However, improved risk stratification is necessary to identify early-stage patients who would benefit most from this approach.<sup>11</sup> These observations clearly indicate that early detection and a better understanding of molecular drivers for CRC metastasis are of critical clinical importance.

Metastasis of CRC is a multi-step process and tumor cells require several different molecular or phenotypic changes to successfully complete the sequence of events.<sup>12,13</sup> Remarkable progress has been made over the last few decades in understanding the genetic basis of primary cancer formation. However, even after extensive cancer genomic analyses, examples of causal genetic events for cancer metastasis are still limited.<sup>12,14</sup> Very recently, non-mutational epigenetic reprogramming has been recognized as a key mechanistic determinant that enables tumor cells to achieve cancer hallmark capabilities.<sup>15</sup> The key role of epigenetic mechanisms, such as DNA methylation, in driving metastatic progression is becoming established.<sup>16–19</sup> Considering the dynamic nature of epigenetic changes, it is highly plausible that epigenetic events play a driver role in CRC metastasis. Indeed, recent studies have provided strong evidence for epigenetic changes accompanying CRC metastasis and poorer patient survival.<sup>20–22</sup> Analysis of

<sup>1</sup>Department of Pathology, Dunedin School of Medicine, University of Otago, Dunedin, New Zealand

<sup>2</sup>Peter MacCallum Cancer Centre, Melbourne, VIC, Australia

<sup>3</sup>The Sir Peter MacCallum Department of Oncology, Faculty of Medicine, Dentistry and Health Sciences, University of Melbourne, Melbourne, VIC, Australia

<sup>4</sup>Department of Medicine, Dunedin School of Medicine, University of Otago, Dunedin, New Zealand

<sup>5</sup>Department of Surgical Sciences, Dunedin School of Medicine, University of Otago, Dunedin, New Zealand

<sup>6</sup>Evotec SE, Hamburg, Germany

<sup>7</sup>Department of Surgery, University of Otago, Christchurch, New Zealand

<sup>8</sup>Honorary Professor, School of Health Sciences and Technology, UPES University, India

<sup>9</sup>Senior authors

<sup>10</sup>Twitter: @euanjrodger

<sup>11</sup>Twitter: @RachelVPurcell

<sup>12</sup>Twitter: @futureliesindha

<sup>13</sup>Lead contact

\*Correspondence: euan.rodger@otago.ac.nz (E.J.R.), rachel.purcell@otago.ac.nz (R.V.P.), aniruddha.chatterjee@otago.ac.nz (A.C.)

<https://doi.org/10.1016/j.isci.2023.106986>



methylation and expression profiles of primary and metastatic CRCs at a single cell level revealed that methylation changes control expression of critical genes during metastasis.<sup>23</sup> In addition, it is now well established that a subgroup of patients with CRC exhibits a particular methylation signature (referred to as CpG island methylation phenotype or CIMP), which possibly affects patient prognosis and response to specific therapies.<sup>24,25</sup> Furthermore, pre-clinical data has demonstrated that DNA methylation-modifying drugs can reduce the proliferation of CRC-initiating cells.<sup>26</sup> This evidence indicates that DNA methylation is a potential key regulator of CRC growth and also suggests the possibility of targeting methylation levels as a potential therapy.

One major issue in identifying epigenetic drivers of CRC metastasis is the extensive molecular heterogeneity exhibited in tumor cells. In this work, we performed whole-genome scale methylation and expression analysis in a well-defined set of paired patient samples (primary CRC and liver metastases). From our analyses, we aimed to identify functionally relevant epigenetic changes that occurred between primary cancers and liver metastases from the same patient, and to identify an epigenetic signature of advanced CRC metastasis. Furthermore, we utilized epigenetic data from normal colon to document progressive changes in potential epigenetic drivers of metastasis and we have validated our results and reproducibility with large additional cohorts of CRC. The development of a robust molecular signature that confers a high risk of metastasis could potentially be used to triage stage II and III patients who would benefit more significantly from adjuvant therapy. Furthermore, and possibly of greater clinical value, would be the development of new effective therapies that specifically target molecular drivers of metastasis.<sup>27</sup>

## RESULTS

### Whole-genome scale DNA methylation profiles are conserved between primary CRC cancers and liver metastases but differ from normal colon

To investigate genome-scale DNA methylation changes of CRC metastasis, we performed reduced representation bisulfite sequencing (RRBS) on primary colonic cancers (CRC,  $n = 10$ , all adenocarcinomas) and matched liver metastases ( $n = 10$ , LM) from the same patients. These samples represented a diverse cohort, including a mixture of different primary tumor sites (left- and right-sided colon and rectum), and three patients received adjuvant chemotherapy before LM resection (Tables 1 and S1). In total, we obtained  $1.4 \times 10^9$  sequence reads for the 20 cancer tissues, allowing for comprehensive genome-wide analysis (Table S2). For comparison, we analyzed our methylation maps alongside independent RRBS data of normal colon tissue ( $n = 10$ ).<sup>28</sup> Global methylome analysis revealed that genomic methylation levels were conserved between primary and matched metastatic tumors (median methylation = 53.37% and 53.32% for CRC and LM respectively), but increased compared to normal colon (median = 49.06%, Figure 1A and Table S3). Most genomic, regulatory and repeat elements had small methylation differences (<4%) in CRC and LM compared to normal colon, except for exons, introns and telomeres, which had increased median methylation of 16.75%, 5.51% and 4.79% respectively (Figures 1A–1F and Table S3). Compared to CRC, most elements had only small methylation differences (<2%) in LM, with the exception of exons, which were 4.16% more methylated (Figure 1A and Table S3).

### Genome-wide analysis identifies late-stage, liver metastasis-specific methylation changes compared to matched primary cancer and normal colon

Locus-specific methylation analysis identified 244 differentially methylated CpGs (DMCs,  $P$ -value  $< 5.0 \times 10^{-4}$  with >20% methylation difference), of which 68 were hypomethylated and 176 were hypermethylated in LM compared to CRC (Figure 2A). Almost two-thirds of the DMCs showed large differences in methylation levels (>40%) and 17 CpGs had methylation differences >70% between CRC and LM (Figure S1A). We also observed that 20 of the DMCs were also differentially methylated between normal colon and CRC, and 44 were also differentially methylated between normal colon and LM. Approximately 50% of the DMCs within these two additional comparisons showed progressive changes in methylation from normal to CRC to LM. Therefore, the majority of identified DMCs were either metastasis-specific or were further altered in metastasis. Principal component analysis (PCA) of the top 1% CpG sites that varied most across all samples showed only segregation between tumor tissues (CRC, LM) and normal colon (Figure S1B). However, PCA of the DMCs showed distinct methylation patterns for each group. LM methylation was separated from CRC and normal colon in dimension 1, which accounted for 53.1% of the variance (Figure 2B). Unsupervised hierarchical clustering of the DMCs also showed that each group had distinct methylation patterns (Figure 2D and Table S4). Although imputation was used for missing methylation data values, the non-imputed methylation patterns of each group were also clearly segregated (Figure S1C). Of the 244

**Table 1. Patient demographics and cancer characteristics**

Characteristics	Number of Patients
<b>Age (years)</b>	
50–59	2
60–69	5
70–79	3
<b>Sex</b>	
Male	7
Female	3
<b>Primary cancer site</b>	
Colon, left side	4
Colon, right side	5
Rectum	1
<b>Treatment prior to liver resection</b>	
Adjuvant chemotherapy	3
No treatment	7
<b>Liver metastasis pathology</b>	
Synchronous	5
Metachronous	5
<b>Other metastases</b>	
Lung metastases (metachronous)	4
Lymph node positive	5

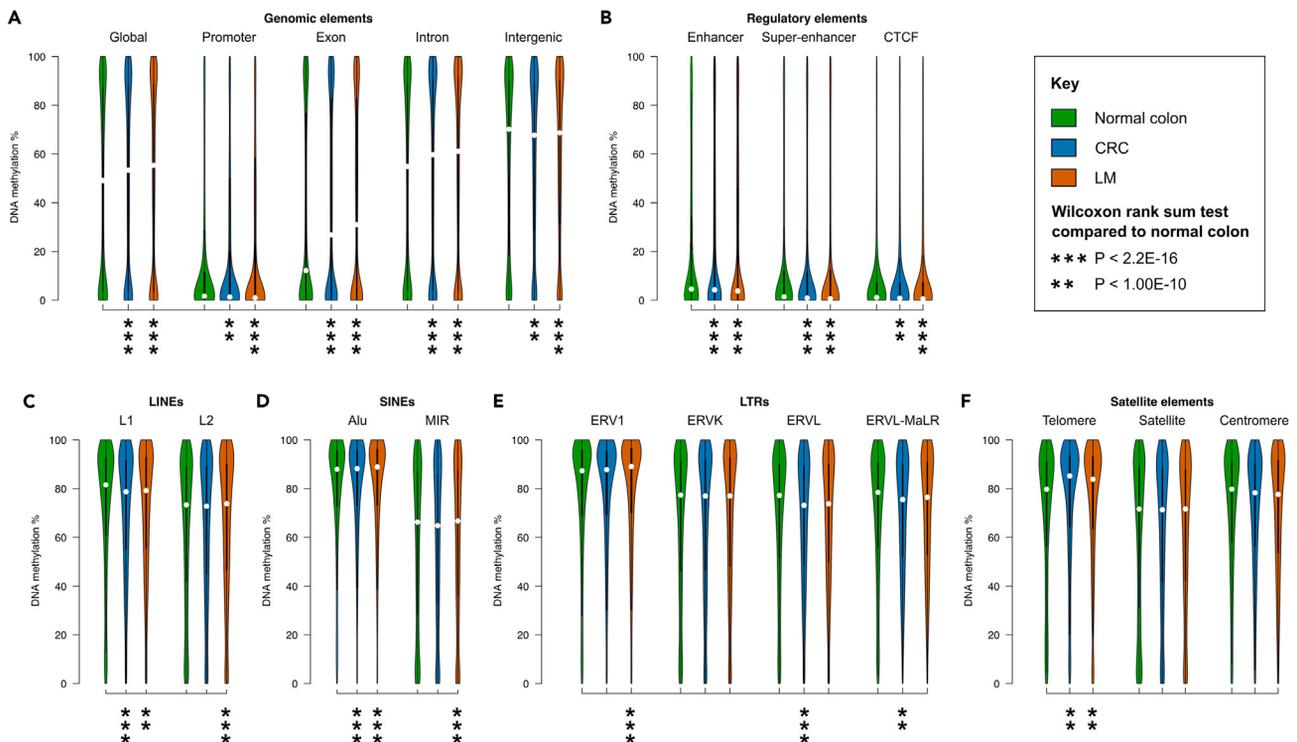
DMCs, 165 were in genes and/or regulatory regions (promoters/enhancers) associated with genes (Figure 2C and Table S4), indicating possible regulatory changes to these genes.

### Independent validation of methylation data and functional analysis of DMCs

The methylation values of our CRC tissues were highly correlated with TCGA colorectal adenocarcinoma (TCGA-COAD) HumanMethylation450K (HM450K) data (Pearson's  $r = 0.97$ , Figure 3A), and with primary colonic tumors from an independent RRBS dataset<sup>28</sup> (Pearson's  $r = 0.96$ , Figure 3A), indicating excellent reproducibility. Using this same RRBS data, PCA of the DMCs showed that the methylation patterns of the primary CRC tissues separated from non-cancerous tissues (normal colon, normal colonic crypt, aberrant crypt foci) and further separated from LM methylation (Figure 3C). To investigate our DMCs in a broader epigenomic context, we utilized several datasets: TCGA-COAD ATAC-seq chromatin accessibility, Epigenomics roadmap histone modification ChIP-Seq, TRANSFAC/JASPAR transcription factor binding profiles, and ENCODE/ChEA consensus target genes. The methylation values of the DMCs in our CRC tissues were negatively correlated with TCGA-COAD chromatin accessibility data (Pearson's  $r = -0.26$ , Figure 3D and Table S5). These loci were highly enriched for H3K27me3, a repressive chromatin mark mediated by the polycomb repressive complex 2 (PRC2). They were also enriched for binding of transcription factors in the Sp/KLF family and of the PRC2 components, EZH2 and SUZ12 (Figure 3E and Table S6). Overall, these findings suggest that the differentially methylated CpG sites between primary colorectal tumors and liver metastases were associated with significant functional regulatory changes in these regions.

### Transcriptomic mapping revealed large expression changes and distinct pathways in liver metastasis

To analyze transcriptomic changes of CRC metastasis, we performed RNA-Seq on the same CRC and LM tissues (Table S7). We identified 424 genes that were significantly differentially expressed (DEG) between CRC and LM (FDR-corrected P-value  $< 0.05$  and fold-change of mean RPKM  $\geq 2$ ). Of these DEGs, 182 genes were down-regulated, whereas the remaining 242 genes were up-regulated in LM (Figure 4A and Table S8). With the exception of one sample (LM270), unsupervised hierarchical clustering of the DEGs separated the



**Figure 1. Global differences in DNA methylation between normal colon, primary colorectal cancers and liver metastases**

(A–F) Equal area violin plots of RRBS data show the distribution of methylation in different genomic elements of paired primary colorectal cancers (CRC,  $n = 10$ ) and liver metastases (LM,  $n = 10$ ) alongside normal colon ( $n = 10$ , Hanley et al. GSE95654). In all cases the Y axis represents DNA methylation level %. See also Table S3.

(A) Gene promoters (-5kb to +1kb from the TSS), exons, introns, and intergenic elements (>5kb upstream from the nearest TSS).

(B) Regulatory elements: enhancers, super-enhancers and CTCF binding sites.

(C) LINES: L1 and L2.

(D) SINES: Alu and MIR.

(E) LTRs: ERV1, ERVK, ERVL and ERV-MaLR.

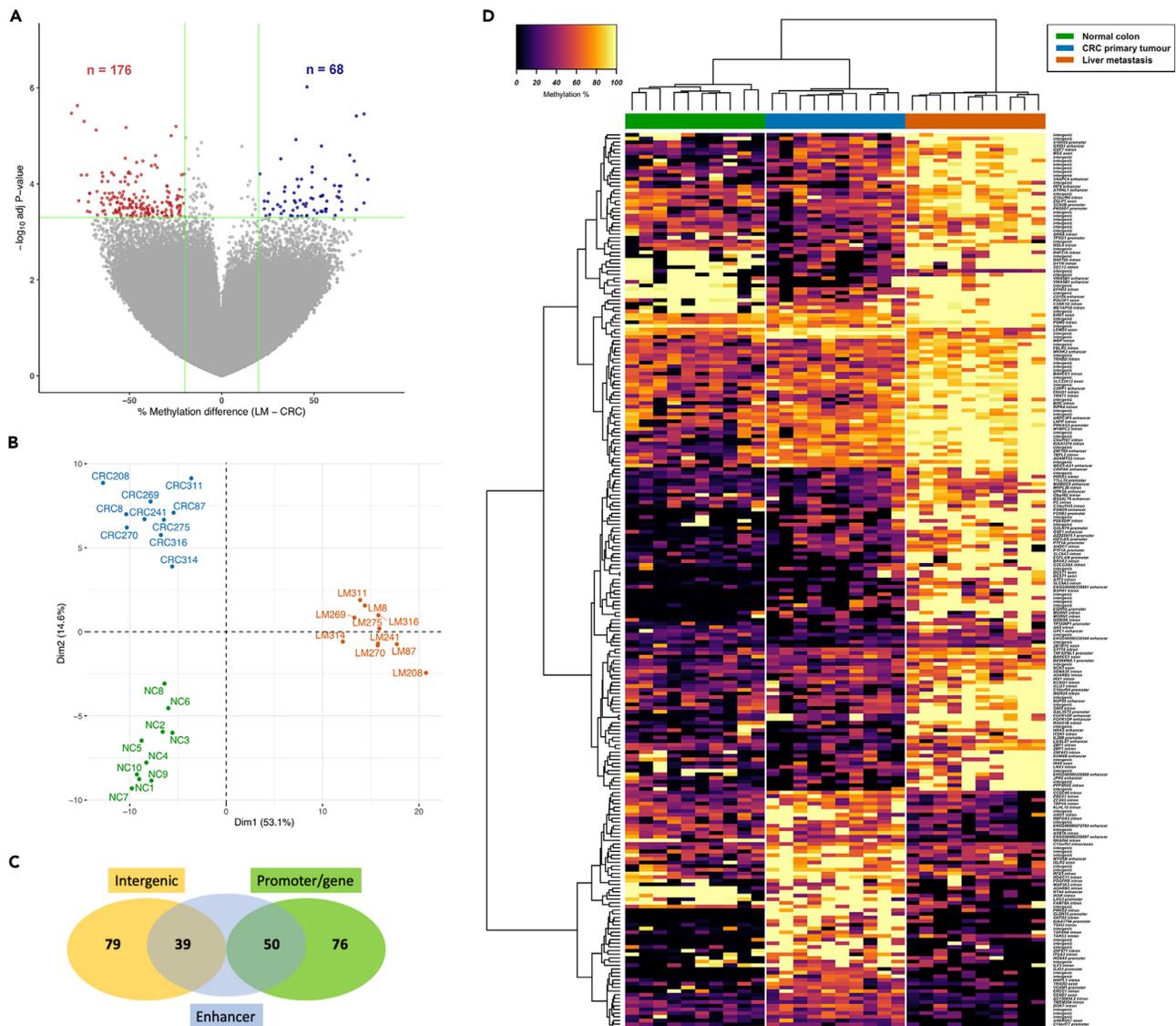
(F) Satellite elements: telomere, satellite, centromere.

expression patterns of the two groups (Figure 4B). PCA of the 500 most variable genes showed only partial segregation of the CRC and LM groups across principal components 1 and 2, which in total accounted for 46% of the variance (Figure 4C).

Functional gene enrichment analysis revealed that the 182 genes down-regulated in LM compared to CRC were mainly involved in gene silencing mechanisms and chromatin organization. These genes were also enriched for regulation of cell differentiation (Figure 4D and Table S9). In contrast, the 242 up-regulated genes were significantly implicated in extracellular matrix organization and cell adhesion. In addition, these genes were enriched for regulation of hemostasis and response to wound healing (Figure 4E and Table S10).

### Immune cell composition and tumor microenvironment in paired primary colorectal cancers and liver metastases

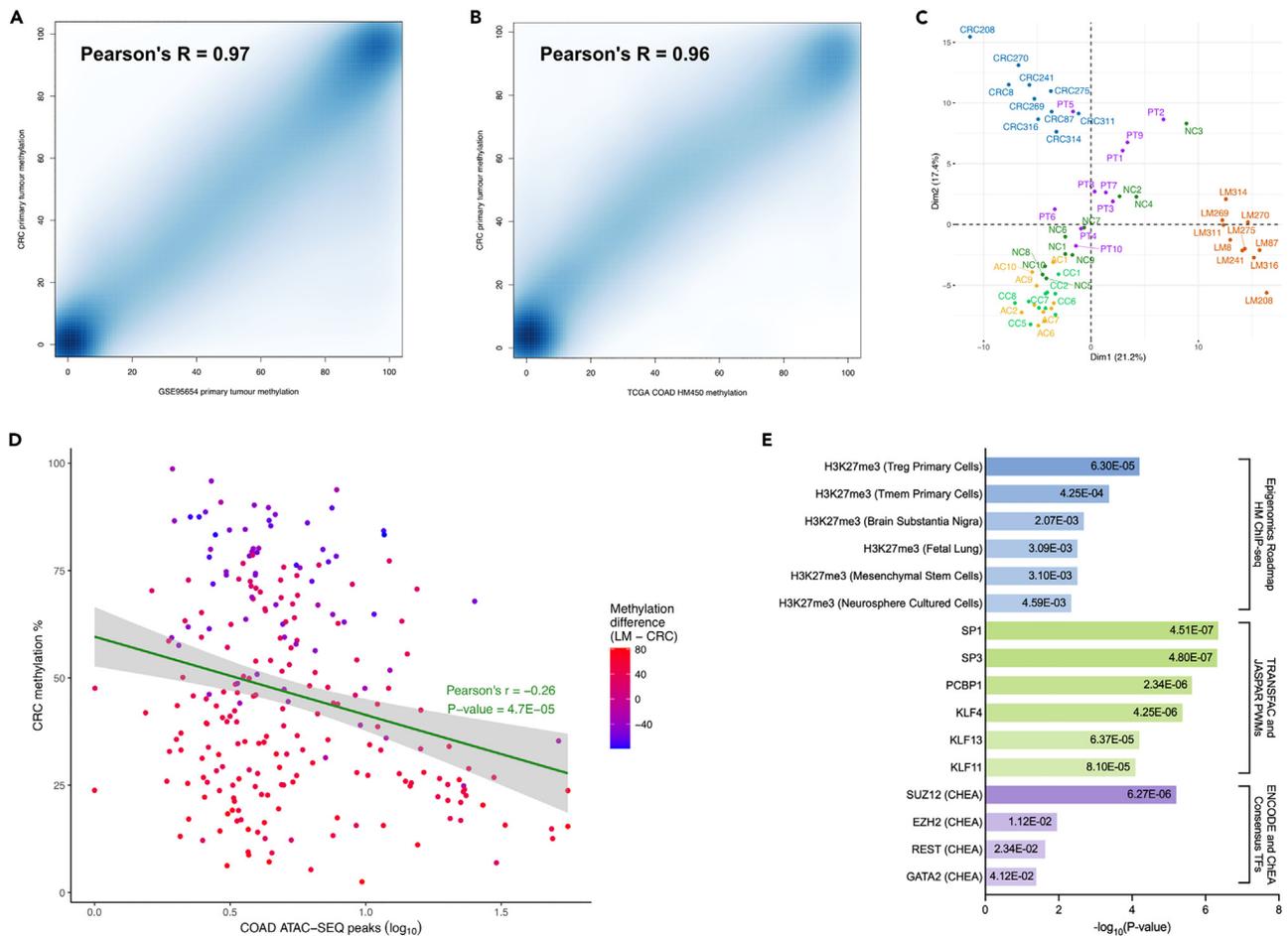
Using expression-based deconvolution, we investigated the fractions of 22 infiltrated immune cell types in the paired CRC and LM tissues. In CRC, CD4 memory resting T cells, M0 macrophages, resting mast cells and resting NK cells were the most infiltrating fractions compared with other immune cells. This pattern is generally what has been previously observed in primary CRCs.<sup>29</sup> In LM, the largest fractions of infiltrating immune cells were M0 macrophages, CD4 memory resting T cells, M2 macrophages and monocytes. The immune cell composition between the paired CRC and LM tumor tissues was mostly similar, although LM had significantly larger fractions of monocytes and M2 macrophages, but fewer resting mast cells than CRC ( $P$ -value <0.05, Figures 5A and 5B, and Table S11).



**Figure 2. Differential methylation patterns between normal colon, primary colorectal cancers and liver metastases**

(A) Volcano plot showing DNA methylation differences between paired primary colorectal cancers (CRC, n = 10) and liver metastases (LM, n = 10). The x axis shows methylation differences of all analyzed CpG sites (LM - CRC) and the y axis shows the  $-\log_{10}$  of P-values. Differentially methylated CpGs (DMCs, P-value  $< 5.0 \times 10^{-4}$  and methylation difference  $\geq 20\%$ ) are shown in either red (increased methylation in LM compared to CRC) or blue (decreased methylation in LM). (B) Principal component analysis (PCA) plot showing dimensionality reduction of the 244 DMCs alongside normal colon methylation data (n = 10, Hanley et al. GSE95654), with imputed missing methylation data values. (C) Venn diagram of the number of DMCs in predicted enhancer elements, associated with genes (promoter/gene body) or in intergenic regions. (D) Methylation heatmap of the 244 DMCs alongside normal colon (black = unmethylated, yellow = fully methylated). See also Figure S1 and Table S4.

To further investigate immune cell composition and tumor microenvironment (TME) in advanced CRC and assess its relationship with the epigenetic signature, we performed a series of additional analyses. We used the ESTIMATE algorithm to obtain immune, stromal and purity scores of each analyzed sample. To benchmark these data from our cohort (CRC and LM), we have analyzed and directly compared them with colorectal adenocarcinoma patients (TCGA-COAD RNA-Seq version 2 dataset) as the standard. We found that the overall immune scores of CRC and LM were not significantly different (P-value = 0.39, Figure 5C and Table S11), which further validates the findings from CIBERSORT analysis that the global immune cell composition of CRC and LM were mostly similar. The immune scores in TCGA CRC patients were higher than for CRC (P-value = 0.04), but not significantly different compared to LM (P-value = 0.63; Figure 5C).



**Figure 3. Validation of genome-scale methylation and functional characterization of differentially methylated CpGs in primary colorectal cancers and liver metastases**

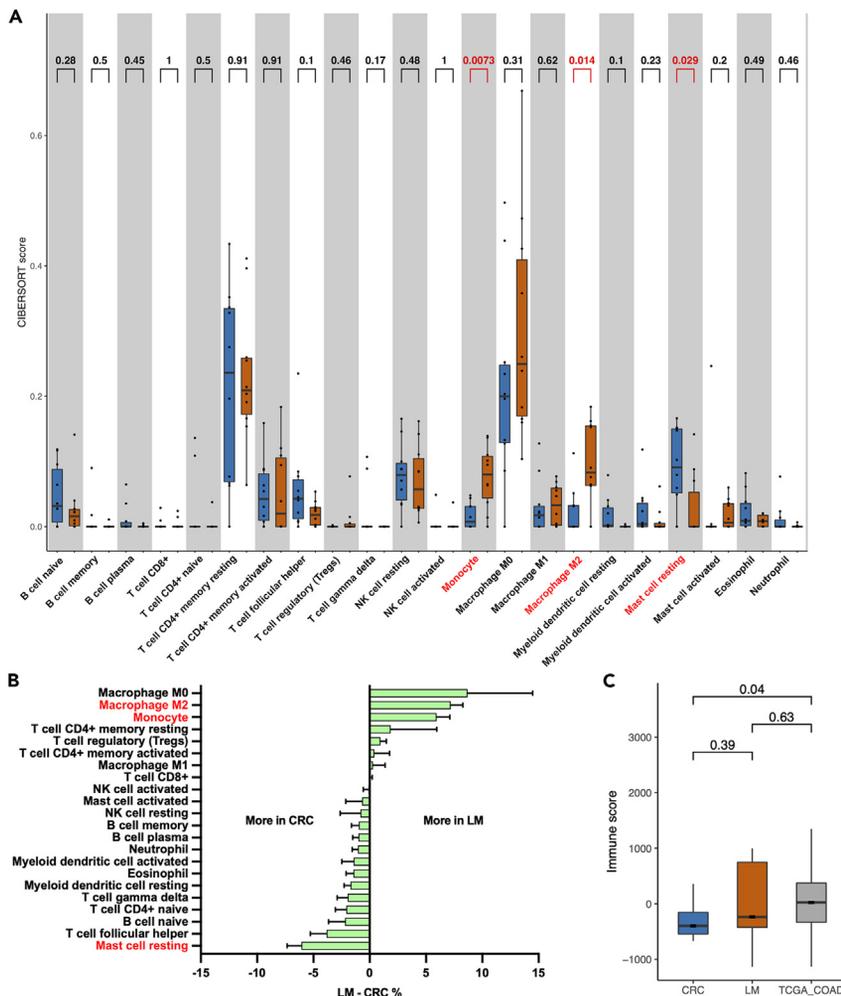
(A and B) Smoothed scatterplot showing methylation correlation of primary colorectal adenocarcinomas (CRC,  $n = 10$ ) compared to A) TCGA colorectal adenocarcinomas (TCGA-COAD,  $n = 410$ ), and B) primary colorectal adenocarcinomas from an independent RRBS dataset ( $n = 10$ , Hanley et al. GSE95654). (C) Principal component analysis (PCA) showing dimensionality reduction of the 244 differentially methylated CpGs (DMCs) alongside normal colon (NC,  $n = 10$ ), normal colonic crypt (CC,  $n = 10$ ), aberrant crypt foci (AC,  $n = 10$ ) and primary colorectal adenocarcinomas (PT,  $n = 10$ ; Hanley et al. GSE95654), with imputed missing methylation data values.

(D) Scatterplot showing correlation of CRC methylation and chromatin accessibility (TCGA-COAD ATAC-seq,  $n = 81$ ) at the DMC loci.

(E) Enrichment analysis of histone modifications (Epigenomics Roadmap), transcription factor binding profiles (TRANSFAC/JASPAR) and consensus target genes (ENCODE/ChEA) associated with the DMCs,  $P\text{-value} < 0.05$ . See also [Tables S5](#) and [S6](#).

For the stromal scores, we also found that the difference between CRC and LM was not significant ( $P\text{-value} = 0.075$ , [Table S11](#)). The TCGA-COAD stromal scores were similar to LM ( $P\text{-value} = 0.39$ ), but significantly higher than the CRC samples ( $P\text{-value} = 0.0097$ ). The tumor purity analysis (ESTIMATE score) also showed that the purity of CRC and LM samples were not significantly different ( $P\text{-value} = 0.11$ , [Table S11](#)) and the purity of our cohort was similar to TCGA – LM scores were not significantly different ( $P\text{-value} = 0.84$ ) whereas CRC samples had higher tumor purity than TCGA ( $P\text{-value} = 0.011$ ). We also found highly significant negative correlations of tumor purity with both immune scores (Pearson  $r = -0.988$ ,  $P\text{-value} = 6.42 \times 10^{-16}$ ) and stromal scores (Pearson  $r = -0.917$ ,  $P\text{-value} = 1.36 \times 10^{-8}$ ). These data are consistent with previous TME analysis using multiple CRC cohorts, which demonstrated significant negative correlations of tumor purity with immune and stromal scores.<sup>30,31</sup> Finally, we analyzed our identified signature in the context of an immune and stromal gene signature related to CRC metastasis. A well-validated 292 gene expression signature associated with immune and stromal components of CRC tumors was previously described using data from TCGA.<sup>32</sup> This immune/stromal gene signature did not overlap with any of the 165 genes associated with the 244 CpG sites in our methylation signature, but did overlap with 44 of the 424 differentially expressed genes between CRC and LM.





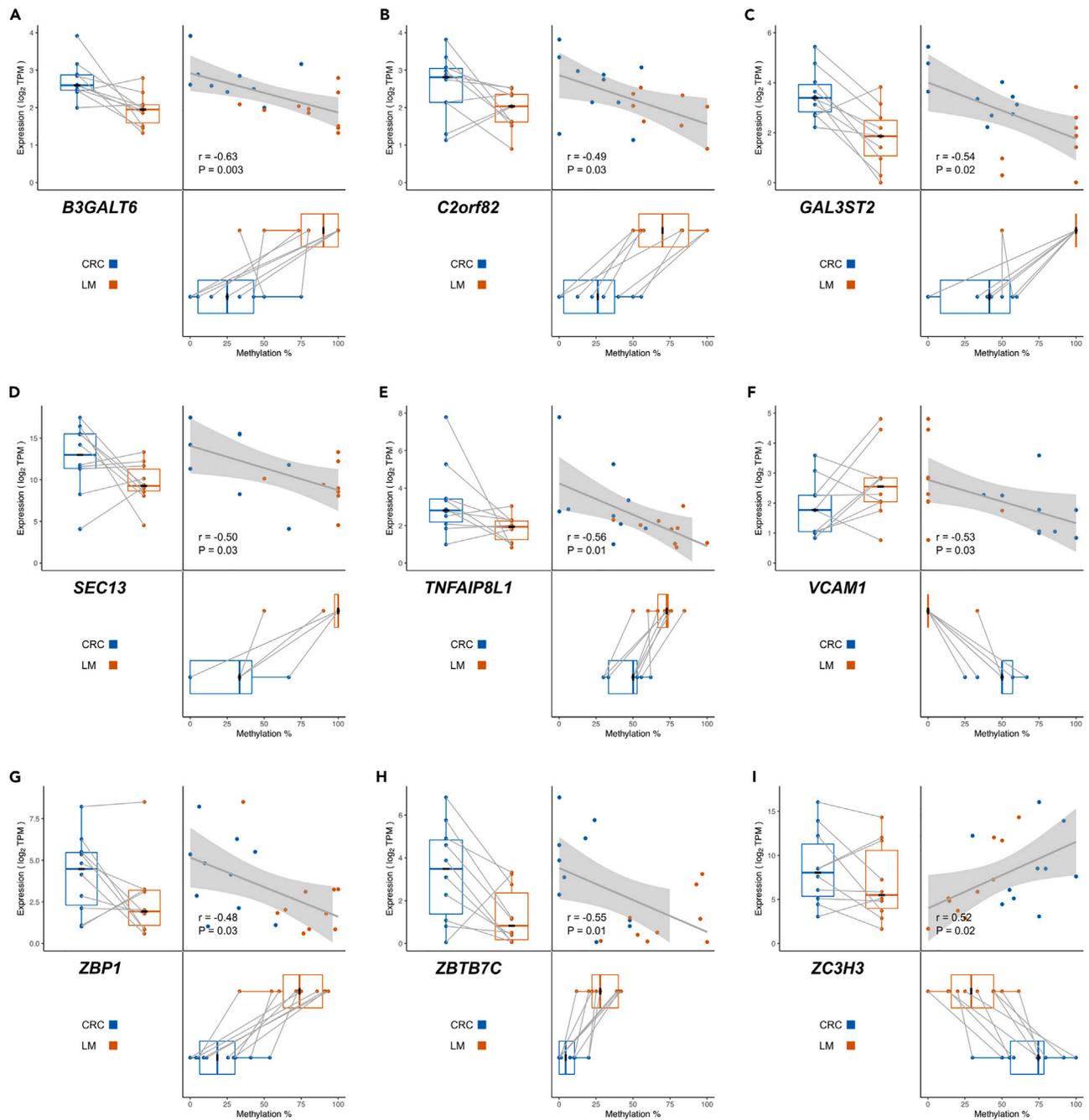
**Figure 5. Expression-based immune composition of paired primary colorectal cancers and liver metastases**

(A) Boxplots of the immune cell composition of paired primary colorectal cancers (CRC, n = 10) and liver metastases (LM, n = 10), estimated by expression-based CIBERSORT deconvolution.

(B) Differences in immune cell composition of LM compared to CRC, based on CIBERSORT scores of paired samples. P-value <0.05 shown in red.

(C) Boxplots showing immune scores of CRC and LM compared to TCGA colorectal adenocarcinomas (TCGA-COAD, n = 263), based on ESTIMATE analysis. See also [Table S11](#).

predicted enhancer elements and 18 had negative correlations between methylation and expression. In most cases, the differences in methylation and expression showed a pairwise pattern between the matched CRC and LM tissues, of which the most significant are shown in [Figure 6](#). The proteins encoded by these genes have a variety of different functions, including: regulation of nucleic acid processing and/or binding (KDM4B, SWI5, ZBP1, ZBTB7C, ZC3H3); intracellular signaling and transport (KLHL10, MAP2K3, NKAIN4, PDGFRB, PRKCZ); ECM/membrane and cell adhesion (EFHD2, EHD1, KCNQ1, SEC13, VCAM1); and lipid/carbohydrate metabolism (B3GALT6, GAL3ST2, TNFAIP8L1). Some of these have been described previously in CRC metastasis to distant organs, such as PDGFRB, PRKCZ, VCAM1 and ZBP1 (aka IMP1).<sup>33–36</sup> Several more have been identified as markers for CRC diagnosis or prognosis, including GAL3ST2, KCNQ1, KDM4B, TNFAIP8L1 (aka *TIPE1*) and ZBTB7C.<sup>37–41</sup> We have also identified several novel putative epigenetic drivers that have not been previously reported in CRC (*C8orf82*, *EFHD2*, *SEC13*, *SWI5* and *ZC3H3*). Furthermore, we validated the reproducibility of the correlated methylation and expression relationship of six genes from TCGA-COAD data. All of them exhibited significantly negative correlated methylation and expression ([Figure S2](#) and [Table S12](#)).



**Figure 6. Concordant DNA methylation and expression differences in paired primary colorectal cancers and liver metastases**

(A–I) Scatterplots and boxplots of the most differentially methylated CpGs (DMCs,  $P$ -value  $< 5.0 \times 10^{-4}$  and methylation difference  $\geq 20\%$ ) with correlated changes in gene expression (Spearman,  $P$ -value  $< 0.05$ ). A) *B3GALT6* enhancer, chr1:1176934–1176935.

(B) *C8orf82* intron, chr8:145753492–145753493.

(C) *GAL3ST2* promoter, chr2:242715169–242715170.

(D) *SEC13* intron, chr3:10346500–10346501.

(E) *TNFAIP8L1* promoter, chr19:4638477–4638478.

(F) *VCAM1* promoter, chr1:101184331–101184332.

(G) *ZBP1* intron, chr20:56194194–56194195.

(H) *ZBTB7C* exon, chr18:45663647–45663648.

(I) *ZC3H3* intron, chr8:144553096–144553097. See also [Figure 6](#) and [Table S12](#).

## DISCUSSION

Metastasis is a complex process which involves several biological and cellular steps that a cancer cell uses to successfully spread<sup>12,13</sup>; it is plausible that epigenetic changes play a driving role in this process.<sup>17,42</sup> Although many studies have provided DNA methylation data for primary CRCs, epigenomic data for tumor metastases remains limited and particularly for paired tumor samples with well-defined clinical features. To our knowledge, here we provide the first sequencing-based genome-wide RRBS methylation and whole transcriptomic map of paired primary CRCs and matched liver metastases (LM) to identify a signature of advanced CRC metastasis. Furthermore, by utilizing multiple patient cohorts we provide extensive independent validation of our results.

We observe that in terms of the global methylation landscape, primary cancers and liver metastases show global hypermethylation when compared to normal colon. The methylome of primary cancers and liver metastases were very similar to each other, suggesting conservation of methylome architecture between primary and matched distant metastases. Global hypomethylation in metastasis compared to primary tumor has been reported for several cancer types,<sup>43,44</sup> however, we did not observe that in our study. The global DNA methylation level is highly dependent on the technical platform used. We have used RRBS, which profiles dynamic and functional regions of the genome, but the true reflection of global methylation status can be most accurately obtained by WGBS (Whole Genome Bisulfite Sequencing), as it allows profiling of large repeat regions which are often hypomethylated in cancer. We have validated our global RRBS methylome by comparison of all common CpGs in independent RRBS data from CRC as well as with Illumina 450K methylation array data from the cancer genome atlas (containing >400 colorectal tumors). In both cases and platforms, we found excellent validation, providing strong evidence for the robustness of our methylation profiles across platforms and cohorts.

Although global methylation levels showed conservation, we have identified strong site-specific methylation changes (244 DMCs) associated with LM in colon cancer and the majority (>70%) of these were hypermethylated in LM. The comparison of normal colon, primary CRC and LM methylomes enabled us to determine the nature of methylation change and whether these DMCs were LM-specific. We revealed that the majority of DMCs showed similar and hypomethylated states in normal colon and primary CRC but became heavily hypermethylated in LM. A relatively small number of DMCs showed dynamic change as these became hypermethylated in primary tumor compared to normal colon and then lost methylation to become hypomethylated in LM. Integration of the methylation profiles of our identified DMCs with chromatin accessibility data (using ATAC-Seq platform) from CRC patients revealed a highly significant correlation of methylation levels and chromatin activity. These data suggest that the methylation levels of the DMCs have a possible functional genomic role, which warrants further investigation. When we analyzed multiple histone modification marks and transcription factor binding profiles, we found that the DMCs were heavily enriched for the repressive chromatin mark H3K27me3 mediated by the PRC2 complex. This was highly consistent with the fact that the DMC signature also showed high enrichment for transcription factors of the core PRC2 complex such as EZH2 and SUZ12. The expression of EZH2 and H3K27me3 were shown to be up-regulated in CRC and associated with advanced tumor stage. H3K27me3 and EZH2 levels were also shown to be positively correlated with the metastasis-free survival of CRC patients.<sup>45–47</sup> We have demonstrated the cooperation of DNA methylation with EZH2 and H3K27me3 levels in other cancers.<sup>48,49</sup> Our consistent results provide further strong evidence of specific and coordinated multilayer epigenetic deregulation in CRC metastasis. This also provides a rationale to explore combinatorial epigenetic therapy (such as DNA methylation inhibitor plus EZH2 inhibitors) for the treatment of CRC in the future.

Metastasis is a complex process that involves multiple molecular steps and is broadly explained by two potential models. In the initiation or founder model, a small population of cells harboring specific genetic or epigenetic profile becomes metastatic over time, whereas in the second and progressive model, the acquisition of specific changes over time provides a selective advantage to the tumor cells to successfully metastasize.<sup>50–52</sup> Our results demonstrate similar global methylomes but highly discriminatory, LM-specific methylation changes in functionally important regions with a role in genomic regulation. These data provide support toward the progressive model where a specific methylation signature is gained and required during the metastasis cascade. Future studies, particularly involving other metastatic sites and appropriate models, could provide important further insights into explaining the possible model of metastasis, which remains a big question in the field of cancer biology. For example, DNA methylation analysis of circulating

tumor cells would reveal dynamic epigenetic changes that are required for the cell to survive in circulation but change again to survive the new metastatic niche.<sup>53</sup>

Our identified signature consisting of 244 DMCs was able to completely segregate LMs from primary CRC and normal colon in PCA analysis. To test the generalizability of our identified signature, we have performed rigorous validation with additional cohorts and analyses. For further validation of our signature, we have included an additional independent cohort (generated by the same RRBS platform),<sup>28</sup> consisting of normal colonic crypt, aberrant crypt foci and additional primary colonic cancers, the 244 CpG signature was able to clearly segregate LMs from all the other groups. Our cohort of tumors consists of matched pairs of primary cancers from both the left and right side of the colon and three patients in the cohort received adjuvant chemotherapy (FOLFOX) before liver resection, whereas the others did not. Therefore, a major finding from this study is our identified methylation signature is able to separate liver metastasis from normal colon, crypt and primary colon adenocarcinoma irrespective of the side of colon cancer and also suggests that the methylation signature may not be affected by treatment pressure. These results provide evidence that our methylation signature represents a universal molecular marker of the CRC metastatic process to the liver irrespective of the specific biology of CRC subtypes. This opens new opportunities to further test these signatures in a clinical setting to predict metastasis. These findings also provide a new avenue to conduct future studies in other cancer types (such as in lung and peritoneal metastasis) to reveal similar true methylation markers that could be identified in cancers originating from other anatomical sites. If this advanced metastasis signature is conserved across different subtypes of CRC, then it is plausible it is the same for other cancer types, which is worth investigating further. Future whole genome-scale studies with a larger number of well-curated and paired metastatic samples are likely to further elucidate the utility of epigenetic signatures in the metastasis process.

Our transcriptomic studies revealed large gene expression changes in LM compared to the matched primary tumors and these changes were associated with metastasis-related and gene regulatory pathways. Further confirming our evidence of multilayer epigenetic deregulation in CRC metastasis, there was an enrichment of differentially expressed genes related to epigenetic regulation of gene expression and chromatin remodeling – particularly H3 and H4 histones. There was a substantial decrease in the expression of different subunit components of the electron transport chain (e.g., complex I subunit gene *NDUFS3*, complex III subunit gene *CYC1* and complex IV subunit gene *COX8A*). This is indicative of metabolic reprogramming, a cancer hallmark known to enhance cancer cell dissemination through EMT activation.<sup>54</sup> Also consistent with alterations to support survival of metastatic tumor cells, there were expression changes in genes involved in extracellular matrix organization (e.g., *ADAMTS10*, *TIMP1*),<sup>55</sup> regulation of wound healing (*FGB*, *SERPINE1*),<sup>56</sup> and regulation of cell-substrate adhesion (e.g., *MACF1*, *NRP1*).<sup>57</sup> However, our PCA analysis of expression profiles was unable to clearly segregate LMs from primary CRC, unlike the DNA methylation signature. Therefore, we provide evidence that DNA methylation is likely to provide a sustained and distinct signature for metastasis in contrast to the transcriptome, which is likely to represent acute and dynamic expression changes in cancer metastasis. A similar observation has been demonstrated in the context of a longitudinal personal multi-omic study, where DNA methylation was temporally stable and changes associated with chronic conditions but transcriptomic changes were related to acute events.<sup>58</sup> By combining the DMCs and their associated gene expression levels from the transcriptome data, we were able to identify 21 loci that each showed a strong methylation-expression relationship in primary and LM tumors. The genes *PDGFRB*, *PRKCZ*, *VCAM1*, and *ZBP1* (aka *IMP1*) have been described previously in CRC metastasis to distant organs.<sup>33–36</sup> It appears that *PDGFRB*,<sup>33</sup> *PRKCZ*,<sup>34</sup> and *VCAM1*<sup>59</sup> all promote liver metastasis in CRC via an EMT-related mechanism. In addition, the genes *GAL3ST2*, *KCNQ1*, *KDM4B*, *TNFAIP8L1* (aka *TIPE1*), and *ZBTB7C* have been identified as markers for CRC diagnosis or prognosis.<sup>37–41</sup> As a histone demethylase, *KDM4B* is an important epigenetic regulator that is more highly expressed in advanced stages of CRC, is associated with poor prognosis, and is regulated by a variety of different cellular stimuli.<sup>60</sup> Notably, we have identified several novel candidates that have not been previously reported in CRC (*C8orf82*, *EFHD2*, *SEC13*, *SWI5* and *ZC3H3*). Our work provides a strong basis for future functional characterization of these candidate epigenetic drivers. Although the candidates described here have gene-specific methylation changes (ie. in genes and/or regulatory regions), we and others have shown previously that methylation changes at intergenic regions and repeat elements can also strongly influence gene expression.<sup>61</sup> Functional studies with *in vitro* and *in vivo* models utilizing cutting-edge epigenomic editing tools<sup>17,62,63</sup> could provide direct evidence for the role of methylation drivers in the metastatic cascade and identify new targets for treating advanced CRC.

The transcriptome profile also allowed us to deconvolute immune cell and tumor microenvironment (TME) composition of primary and matched LMs. The results broadly show that the immune cell make-up is similar in paired CRC and LMs, except for tumor-associated macrophages (TAMs). Orthogonal analysis using ESTIMATE also confirmed the overall similarity of immune scores between CRC and LM. Although TAMs represent a highly heterogeneous population, they are thought to pave the way for tissue invasion and intravasation and to provide favorable conditions for metastasis.<sup>64</sup> We found both classically activated M1 (immune-stimulating and tumor-suppressive) and alternatively activated M2 macrophages (immune-suppressive and tumor-promoting) were increased in LM compared to primary CRC. However, only the increase in M2 macrophages was statistically significant. The proportion of M2/M1 type TAMs has been shown to be positively correlated with liver metastasis in patients with CRC.<sup>65</sup> Our analysis suggests that LM contains elevated levels of mixed subtypes of TAMs compared to matched primary CRC, supporting a line of work that proposes polarizing immune-suppressive and metastasis-promoting macrophages (M2) into immune-stimulating and tumor-suppressive (M1) macrophages as a possible modulator of TME and therapeutic strategy in CRC.<sup>66</sup> Future experimental studies focusing on specific immune cells and their phenotype in the CRC microenvironment are likely to provide an experimental basis for this hypothesis. Our ESTIMATE analysis demonstrated high tumor purity of CRC and LM, which we benchmarked against TCGA data and also showed no significant difference between CRC and LM in terms of purity and stromal score. Furthermore, our identified methylation signature did not show any overlap with the previously established immune/stromal CRC gene signature. These multiple lines of investigation provide evidence that our identified methylation signature and the differences we observed in methylation status are very likely to be characteristics of tumor cells within CRC and LM rather than the influence of infiltrating immune and stromal cells.

In conclusion, we have identified a signature of advanced CRC metastasis by performing unbiased whole genome-scale DNA methylation and full transcriptome analyses of paired primary and liver metastases from CRC patients. Although our genome-wide analyses were performed in well-defined, discrete clinical samples, we show high technical validation as well as excellent validation of our methylation signature in independent cohorts. We also show that the methylation signature is independent of immune and stromal components. Our analysis uncovered universal markers of CRC metastasis to the liver that are independent of clinical subtype and future large-scale studies will aid the assessment of clinical utility of these markers. The metastasis signature was associated with functionally relevant changes to genomic regulation. Furthermore, these markers could represent epigenetic changes that facilitate a more generalized mechanism of metastasis. A more comprehensive validation of this signature in other cancer types and metastatic sites could provide a target for inhibiting metastatic spread.

### Limitations of the study

A limitation of our study is that we have derived the epigenetic signature based on the analysis of only 20 tumors. Paired primary and metastatic tumor samples (especially with good clinical information) are challenging to obtain and therefore limited data is available for paired epigenetic analysis. To test the generalizability of our identified signature, we have performed rigorous validation with additional cohorts and analyses.

### STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- [KEY RESOURCES TABLE](#)
- [RESOURCE AVAILABILITY](#)
  - Lead contact
  - Materials availability
  - Data and code availability
- [EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS](#)
  - Sample collection and analysis
- [METHOD DETAILS](#)
  - RRBS library preparation and sequencing
  - RNA isolation and sequencing
- [QUANTIFICATION AND STATISTICAL ANALYSIS](#)
  - DNA methylation data analysis

- Comparison with external datasets
- Analysis of transcriptomic data
- Pathway, tumour microenvironment and tumour purity analysis

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2023.106986>.

## ACKNOWLEDGMENTS

The authors are grateful to the patients and their families for contributing the samples used in this study, The Cancer Society Tissue Bank (CSTB), and the Otago Genomics Facility at the University of Otago for providing the sequencing service and assistance. This research was funded by the Royal Society of New Zealand Te Aparangi (for a Rutherford Discovery Fellowship to A.C. and Marsden Fund to E.J.R. and A.C.), and grants from Lottery Health Research New Zealand, the University of Otago (A.C.), Maurice and Phyllis Paykel Trust, and Gut Cancer Foundation NZ (to R.V.P.). We are also grateful for additional support from the T.D. Scott Chair in Urology and Department of Surgical Sciences (to A.C. and E.J.R.), Hugh Green Foundation, Colorectal Surgical Society of Australia and New Zealand (to R.V.P.), Maurice Wilkins Centre for Molecular Biodiscovery, Cancer Society of New Zealand, Health Research Council of New Zealand, New Zealand Institute for Cancer Research Trust (to M.R.E. – salary support for A.C., E.J.R., S.A., A.L.), and John Gavin Postdoctoral Fellowship from the Cancer Research Trust New Zealand (to A.A.).

## AUTHOR CONTRIBUTIONS

Conceptualization: E.R., M.E., S.P., J.M., R.P., and A.C.; Formal analysis: E.R., G.G., P.A., P.S., and S.A.; Investigation: E.R., S.A., S.B., A.L., A.A., and S.S.; Resources: M.E., F.F., R.P. and A.C.; Data curation (access and verification of the data): E.R., G.G., P.A., P.S., and S.S.; Writing—original draft: E.R. and A.C.; Writing—review and editing: All authors. Visualization: E.R., G.G., and P.A.; Supervision: M.E., R.P., and A.C.; Project Administration: E.R., R.P., and A.C.; Funding acquisition: R.P. and A.C. All authors read and approved the final version of the manuscript.

## DECLARATION OF INTERESTS

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

## INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: February 13, 2023

Revised: April 12, 2023

Accepted: May 24, 2023

Published: May 28, 2023

## REFERENCES

1. Verma, M., and Kumar, V. (2017). Epigenetic biomarkers in colorectal cancer. *Mol. Diagn. Ther.* 21, 153–165. <https://doi.org/10.1007/s40291-016-0244-x>.
2. Siegel, R.L., Miller, K.D., and Jemal, A. (2016). Cancer statistics, 2016. *CA A Cancer J. Clin.* 66, 7–30. <https://doi.org/10.3322/caac.21332>.
3. Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., and Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA A Cancer J. Clin.* 71, 209–249. <https://doi.org/10.3322/caac.21660>.
4. Stoffel, E.M., and Murphy, C.C. (2020). Epidemiology and mechanisms of the increasing incidence of colon and rectal cancers in young adults. *Gastroenterology* 158, 341–353. <https://doi.org/10.1053/j.gastro.2019.07.055>.
5. Siegel, R., Naishadham, D., and Jemal, A. (2013). Cancer statistics, 2013. *CA A Cancer J. Clin.* 63, 11–30. <https://doi.org/10.3322/caac.21166>.
6. Bupathi, M., and Wu, C. (2016). Biomarkers for immune therapy in colorectal cancer: mismatch-repair deficiency and others. *J. Gastrointest. Oncol.* 7, 713–720. <https://doi.org/10.21037/jgo.2016.07.03>.
7. Tauriello, D.V.F., Calon, A., Lonardo, E., and Batlle, E. (2017). Determinants of metastatic competency in colorectal cancer. *Mol. Oncol.* 11, 97–119. <https://doi.org/10.1002/1878-0261.12018>.
8. Vatandoust, S., Price, T.J., and Karapetis, C.S. (2015). Colorectal cancer: metastases to a single organ. *World J. Gastroenterol.* 21, 11767–11776. <https://doi.org/10.3748/wjg.v21.i41.11767>.
9. Zhou, H., Liu, Z., Wang, Y., Wen, X., Amador, E.H., Yuan, L., Ran, X., Xiong, L., Ran, Y., Chen, W., and Wen, Y. (2022). Colorectal liver metastasis: molecular mechanism and interventional therapy. *Signal Transduct. Targeted Ther.* 7, 70. <https://doi.org/10.1038/s41392-022-00922-2>.
10. Casadaban, L., Rauscher, G., Aklilu, M., Villenes, D., Freels, S., and Maker, A.V. (2016).

- Adjuvant chemotherapy is associated with improved survival in patients with stage II colon cancer. *Cancer* 122, 3277–3287. <https://doi.org/10.1002/cncr.30181>.
11. Chan, G.H.J., and Chee, C.E. (2019). Making sense of adjuvant chemotherapy in colorectal cancer. *J. Gastrointest. Oncol.* 10, 1183–1192. <https://doi.org/10.21037/jgo.2019.06.03>.
  12. Lambert, A.W., Pattabiraman, D.R., and Weinberg, R.A. (2017). Emerging biological principles of metastasis. *Cell* 168, 670–691. <https://doi.org/10.1016/j.cell.2016.11.037>.
  13. Shen, M., and Kang, Y. (2020). Stresses in the metastatic cascade: molecular mechanisms and therapeutic opportunities. *Genes Dev.* 34, 1577–1598. <https://doi.org/10.1101/gad.343251.120>.
  14. Makohon-Moore, A.P., Zhang, M., Reiter, J.G., Bozic, I., Allen, B., Kundu, D., Chatterjee, K., Wong, F., Jiao, Y., Kohutek, Z.A., et al. (2017). Limited heterogeneity of known driver gene mutations among the metastases of individual patients with pancreatic cancer. *Nat. Genet.* 49, 358–366. <https://doi.org/10.1038/ng.3764>.
  15. Hanahan, D. (2022). Hallmarks of cancer: new dimensions. *Cancer Discov.* 12, 31–46. <https://doi.org/10.1158/2159-8290.CD-21-1059>.
  16. Chen, J.F., and Yan, Q. (2021). The roles of epigenetics in cancer progression and metastasis. *Biochem. J.* 478, 3373–3393. <https://doi.org/10.1042/BCJ20210084>.
  17. Banerjee, R., Smith, J., Eccles, M.R., Weeks, R.J., and Chatterjee, A. (2022). Epigenetic basis and targeting of cancer metastasis. *Trends Cancer* 8, 226–241. <https://doi.org/10.1016/j.trecan.2021.11.008>.
  18. Pomerantz, M.M., Qiu, X., Zhu, Y., Takeda, D.Y., Pan, W., Baca, S.C., Gusev, A., Korthauer, K.D., Severson, T.M., Ha, G., et al. (2020). Prostate cancer reactivates developmental epigenomic programs during metastatic progression. *Nat. Genet.* 52, 790–799. <https://doi.org/10.1038/s41588-020-0664-8>.
  19. Lu, Z., Zou, J., Li, S., Topper, M.J., Tao, Y., Zhang, H., Jiao, X., Xie, W., Kong, X., Vaz, M., et al. (2020). Epigenetic therapy inhibits metastases by disrupting premetastatic niches. *Nature* 579, 284–290. <https://doi.org/10.1038/s41586-020-2054-x>.
  20. Beggs, A.D., Jones, A., El-Bahrawy, M., Abulafi, M., Hodgson, S.V., and Tomlinson, I.P.M. (2013). Whole-genome methylation analysis of benign and malignant colorectal tumours. *J. Pathol.* 229, 697–704. <https://doi.org/10.1002/path.4132>.
  21. Martínez-Cardús, A., Moran, S., Musulen, E., Moutinho, C., Manzano, J.L., Martínez-Balibrea, E., Tierno, M., Élez, E., Landolfi, S., Lorden, P., et al. (2016). Epigenetic homogeneity within colorectal tumors predicts shorter relapse-free and overall survival times for patients with locoregional cancer. *Gastroenterology* 151, 961–972. <https://doi.org/10.1053/j.gastro.2016.08.001>.
  22. Vidal, E., Sayols, S., Moran, S., Guillaumet-Adkins, A., Schroeder, M.P., Royo, R., Orozco, M., Gut, M., Gut, I., Lopez-Bigas, N., et al. (2017). A DNA methylation map of human cancer at single base-pair resolution. *Oncogene* 36, 5648–5657. <https://doi.org/10.1038/onc.2017.176>.
  23. Bian, S., Hou, Y., Zhou, X., Li, X., Yong, J., Wang, Y., Wang, W., Yan, J., Hu, B., Guo, H., et al. (2018). Single-cell multiomics sequencing and analyses of human colorectal cancer. *Science* 362, 1060–1063. <https://doi.org/10.1126/science.aao3791>.
  24. Cancer Genome Atlas Network (2012). Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 487, 330–337. <https://doi.org/10.1038/nature11252>.
  25. Jia, M., Gao, X., Zhang, Y., Hoffmeister, M., and Brenner, H. (2016). Different definitions of CpG island methylator phenotype and outcomes of colorectal cancer: a systematic review. *Clin. Epigenet.* 8, 25. <https://doi.org/10.1186/s13148-016-0191-8>.
  26. Roulois, D., Loo Yau, H., Singhanian, R., Wang, Y., Danesh, A., Shen, S.Y., Han, H., Liang, G., Jones, P.A., Pugh, T.J., et al. (2015). DNA-demethylating agents target colorectal cancer cells by inducing viral mimicry by endogenous transcripts. *Cell* 162, 961–973. <https://doi.org/10.1016/j.cell.2015.07.056>.
  27. Anderson, R.L., Balasas, T., Callaghan, J., Coombes, R.C., Evans, J., Hall, J.A., Kinrade, S., Jones, D., Jones, P.S., Jones, R., et al. (2019). A framework for the development of effective anti-metastatic agents. *Nat. Rev. Clin. Oncol.* 16, 185–204. <https://doi.org/10.1038/s41571-018-0134-8>.
  28. Hanley, M.P., Hahn, M.A., Li, A.X., Wu, X., Lin, J., Wang, J., Choi, A.H., Ouyang, Z., Fong, Y., Pfeifer, G.P., et al. (2017). Genome-wide DNA methylation profiling reveals cancer-associated changes within early colonic neoplasia. *Oncogene* 36, 5035–5044. <https://doi.org/10.1038/onc.2017.130>.
  29. Ge, P., Wang, W., Li, L., Zhang, G., Gao, Z., Tang, Z., Dang, X., and Wu, Y. (2019). Profiles of immune cell infiltration and immune-related genes in the tumor microenvironment of colorectal cancer. *Biomed. Pharmacother.* 118, 109228. <https://doi.org/10.1016/j.biopha.2019.109228>.
  30. Wu, Y., Jia, H., Zhou, H., Liu, X., Sun, J., Zhou, X., and Zhao, H. (2022). Immune and stromal related genes in colon cancer: analysis of tumour microenvironment based on the cancer genome atlas (TCGA) and gene expression omnibus (GEO) databases. *Scand. J. Immunol.* 95, e13119. <https://doi.org/10.1111/sji.13119>.
  31. Chen, W., Huang, J., Xiong, J., Fu, P., Chen, C., Liu, Y., Li, Z., Jie, Z., and Cao, Y. (2021). Identification of a tumor microenvironment-related gene signature indicative of disease prognosis and treatment response in colon cancer. *Oxid. Med. Cell. Longev.* 2021, 6290261. <https://doi.org/10.1155/2021/6290261>.
  32. Liu, Y., Cheng, L., Li, C., Zhang, C., Wang, L., and Zhang, J. (2021). Identification of tumor microenvironment-related prognostic genes in colorectal cancer based on bioinformatic methods. *Sci. Rep.* 11, 15040. <https://doi.org/10.1038/s41598-021-94541-6>.
  33. Steller, E.J.A., Raats, D.A., Koster, J., Rutten, B., Govaert, K.M., Emmink, B.L., Snoeren, N., van Hooff, S.R., Holstege, F.C.P., Maas, C., et al. (2013). PDGFRB promotes liver metastasis formation of mesenchymal-like colorectal tumor cells. *Neoplasia* 15, 204–217. <https://doi.org/10.1593/neo.121726>.
  34. Shelton, P.M., Duran, A., Nakanishi, Y., Reina-Campos, M., Kasashima, H., Llado, V., Ma, L., Campos, A., García-Olmo, D., García-Arnan, M., et al. (2018). The secretion of miR-200s by a PKCzeta/ADAR2 signaling Axis promotes liver metastasis in colorectal cancer. *Cell Rep.* 23, 1178–1191. <https://doi.org/10.1016/j.celrep.2018.03.118>.
  35. Alexiou, D., Karayiannakis, A.J., Syrigos, K.N., Zbar, A., Kremmyda, A., Bramis, I., and Tsigris, C. (2001). Serum levels of E-selectin, ICAM-1 and VCAM-1 in colorectal cancer patients: correlations with clinicopathological features, patient survival and tumour surgery. *Eur. J. Cancer* 37, 2392–2397. [https://doi.org/10.1016/s0959-8049\(01\)00318-5](https://doi.org/10.1016/s0959-8049(01)00318-5).
  36. Hamilton, K.E., Noubissi, F.K., Katti, P.S., Hahn, C.M., Davey, S.R., Lundsmith, E.T., Klein-Szanto, A.J., Rhim, A.D., Spiegelman, V.S., and Rustgi, A.K. (2013). IMP1 promotes tumor growth, dissemination and a tumor-initiating cell phenotype in colorectal cancer cell xenografts. *Carcinogenesis* 34, 2647–2654. <https://doi.org/10.1093/carcin/bgt217>.
  37. Snow, S.M., Matkowskyj, K.A., Maresh, M., Clipson, L., Vo, T.N., Johnson, K.A., Deming, D.A., Newton, M.A., Grady, W.M., Pickhardt, P.J., and Halberg, R.B. (2022). Validation of genetic classifiers derived from mouse and human tumors to identify molecular subtypes of colorectal cancer. *Hum. Pathol.* 119, 1–14. <https://doi.org/10.1016/j.humpath.2021.10.002>.
  38. den Uil, S.H., Coupé, V.M.H., Linnekamp, J.F., van den Broek, E., Goos, J.A.C.M., Delis-van Diemen, P.M., Belt, E.J.T., van Grieken, N.C.T., Scott, P.M., Vermeulen, L., et al. (2016). Loss of KCNQ1 expression in stage II and stage III colon cancer is a strong prognostic factor for disease recurrence. *Br. J. Cancer* 115, 1565–1574. <https://doi.org/10.1038/bjc.2016.376>.
  39. Chen, B., Zhu, Y., Chen, J., Feng, Y., and Xu, Y. (2021). Activation of TC10-like transcription by lysine demethylase KDM4B in colorectal cancer cells. *Front. Cell Dev. Biol.* 9, 617549. <https://doi.org/10.3389/fcell.2021.617549>.
  40. Ye, T., Yang, B., Wang, C., Su, C., Luo, J., Yang, X., Yu, H., Yuan, Z., Meng, Z., and Xia, J. (2020). TIE1 impairs stemness maintenance in colorectal cancer through directly targeting beta-catenin. *Carcinogenesis* 41, 25–35. <https://doi.org/10.1093/carcin/bgz079>.
  41. Chen, X., Jiang, Z., Pu, Y., Jiang, X., Xiang, L., and Jiang, Z. (2020). Zinc finger and BTB domain-containing 7C (ZBTB7C) expression as an independent prognostic factor for colorectal cancer and its relevant molecular mechanisms. *Am. J. Transl. Res.* 12, 4141–4159.

42. Chatterjee, A., Rodger, E.J., and Eccles, M.R. (2018). Epigenetic drivers of tumourigenesis and cancer metastasis. *Semin. Cancer Biol.* 51, 149–159. <https://doi.org/10.1016/j.semcancer.2017.08.004>.
43. Vidal, E., Sayols, S., Moran, S., Guillaumet-Adkins, A., Schroeder, M.P., Royo, R., Orozco, M., Gut, M., Gut, I., Lopez-Bigas, N., et al. (2017). A DNA methylation map of human cancer at single base-pair resolution. *Oncogene* 36, 5648–5657. <https://doi.org/10.1038/onc.2017.176>.
44. Chatterjee, A., Stockwell, P.A., Ahn, A., Rodger, E.J., Leichter, A.L., and Eccles, M.R. (2017). Genome-wide methylation sequencing of paired primary and metastatic cell lines identifies common DNA methylation changes and a role for EBF3 as a candidate epigenetic driver of melanoma metastasis. *Oncotarget* 8, 6085–6101. <https://doi.org/10.18632/oncotarget.14042>.
45. Wang, Q., Chen, X., Jiang, Y., Liu, S., Liu, H., Sun, X., Zhang, H., Liu, Z., Tao, Y., Li, C., et al. (2020). Elevating H3K27me3 level sensitizes colorectal cancer to oxaliplatin. *J. Mol. Cell Biol.* 12, 125–137. <https://doi.org/10.1093/jmcb/mjz032>.
46. Wang, C.G., Ye, Y.J., Yuan, J., Liu, F.F., Zhang, H., and Wang, S. (2010). EZH2 and STAT6 expression profiles are correlated with colorectal cancer stage and prognosis. *World J. Gastroenterol.* 16, 2421–2427. <https://doi.org/10.3748/wjg.v16.i19.2421>.
47. Vilorio-Marqués, L., Martín, V., Díez-Tascón, C., González-Sevilla, M.F., Fernández-Villa, T., Honrado, E., Davila-Batista, V., and Molina, A.J. (2017). The role of EZH2 in overall survival of colorectal cancer: a meta-analysis. *Sci. Rep.* 7, 13806. <https://doi.org/10.1038/s41598-017-13670-z>.
48. Tiffen, J., Gallagher, S.J., Filipp, F., Gunatilake, D., Emran, A.A., Cullinane, C., Dutton-Register, K., Aoude, L., Hayward, N., Chatterjee, A., et al. (2020). EZH2 cooperates with DNA methylation to downregulate key tumour suppressors and interferon gene signatures in melanoma. *J. Invest. Dermatol.* 140, 2442–2454.e5. <https://doi.org/10.1016/j.jid.2020.02.042>.
49. Emran, A.A., Chatterjee, A., Rodger, E.J., Tiffen, J.C., Gallagher, S.J., Eccles, M.R., and Hersey, P. (2019). Targeting DNA methylation and EZH2 activity to overcome melanoma resistance to immunotherapy. *Trends Immunol.* 40, 328–344. <https://doi.org/10.1016/j.it.2019.02.004>.
50. Nowell, P.C. (1976). The clonal evolution of tumor cell populations. *Science* 194, 23–28. <https://doi.org/10.1126/science.959840>.
51. Ramaswamy, S., Ross, K.N., Lander, E.S., and Golub, T.R. (2003). A molecular signature of metastasis in primary solid tumors. *Nat. Genet.* 33, 49–54. <https://doi.org/10.1038/ng1060>.
52. Fares, J., Fares, M.Y., Khachfe, H.H., Salhab, H.A., and Fares, Y. (2020). Molecular principles of metastasis: a hallmark of cancer revisited. *Signal Transduct. Targeted Ther.* 5, 28. <https://doi.org/10.1038/s41392-020-0134-x>.
53. Vasantharajan, S.S., Eccles, M.R., Rodger, E.J., Pattison, S., McCall, J.L., Gray, E.S., Calapre, L., and Chatterjee, A. (2021). The Epigenetic landscape of Circulating tumour cells. *Biochim. Biophys. Acta Rev. Canc* 1875, 188514. <https://doi.org/10.1016/j.bbcan.2021.188514>.
54. Sun, L., Zhang, H., and Gao, P. (2022). Metabolic reprogramming and epigenetic modifications on the path to cancer. *Protein Cell* 13, 877–919. <https://doi.org/10.1007/s13238-021-00846-7>.
55. Winkler, J., Abisoye-Ogunniyan, A., Metcalf, K.J., and Werb, Z. (2020). Concepts of extracellular matrix remodelling in tumour progression and metastasis. *Nat. Commun.* 11, 5120. <https://doi.org/10.1038/s41467-020-18794-x>.
56. Deyell, M., Garris, C.S., and Laughney, A.M. (2021). Cancer metastasis as a non-healing wound. *Br. J. Cancer* 124, 1491–1502. <https://doi.org/10.1038/s41416-021-01309-w>.
57. Janiszewska, M., Primi, M.C., and Izard, T. (2020). Cell adhesion in cancer: beyond the migration of single cells. *J. Biol. Chem.* 295, 2495–2505. <https://doi.org/10.1074/jbc.REV119.007759>.
58. Chen, R., Xia, L., Tu, K., Duan, M., Kukurba, K., Li-Pook-Tham, J., Xie, D., and Snyder, M. (2018). Longitudinal personal DNA methylome dynamics in a human with a chronic condition. *Nat. Med.* 24, 1930–1939. <https://doi.org/10.1038/s41591-018-0237-x>.
59. Zhang, D., Bi, J., Liang, Q., Wang, S., Zhang, L., Han, F., Li, S., Qiu, B., Fan, X., Chen, W., et al. (2020). VCAM1 promotes tumor cell invasion and metastasis by inducing EMT and transendothelial migration in colorectal cancer. *Front. Oncol.* 10, 1066. <https://doi.org/10.3389/fonc.2020.01066>.
60. Wilson, C., and Krieg, A.J. (2019). KDM4B: a nail for every hammer? *Genes* 10, 134. <https://doi.org/10.3390/genes10020134>.
61. Chatterjee, A., Rodger, E.J., Ahn, A., Stockwell, P.A., Parry, M., Motwani, J., Gallagher, S.J., Shklovskaya, E., Tiffen, J., Eccles, M.R., and Hersey, P. (2018). Marked global DNA hypomethylation is associated with constitutive PD-L1 expression in melanoma. *iScience* 4, 312–325. <https://doi.org/10.1016/j.isci.2018.05.021>.
62. Smith, J., Banerjee, R., Weeks, R.J., and Chatterjee, A. (2022). Editing of DNA methylation patterns using CRISPR-based tools. *Methods Mol. Biol.* 2458, 63–74. [https://doi.org/10.1007/978-1-0716-2140-0\\_4](https://doi.org/10.1007/978-1-0716-2140-0_4).
63. Urbano, A., Smith, J., Weeks, R.J., and Chatterjee, A. (2019). Gene-specific targeting of DNA methylation in the mammalian genome. *Cancers* 11, 1515. <https://doi.org/10.3390/cancers11101515>.
64. Cortese, N., Soldani, C., Franceschini, B., Barbagallo, M., Marchesi, F., Torzilli, G., and Donadon, M. (2019). Macrophages in colorectal cancer liver metastases. *Cancers* 11, 633. <https://doi.org/10.3390/cancers11050633>.
65. Cui, Y.L., Li, H.K., Zhou, H.Y., Zhang, T., and Li, Q. (2013). Correlations of tumor-associated macrophage subtypes with liver metastases of colorectal cancer. *Asian Pac. J. Cancer Prev. APJCP* 14, 1003–1007. <https://doi.org/10.7314/apjcp.2013.14.2.1003>.
66. Wang, H., Tian, T., and Zhang, J. (2021). Tumor-associated macrophages (TAMs) in colorectal cancer (CRC): from mechanism to therapy and prognosis. *Int. J. Mol. Sci.* 22, 8470. <https://doi.org/10.3390/ijms22168470>.
67. Al Momeni, S., Rodger, E.J., Stockwell, P.A., Eccles, M.R., and Chatterjee, A. (2022). Generating sequencing-based DNA methylation maps from low DNA input samples. *Methods Mol. Biol.* 2458, 3–21. [https://doi.org/10.1007/978-1-0716-2140-0\\_1](https://doi.org/10.1007/978-1-0716-2140-0_1).
68. Chatterjee, A., Rodger, E.J., Stockwell, P.A., Weeks, R.J., and Morison, I.M. (2012). Technical considerations for reduced representation bisulfite sequencing with multiplexed libraries. *J. Biomed. Biotechnol.* 2012, 741542. <https://doi.org/10.1155/2012/741542>.
69. Purcell, R.V., Visnovska, M., Biggs, P.J., Schmeier, S., and Frizelle, F.A. (2017). Distinct gut microbiome patterns associate with consensus molecular subtypes of colorectal cancer. *Sci. Rep.* 7, 11590. <https://doi.org/10.1038/s41598-017-11237-6>.
70. Chatterjee, A., Stockwell, P.A., Rodger, E.J., and Morison, I.M. (2012). Comparison of alignment software for genome-wide bisulphite sequence data. *Nucleic Acids Res.* 40, e79. <https://doi.org/10.1093/nar/gks150>.
71. Stockwell, P.A., Chatterjee, A., Rodger, E.J., and Morison, I.M. (2014). DMAP: differential methylation analysis package for RRBS and WGBS data. *Bioinformatics* 30, 1814–1822. <https://doi.org/10.1093/bioinformatics/btu126>.
72. Krueger, F., and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 27, 1571–1572. <https://doi.org/10.1093/bioinformatics/btr167>.
73. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. <https://doi.org/10.1093/bioinformatics/btq033>.
74. Fishilevich, S., Nudel, R., Rappaport, N., Hadar, R., Plaschkes, I., Iny Stein, T., Rosen, N., Kohn, A., Twik, M., Safran, M., et al. (2017). GeneHancer: Genome-wide Integration of Enhancers and Target Genes in GeneCards Database. <https://doi.org/10.1093/database/bax028>.
75. Orouji, E., Raman, A.T., Singh, A.K., Sorokin, A., Arslan, E., Ghosh, A.K., Schulz, J., Terranova, C., Jiang, S., Tang, M., et al. (2022). Chromatin state dynamics confers specific therapeutic strategies in enhancer subtypes of colorectal cancer. *Gut* 71, 938–949. <https://doi.org/10.1136/gutjnl-2020-322835>.

76. Wang, H., Maurano, M.T., Qu, H., Varley, K.E., Gertz, J., Pauli, F., Lee, K., Canfield, T., Weaver, M., Sandstrom, R., et al. (2012). Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res.* 22, 1680–1688. <https://doi.org/10.1101/gr.136101.111>.
77. Smit, A.F.A., Hubley, R., and Green, P. (2013). *RepeatMasker Open-4.0*.
78. Díez-Villanueva, A., Mallona, I., and Peinado, M.A. (2015). Wanderer, an interactive viewer to explore DNA methylation and gene expression data in human cancer. *Epigenet. Chromatin* 8, 22. <https://doi.org/10.1186/s13072-015-0014-8>.
79. Xie, Z., Bailey, A., Kuleshov, M.V., Clarke, D.J.B., Evangelista, J.E., Jenkins, S.L., Lachmann, A., Wojciechowicz, M.L., Kropiwnicki, E., Jagodnik, K.M., et al. (2021). Gene set knowledge discovery with Enrichr. *Curr. Protoc.* 1, e90. <https://doi.org/10.1002/cpz1.90>.
80. Gimenez, G., Stockwell, P.A., Rodger, E.J., and Chatterjee, A. (2023). Strategy for RNA-seq experimental design and data analysis. *Methods Mol. Biol.* 2588, 249–278. [https://doi.org/10.1007/978-1-0716-2780-8\\_16](https://doi.org/10.1007/978-1-0716-2780-8_16).
81. Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. <https://doi.org/10.1038/nmeth.3317>.
82. Li, T., Fu, J., Zeng, Z., Cohen, D., Li, J., Chen, Q., Li, B., and Liu, X.S. (2020). TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res.* 48, W509–W514. <https://doi.org/10.1093/nar/gkaa407>.
83. Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-García, W., Treviño, V., Shen, H., Laird, P.W., Levine, D.A., et al. (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* 4, 2612. <https://doi.org/10.1038/ncomms3612>.

## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Biological samples</b>		
Tissue samples of primary colorectal adenocarcinomas and liver metastases from the same patients	Cancer Society Tissue Bank (University of Otago, Christchurch, New Zealand)	<a href="https://www.otago.ac.nz/mackenzie-cancer/tissue-bank/">https://www.otago.ac.nz/mackenzie-cancer/tissue-bank/</a>
<b>Critical commercial assays</b>		
TruSeq DNA Nano Low Throughput Library Prep Kit	Illumina	Cat#: TG-202-1001
EZ DNA Methylation Kit	Zymo Research	Cat#: D5001
TruSeq RNA Library Preparation Kit v2	Illumina	Cat#: RS-122-2001
<b>Deposited data</b>		
Raw and analyzed RRBS and RNA-seq data of primary colorectal adenocarcinomas and liver metastases from the same patients	This paper	GEO: GSE213402
Human reference genome NCBI build 37, GRCh37	Genome Reference Consortium	<a href="http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/">http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/</a>
Raw RRBS data of primary colorectal adenocarcinomas, adjacent normal colon, normal colonic crypt, and aberrant crypt foci	Hanley et al. <sup>1</sup>	GEO: GSE95654
GeneHancer	Fishilevich et al. <sup>2</sup>	<a href="https://genome.ucsc.edu/">https://genome.ucsc.edu/</a>
H3K27Ac CHIP-seq data of HCT116 cells	Orouji et al. <sup>3</sup>	GEO: GSE136888
CTCF CHIP-seq data of HCT116 cells	Wang et al. <sup>4</sup>	GEO: GSE30263
RepeatMasker	Smit et al. <sup>5</sup>	<a href="https://genome.ucsc.edu/">https://genome.ucsc.edu/</a>
HumanMethylation450K methylation data from colon adenocarcinoma and rectum adenocarcinoma cohorts	TCGA	<a href="https://portal.gdc.cancer.gov">https://portal.gdc.cancer.gov</a>
ATAC-Seq data from colon adenocarcinoma cohort	TCGA	<a href="https://portal.gdc.cancer.gov/">https://portal.gdc.cancer.gov/</a>
Expression-based ESTIMATE scores from colon adenocarcinoma cohort (RNA-Seq version 2)	TCGA	<a href="https://bioinformatics.mdanderson.org/estimate/">https://bioinformatics.mdanderson.org/estimate/</a>
<b>Software and algorithms</b>		
Bismark	Krueger et al. <sup>6</sup>	<a href="https://www.bioinformatics.babraham.ac.uk/projects/bismark/">https://www.bioinformatics.babraham.ac.uk/projects/bismark/</a>
DMAP	Stockwell et al. <sup>7</sup>	<a href="https://www.otago.ac.nz/chatterjee-lab/tools/">https://www.otago.ac.nz/chatterjee-lab/tools/</a>
R packages: hclust, prcomp, methylImp, TCGAbiolinks, deeptools, featureCounts, DESeq2, ClusterProfiler	RStudio	<a href="https://posit.co">https://posit.co</a>
BEDTools	Quinlan et al. <sup>8</sup>	<a href="https://bedtools.readthedocs.io/en/latest/">https://bedtools.readthedocs.io/en/latest/</a>
Wanderer	Díez-Villanueva et al. <sup>9</sup>	<a href="http://maplab.cat/wanderer">http://maplab.cat/wanderer</a>
Enrichr	Xie et al. <sup>10</sup>	<a href="https://maayanlab.cloud/Enrichr/">https://maayanlab.cloud/Enrichr/</a>
HISAT2	Kim et al. <sup>11</sup>	<a href="http://daehwankimlab.github.io/hisat2/">http://daehwankimlab.github.io/hisat2/</a>
TIMER2.0	Li et al. <sup>12</sup>	<a href="http://timer.cistrome.org">http://timer.cistrome.org</a>
ESTIMATE	Yoshihara et al. <sup>13</sup>	<a href="https://bioinformatics.mdanderson.org/public-software/estimate/">https://bioinformatics.mdanderson.org/public-software/estimate/</a>

## RESOURCE AVAILABILITY

### Lead contact

Further information and requests should be directed to and will be fulfilled by the lead contact, Aniruddha Chatterjee: [aniruddha.chatterjee@otago.ac.nz](mailto:aniruddha.chatterjee@otago.ac.nz).

### Materials availability

This study did not generate any new unique materials or reagents.

### Data and code availability

- RRBS and RNA-seq data have been deposited at GEO and are publicly available as of the date of publication; the series accession number is listed in the [key resources table](#). De-identified patient data is available in this paper's [supplemental information](#). This paper also analyses existing, publicly available data. The accession numbers for these datasets are listed in the [key resources table](#).
- This paper does not report original code.
- Any additional information required to reanalyse the data reported in this paper is available from the [lead contact](#) upon request.

## EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

### Sample collection and analysis

Tissue samples of primary colorectal adenocarcinomas (n=10) and liver metastases from the same patients (n=10) were accessed from the Cancer Society Tissue Bank (University of Otago, Christchurch). Ethical approval was granted for the study from the University of Otago, Human Ethics Committee (Approval number: H16/037). Patients provided written, informed consent prior to biobanking of specimens. Patient data, including staging, recurrence, metastases, treatment and histology were retrospectively collected from patient medical records ([Table S1](#)). Routine clinical testing of MMR, *BRAF* and *KRAS* was not undertaken at the time these patients underwent resection and therefore this data is not available. Exclusion criteria included patients with hereditary CRC, and patients who had received pre-operative chemotherapy or radiation therapy, prior to resection of their primary cancers. Patients were included if resection specimens were available from their liver metastases. Tumour core samples were dissected from surgical specimens and immediately frozen in liquid nitrogen and initially stored at  $-80^{\circ}\text{C}$ .

## METHOD DETAILS

### RRBS library preparation and sequencing

We used reduced representation bisulfite sequencing (RRBS) to map genome-scale DNA methylation at single nucleotide resolution as we have described previously.<sup>67,68</sup> Briefly, genomic DNA was extracted from <20 mg of tissue using the Qiagen QIAamp DNA Mini Kit. The DNA was digested with NEB *MspI* enzyme followed by end-repair and ligation of Illumina TruSeq sequencing adaptors. Fragments were then size-selected and bisulfite-converted using the Zymo EZ DNA Methylation Kit prior to 16–18 rounds of PCR amplification. The quality and size distribution of the libraries was determined using an Agilent 2100 Bioanalyzer and libraries were sequenced on an Illumina HiSeq2500 machine (100 bp reads, single-ended). Base-calling of the reads was performed with Illumina Real Time Analyzer (RTA) software.

### RNA isolation and sequencing

RNA extraction and sequencing were carried out as detailed previously.<sup>69</sup> Briefly, RNA was extracted from <20 mg of tissue using the RNeasy Plus Mini Kit, including DNase treatment, following tissue disruption using a Retsch Mixer Mill. RNA-seq libraries were prepared using Illumina TruSeq V2 reagents, including ribosomal RNA depletion using RiboZero Gold. Sequencing was performed using the Illumina HiSeq 2500 V4 platform to produce 125 bp paired-end reads.

## QUANTIFICATION AND STATISTICAL ANALYSIS

### DNA methylation data analysis

The quality check and processing of the sequenced RRBS reads was performed using in-house developed bioinformatics tools as previously described.<sup>70,71</sup> The sequencing coverage and quality statistics for each sample are summarised in [Table S2](#). Processed sequence reads were aligned to the reference human

genome (GRCh37) using Bismark<sup>72</sup> with stringent mapping criteria by allowing only one mismatch (default = 2) in the seed. The median non-CpG DNA methylation of all samples was 0.41%, indicating effective bisulfite conversion and low levels of true non-CpG methylation. CpG sites overlapping with SNPs (dbSNP build 151) were filtered out. For CpGs with at least 5× coverage in at least eight paired samples, a two-sample t-test with Welch correction identified 244 differentially methylated CpGs (DMCs) with at least 20% difference in methylation and P-value < 5.0 × 10<sup>-04</sup>. Gene information for all CpGs, including location of gene promoters, introns and exons, was derived using the *identgeneloc* program of the DMAP tool.<sup>71</sup> R packages were used for unsupervised hierarchical clustering of CpG site methylation, using ward linkage (hclust method = "ward.D2") and Euclidean distance (dist method = "euclidean"), and Principal Component Analysis (PCA) using default parameters (prcomp). The methylImp package was used to impute missing methylation data values (maximum 2 per group for each CpG).

### Comparison with external datasets

We utilised previously described RRBS data from primary colorectal adenocarcinomas (n=10), adjacent normal colon (n=10), normal colonic crypt (n=10) and aberrant crypt foci (n=10; GSE95654).<sup>28</sup> There were 4,733,522 CpGs in common with our data. We used BEDTools<sup>73</sup> to overlap the methylation data with several genomic elements and regulatory features: enhancers (GeneHancer<sup>74</sup>), super-enhancers (H3K27Ac ChIP-seq data of HCT116 cells, GSE136888),<sup>75</sup> CTCF binding sites (CTCF ChIP-seq data of HCT116 cells, GSE30263)<sup>76</sup> and specific repeat classes (RepeatMasker).<sup>77</sup> Level 2 normalised TCGA Infinium HumanMethylation450K (HM450K) methylation data from colon adenocarcinoma (TCGA-COAD) and rectum adenocarcinoma (TCGA-READ) projects (n = 410) were obtained using the R package TCGAAbiolinks. There were 136,838 CpGs in common with our data. The Wanderer tool<sup>78</sup> was used to obtain level 3 TCGA-COAD methylation array data of the HM450K probes closest to each of the DMCs and corresponding RNA-Seq gene expression data (n=249). ATAC-Seq bigwig files from the TCGA-COAD cohort (n=81) were downloaded from the National Cancer Institute Genomic Data Commons (<https://portal.gdc.cancer.gov/>). The multiBamSummary function from the deeptools package was used to profile the chromatin accessibility at the loci of the 244 DMCs. The online platform Enrichr<sup>79</sup> was used to infer the enrichment of the DMCs within several datasets (modified Fisher's exact test, P-value < 0.05): Epigenomics Roadmap histone modifications, TRANSFAC/JASPAR transcription factor binding profiles and ENCODE/ChEA consensus target genes.

### Analysis of transcriptomic data

The quality check and processing of the RNA-seq data was performed as previously described.<sup>80</sup> The sequencing coverage and quality statistics for each sample are summarised in Table S7. Adaptor and low quality (<Q20) sequences were removed using fastq-mcf tools. Cleaned reads were then aligned to the human genome (GRCh37) using HISAT2 version 2.0.5.<sup>81</sup> Read counts were retrieved using featureCounts based on Ensembl gene annotation. Read counts were normalised and differential expression analysis was conducted using DESeq2. The design formula integrated tumour type, controlling for "patient-specific effect". A Wald test identified 424 genes significantly differentially expressed genes (DEGs) with Benjamini-Hochberg (BH) adjusted P-value < 0.05 and fold-change of mean RPKM ≥ 2.

### Pathway, tumour microenvironment and tumour purity analysis

Pathway analysis was performed using an over-representation approach. DEGs were tested against Gene Ontology terms using ClusterProfiler and all annotated genes as background (BH-adjusted P-value < 0.05). The online platform TIMER2.0<sup>82</sup> was used to perform expression-based deconvolution, estimating the proportion of 22 different immune cell types in each tissue sample, using the CIBERSORT-ABS values from the estimation matrix for comparisons (Wilcoxon signed-rank test, P-value < 0.05). The tidyestimate R package (URL: <https://cran.r-project.org/web/packages/tidyestimate/vignettes/using-tidyestimate.html>; based on ESTIMATE algorithm<sup>83</sup>) was used to calculate tumour purity, immune and stromal scores from expression data. To compare our results with TCGA data, tumour purity, immune and stromal scores were downloaded from MD Anderson Cancer Centre curated TCGA database (URL: <https://bioinformatics.mdanderson.org/estimate/disease.html>), using RNA-Seq version 2 to match our platform of data. As recommended, we used ESTIMATE as a measure of tumour purity – higher purity is indicated by a lower ESTIMATE score. For consistent comparisons, a Wilcoxon signed-rank test (P-value < 0.05) was used between different groups to calculate the significance of immune, stromal and purity components.