

1       **Single-cell chromatin accessibility analysis reveals the epigenetic  
2       basis and signature transcription factors for the molecular subtypes  
3       of colorectal cancers**

4

5       Zhenyu Liu<sup>1,2,7</sup>, Yuqiong Hu<sup>1,2,5,6,7</sup>, Haoling Xie<sup>1,4,7</sup>, Kexuan Chen<sup>1,2,7</sup>, Lu Wen<sup>1,2</sup>, Wei Fu<sup>1,3</sup>,  
6       Xin Zhou<sup>1,3\*</sup>, and Fuchou Tang<sup>1,2,4\*</sup>

7

8       <sup>1</sup>School of Life Sciences, Biomedical Pioneering Innovation Center, Department  
9       of General Surgery, Third Hospital, Peking University, Beijing, China

10      <sup>2</sup>Beijing Advanced Innovation Center for Genomics (ICG), Ministry of Education Key  
11      Laboratory of Cell Proliferation and Differentiation, Beijing, China

12      <sup>3</sup>Peking University Third Hospital Cancer Center, Beijing, China

13      <sup>4</sup>Academy for Advanced Interdisciplinary Studies, Peking University, Beijing 100871, China

14      <sup>5</sup>Key Laboratory of Organ Regeneration and Reconstruction, Institute of Zoology, Chinese  
15      Academy of Sciences, Beijing, China.

16      <sup>6</sup>Beijing Institute for Stem Cell and Regenerative Medicine, Beijing, China.

17      <sup>7</sup>These authors contributed equally: Zhenyu Liu, Yuqiong Hu, Haoling Xie, Kexuan Chen

18      \*Correspondence:     Xin       Zhou       (zhouxin@pku.edu.cn)     and     Fuchou     Tang  
19       (tangfuchou@pku.edu.cn)

20

21      **Running title:** Epigenetic classification of CRCs by single-cell ATAC-seq

22

23      **Corresponding author information**

24      Fuchou Tang

25      Addresses: Biomedical Pioneering Innovation Center, Peking University, 5 Yiheyuan Road,  
26      Haidian District, Beijing 100871, China

27      Email: tangfuchou@pku.edu.cn

28      Phone number: +86 10 62744062

29      Xin Zhou

30      Addresses: Department of General Surgery, Third Hospital, Peking University, 49 Huayuan  
31      North Road, Haidian District, Beijing 100191, China

32      Email: zhouxin@pku.edu.cn

33      Phone number: +86 10 82267338

34

35      **Competing interests**

36 The authors declare no competing interests.

37   **Abstract**

38   Colorectal cancer (CRC) is a highly heterogeneous disease, with well-characterized subtypes  
39   based on genome, DNA methylome, and transcriptome signatures. To chart the epigenetic  
40   landscape of CRCs, we generated a high-quality single-cell chromatin accessibility atlas of  
41   epithelial cells for 29 patients. Abnormal chromatin states acquired in adenomas were largely  
42   retained in CRCs, which were tightly accompanied by opposite changes of DNA methylation.  
43   Unsupervised analysis on malignant cells revealed two epigenetic subtypes, exactly matching  
44   iCMS classification, and key iCMS-specific transcription factors were identified, including  
45   HNF4A, PPARA for iCMS2 tumors, and FOXA3, MAFK for iCMS3 tumors. Notably,  
46   subtype-specific TFs bind to distinct target gene sets and contribute to both inter-patient  
47   similarities and diversities for both chromatin accessibilities and RNA expressions. Moreover, we  
48   identified CpG-island methylator phenotypes and pinpointed chromatin state signatures and TF  
49   regulators for CIMP-High subtype. Our work systematically revealed the epigenetic basis of the  
50   well-known iCMS and CIMP classifications of CRCs.

51

52   **Statement of significance**

53   Our work revealed the epigenetic basis of the well-known iCMS and CIMP classifications of  
54   CRCs. Moreover, the inter-patient minor similarities and major diversities of chromatin  
55   accessibility signatures of TF target genes can faithfully explain the corresponding inter-patient  
56   minor similarities and major diversities of RNA expression signatures of CRCs respectively.

57

58   **Keywords**

59   colorectal cancer, scATAC-seq, transcription factors

60 **Introduction**

61 Colorectal cancer forms a highly heterogeneous group of tumors classified into distinct genomic,  
62 epigenomic, and more recently, transcriptomic subtypes (1-4). Among different aspects of CRC  
63 classifications, genetic and transcriptomic changes have been more extensively studied and better  
64 understood (5). In 2015, based on the transcriptome profile of bulk tumor samples, the consensus  
65 molecular subtype (CMS) of CRC was established (1). CRCs are classified into four CMS  
66 categories (CMS1-4) characterized by immune infiltration, WNT and MYC activation, metabolic  
67 deregulation, and mesenchymal infiltration, respectively. However, bulk transcriptome reflects the  
68 averaged gene expression across heterogeneous tumor tissues, thus obscuring the complicated cell  
69 type composition in the tumor microenvironment.

70

71 Recently, by analyzing single-cell RNA sequencing (scRNA-seq) data exclusively derived from  
72 CRC malignant epithelial cells, the intrinsic-consensus molecular subtypes (iCMSs) were  
73 identified (6). iCMS2 cells, corresponding to CMS2 tumors, are characterized by greater WNT  
74 and MYC activation, higher levels of copy number variations (CNVs), and prevalence in the left  
75 colon. Conversely, iCMS3 cells, corresponding to CMS1 and CMS3 tumors, are characterized by  
76 greater MAPK pathway activity, fewer copy number changes, and pervasiveness in the right colon.  
77 Notably, the CMS4 subtype is not observed in the iCMS classification, suggesting that fibrosis is  
78 decoupled from the intrinsic characteristics of malignant epithelial cells. This underscores the  
79 substantial potential of single-cell omics techniques in unraveling the complexities of tumor  
80 subtypes.

81

82 Epigenetic regulatory mechanisms also exert profound influences on phenotypes of tumor cells (7).  
83 Previous studies on epigenetic classification of CRCs have primarily centered on DNA  
84 methylation due to the technical feasibility of DNA methylation assays compared with  
85 chromatin-related assays (8). A subset of CRCs has an exceptionally increased frequency of  
86 hypermethylation at certain CpG islands, forming a distinct subtype referred to as CpG-island  
87 methylator phenotype (CIMP). Based on the degree of genome-wide hypermethylation, CRCs  
88 could be classified into CIMP-High, CIMP-Low, and CIMP-Negative groups (9). The CIMP-High  
89 subtype is associated with a variety of specific clinical and histopathological features and  
90 encompasses almost all cases with *BRAF* mutation (9-11). A recent study utilized single-cell assay  
91 for transposase-accessible chromatin using sequencing (scATAC-seq) to map the regulatory  
92 trajectory of CRC malignant transformation (12). However, this study mainly concentrated on the  
93 phenotypic continuum from healthy colon to carcinoma, with most samples collected from  
94 precancerous polyps, thus leaving the epigenetic characteristics among malignant tumor cells  
95 largely unexplored (12).

96

97 To comprehensively profile the epigenetic subtypes and regulatory landscapes of CRCs at  
98 single-cell resolution, we conducted high-precision scATAC-seq analysis across a spectrum that  
99 included healthy colon tissue, precancerous adenoma, and CRC samples from 29 patients. We  
100 observed that early changes in chromatin accessibility in benign adenomas were predominantly  
101 retained in malignant cells, which was clearly accompanied by DNA methylation changes in the  
102 opposite direction in the same peaks. Unsupervised analysis of scATAC-seq data identified two  
103 distinct epigenetic subtypes of malignant cells, closely mirroring the previously reported

104 transcriptome-based iCMS classifications (iCMS2 & iCMS3). Remarkably, despite considerable  
105 diversities in chromatin accessibility profiles among different patients, the transcription factors  
106 (TFs) associated with each subtype exhibited a striking level of conservation. However, within  
107 each subtype, distinct tumors exhibited varying activations of downstream targets of consistently  
108 activated TFs, thus contributing to intra-subtype patient-specific heterogeneous features. In  
109 addition, leveraging known markers for the epigenetic CIMP categories, we were able to identify  
110 different CIMP subtypes in our cohort, as well as the chromatin dynamics and transcription factors  
111 associated with CIMP-High program. Our results underscored the fundamental role of abnormal  
112 expressions and activities of transcription factors in shaping different aspects of CRC phenotypes.  
113

## 114 **Results**

### 115 **scATAC-seq profiling of CRC samples.**

116 To systematically profile the epigenetic regulatory landscape of CRC malignant epithelial cells,  
117 we generated high-quality scATAC-seq data from a total of 80 samples from 28 CRC patients and  
118 one patient with only rectal adenoma (Fig. 1A and Supplementary Table 1, Methods). Epithelial  
119 cells were specifically enriched using magnetic-activated cell sorting (MACS) with EpCAM  
120 antibody before library preparation (except P09, Methods). After stringent quality control, 16,778  
121 high-quality cells were retained for downstream analysis, with a mean transcription start site (TSS)  
122 enrichment of approximately 11 (Supplementary Fig. 1A). Epithelial cells, immune cells, and  
123 fibroblasts were clearly identified with the gene activity score of well-established  
124 cell-type-specific markers (Fig. 1B and Supplementary Fig. 1B). Immune cells and fibroblasts  
125 from different patients were well clustered together in low-dimensional subspaces, suggesting  
126 minimal batch effect in our data (Fig. 1B and Supplementary Fig. 1C). Epithelial cell consistently  
127 represented the largest proportion across all sample locations (Supplementary Fig. 1D), with a  
128 significant increase in MACS-enriched samples (Supplementary Fig. 1E), indicating the efficacy  
129 of MACS in enriching epithelial cells.  
130

131 To further validate our cell type annotation based on gene activity scores, we additionally  
132 generated scATAC-seq data from fluorescence activated cell sorting (FACS) sorted B cells ( $CD45^+$   
133  $CD3^- CD19^+$ ), T cells ( $CD45^+ CD3^+$ ), fibroblasts ( $CD90^+ CD45^-$ ), and epithelial cells (EpCAM $^+$   
134  $CD45^-$ ) (Supplementary Fig. 1F, G). FACS-sorted cells exhibited robust clustering congruence  
135 with the manually annotated cell types, and displayed high correlation of genome-wide chromatin  
136 accessibility profiles (Supplementary Fig. 1H, I), confirming the accuracy of our cell type  
137 annotations.  
138

### 139 **Increased inter-patient heterogeneities along malignant transformation of CRC.**

140 We next focused on the epithelial compartment. Unsupervised clustering identified 29 distinct  
141 epithelial cell clusters (C1-29) from a total of 11,249 epithelial cells. Based on the sampling  
142 positions and gene activities, epithelial cells were further categorized into healthy epithelial cells  
143 (C1-2), adenoma cells (C3-4), and malignant cells (C5-29) (Fig. 1C and Supplementary Table 2).  
144 Single-cell copy number variation analysis of scATAC-seq data (13) also confirmed our epithelial  
145 cell type annotation, where chromosome arm level CNVs were observed in the majority of  
146 malignant cells while no CNVs were detected in healthy epithelial cells (Supplementary Fig. 2A,  
147 B). A low level of CNVs was also observed in a subset of adenoma cells. These adenoma CNVs

148 were also observed in the malignant cells (Supplementary Fig. 2C), encompassing a small fraction  
149 of all CNVs in malignant cells. This suggested that chromosome-level copy number alterations  
150 started to emerge in precancerous adenomas, consistent with previous studies (14,15).

151  
152 To explore the genome-wide accessibility of each epithelial cell type, we aggregated cells within  
153 the same cluster to create pseudo-bulk chromatin accessibility tracks for every cluster of cells.  
154 When compared with adenoma, healthy epithelial cells, and immune cells, higher chromatin  
155 accessibilities were consistently observed in all malignant cell clusters near the TSS of known  
156 CRC marker genes, including *MYC* (2), *S100P* (16), and *CDH3* (17,18) (Fig. 1D). The chromatin  
157 accessibility profiles of our pseudo-bulk clusters exhibited a high degree of consistency with bulk  
158 ATAC-seq signals from The Cancer Genome Atlas (TCGA) program (19), further affirming the  
159 reliability of our single-cell ATAC-seq data (Fig. 1D). We also identified cell-type specific marker  
160 peaks for healthy, adenoma and malignant cells. Malignant cells exhibited a significant number of  
161 peaks with increased accessibility, shared across the majority of these 28 CRC patients (Fig. 1E).  
162 This suggested a globally more accessible chromatin structure in malignant cells of CRC patients.  
163

164 Notably, malignant cells in our cohort were mostly clustered by patients, while healthy epithelial  
165 cells from different patients tended to coalesce into shared clusters (Fig. 1C). We utilized adjusted  
166 mutual information (AMI) as a quantitative assessment of the congruence between unsupervised  
167 clusters and patients of origin. The AMI reached 0.918 for malignant cells and was only 0.003 for  
168 healthy colon cells, signifying high levels of patient-specific chromatin profiles in malignant cells.  
169 The AMI for adenoma cells was 0.175, indicating a sharp increase of inter-tumor heterogeneities  
170 among patients during malignant transformation from benign adenoma. To further address this  
171 phenomenon, we computed the similarity of each epithelial cell type from different patients, as  
172 measured by the distance in low-dimensional peak accessibility space. Healthy epithelial cells  
173 were largely homogeneous while malignant cells showed low patient-to-patient similarities.  
174 Adenoma cells, on the other hand, occupied an intermediate position between these two extremes  
175 (Fig. 1F).

176  
177 We further interrogated this phenomenon in the aspect of TF activities. TF activity score was  
178 defined as the normalized genome-wide accessibility on the motif of a specific TF, computed on a  
179 per-cell basis using chromVAR (Methods) (20). In the TF activity space, healthy, adenoma, and  
180 malignant cells could be distinctly separated, with each cell type exhibiting its set of  
181 cell-type-specific TFs. Patient-specific clustering was robustly observed in malignant cells, while  
182 healthy epithelial cells were largely homogeneous (Fig. 1G). Taken together, these results  
183 indicated that epigenetic inter-tumor heterogeneities progressively intensified throughout the  
184 malignant transformation of CRCs.

185

#### 186 **Early chromatin accessibility changes in adenoma were retained in CRC.**

187 CRCs arise from adenomatous polyps. Therefore, the detection of early colorectal adenomas is  
188 crucial for improving survival rates of CRCs. To better understand the chromatin dynamics during  
189 the generation of adenomas, we conducted pairwise differential accessibility analysis of adenoma  
190 and healthy epithelial cells, which revealed 20,044 accessible peaks that were gained and 4,028  
191 genomic peaks that were lost in adenoma cells (Fig. 2A, Supplementary Table 3). The gained

192 peaks in adenoma cells were primarily enriched for intergenic regions and distal enhancers,  
193 whereas lost peaks were enriched for promoters, gene bodies, and proximal enhancers (Fig. 2B),  
194 suggesting that lost peaks might be closely associated with gene silencing events. Notably, lost  
195 peaks displayed a significant enrichment for CpG islands (CGIs), with 19.4% of all lost peaks and  
196 64.6% of lost promoter peaks overlapping with CGIs (Fig. 2C). We also validated the identified  
197 differential peaks using the scATAC-seq data from the continuum of CRC malignant  
198 transformation (Supplementary Fig. 3A, B) (12). Enrichment analysis of TF motifs indicated that  
199 gained peaks in adenomas were enriched for LEF1 and TCF3 motifs, while lost peaks were  
200 enriched for KLF and HOX family motifs (Supplementary Fig. 3C). This indicates that aberrantly  
201 losing activities of KLF and HOX family transcription factors and abnormally gaining activities of  
202 transcription factors LEF1 and TCF3 potentially played important roles for the formation of  
203 adenoma.

204

205 Since adenoma presents a transition state during CRC tumorigenesis, we were interested in  
206 determining whether early chromatin accessibility changes in adenoma were preserved in CRCs.  
207 To address this, we identified differentially accessible peaks of each malignant cluster and  
208 compared them with those identified in adenomas. For most differential peaks in adenomas,  
209 similar levels of chromatin accessibilities were consistently observed across the majority of CRCs  
210 (Fig. 2D). However, adenoma-derived differential peaks constituted only a small fraction (11.6%)  
211 of all CRC-related chromatin programs (Fig. 2E), with an average of 88.4% of the differential  
212 peaks specific to CRCs.

213

214 We also performed similar analyses on gene activity scores and identified 6 gene clusters with  
215 distinct expression dynamics through malignant transformation. The majority of the differential  
216 genes identified could be validated in TCGA bulk RNA-seq data or an independent scRNA-seq  
217 dataset of CRC samples (21) (Supplementary Fig. 4A, B). Consistent with chromatin accessibility,  
218 differentially expressed genes in adenomas, both upregulated (G3) and downregulated (G1),  
219 mostly exhibited similar expression patterns in malignant cells (Supplementary Fig. 4C). We  
220 further identified differentially expressed genes consistently detected in all patients in each gene  
221 cluster (Supplementary Fig. 4D, E). Both c-MET and AP-2 $\alpha$ , encoded by MET and TFAP2A genes,  
222 respectively, were known to promote angiogenesis, growth, and metastasis (22-25). Claudin-1,  
223 encoded by the CLDN1 gene, could increase the invasive activity of CRCs and reduce anoikis (26).  
224 These genes were all prevalently activated in early adenomas. On the contrary, HOXA2, the  
225 promoter of which is frequently hypermethylated in most CRC cases (27), was consistently  
226 downregulated in adenomas.

227

228 **Chromatin accessibility changes in adenoma were tightly accompanied by DNA methylation  
229 changes.**

230 Closed chromatin regions in adenomas exhibited a notable enrichment on CGIs and were enriched  
231 for methylation-related KLF family TFs (28-30), implying that DNA methylation might  
232 potentially play a pivotal role in the initiation of early adenoma. To explore the relationships  
233 between chromatin accessibility and DNA methylation, we re-analyzed previously published DNA  
234 methylation array data (31). As anticipated, lost chromatin accessibility peaks, although exhibiting  
235 different absolute DNA methylation ratios, were consistently hypermethylated in adenomas. In

236 contrast, gained peaks underwent a significant DNA demethylation process in adenomas (Fig. 2F,  
237 G). On average, 68.7% of gained peaks and 64.0% of lost peaks were coupled with significant  
238 changes in DNA methylation in the reverse direction, indicating a clear anti-correlation pattern  
239 between chromatin accessibility and DNA methylation during the initiation of adenomas. The  
240 hypermethylation of lost peaks was largely maintained in malignant samples, mirroring the trends  
241 observed in chromatin accessibility (Fig. 2F, G). However, hypomethylation of gained peaks was  
242 partially reversed in a subset of malignant samples (Fig. 2F). This group of CRCs was  
243 characterized by genome-wide hypermethylation, potentially corresponding to the CIMP-High  
244 subtype of CRCs (4).

245

246 We explored deeper into the genes regulated by this anti-correlation of chromatin accessibility and  
247 DNA methylation. Concerning the gained peaks, we noted several genes known to be involved in  
248 CRC progression and showed significant demethylation in the promoters or gene bodies (Fig. 2H).  
249 *AXIN2*, a component of the WNT signaling pathway, is frequently mutated in CRCs and is  
250 associated with DNA mismatch repair and immune suppression (32,33). *KRT80* could activate the  
251 AKT pathway and promote migration and invasion of CRCs (34). *NKD2*, the target of key CRC  
252 driver *ATF1*, was upregulated in CRCs and contributed to CRC cell proliferation (35). All these  
253 genes consistently exhibited increased chromatin accessibility during the transition to adenomas in  
254 all patients (Fig. 2I). However, the dynamics of chromatin accessibility during the transition from  
255 adenomas to CRCs were rather complex. For instance, *AXIN2* remained upregulated in P11 but  
256 was downregulated in malignant cells of the other 3 patients. *NKD2* was downregulated in the  
257 malignant cells of P19 but not in other patients. This underscored the inter-patient heterogeneities  
258 and intricacies of CRC phenotypes.

259

260 Similar analyses were also conducted on lost peaks. HOXA genes, including *HOXA2* and *HOXA5*,  
261 were recognized for an extremely high frequency of hypermethylation in CRCs and excellent  
262 diagnostic ability (27,36). We observed that DNA hypermethylation of *HOXA2* and *HOXA5*  
263 emerged early in adenomas (Fig. 2J), coupled with consistently decreased chromatin accessibility  
264 across different patients (Fig. 2K). Similar patterns were also observed on *CHST2*, a methylated  
265 DNA marker used for the detection of both primary and distantly recurrent CRCs by plasma assay  
266 (37). These results indicated that DNA methylation gain or loss already happened during the  
267 formation of benign adenomas, which were accompanied by loss or gain of chromatin  
268 accessibility accordingly to regulate the expression of key CRC-related genes.

269

## 270 **Unsupervised analysis identified two epigenetic subtypes aligning with iCMS classification.**

271 CRCs have previously been categorized into different subtypes based on DNA methylation (CIMP)  
272 (4,10) and transcriptome profiles (CMS, iCMS) (1,6). We further delved deeper into the  
273 characteristics and subtype classifications of CRC at chromatin accessibility level. Non-negative  
274 matrix factorization (NMF) of all malignant clusters identified 2 distinct subtypes, namely  
275 Group-1 and Group-2 (Fig. 3A, Supplementary Fig. 5A, B and Supplementary Table 4). Both  
276 groups were composed of several patient-specific malignant cell clusters and were clearly  
277 distinguishable from each other in low dimensional embeddings of chromatin accessibility space  
278 (Fig. 3B) and TF activity space (Fig. 3C).

279

We further explored the characteristics of two malignant cell groups and their correlation with clinical traits. Hierarchical clustering of all malignant cells in single-cell CNV space also predominantly aligned with the differentiation of Group-1 and Group-2 cells (Fig. 3D). Group-1 cells exhibited high levels of CNVs, with frequent gains of 7pq, 8q, 13q and 20pq, as well as frequent loss of 1p, 5q, 8p, 14q, 15q, 17p and 18pq, while Group-2 cells were mostly diploid or exhibited fewer CNVs (Fig. 3D). The CNV profile of Group-1 and Group-2 tumors closely resembled that of iCMS2 and iCMS3 (6). Furthermore, Group-1 cancers were all microsatellite-stable (MSS) and predominantly originated from the left colon, whereas Group-2 comprised all microsatellite instable (MSI) cancers, and primarily originated from the right colon (Fig. 3E, Supplementary Fig. 5C), which also closely aligned with the clinical features associated with iCMS2 and iCMS3 (6). These results underscored a robust concordance between our epigenetic classification and the iCMS classification.

292

To provide more conclusive evidence of the association between these two different classifications, we computed gene module scores for iCMS2 and iCMS3 signatures in each single cell using previously identified differentially expressed genes of iCMS subtypes (6) (Methods, Supplementary Table 4). As anticipated, the distribution of iCMS2 and iCMS3 scores mirrored the differentiation between Group-1 and Group-2 cells (Fig. 3F), with Group-1 cells displaying higher iCMS2 scores and Group-2 cells displaying higher iCMS3 scores (Fig. 3G). Expression of key module genes, including *MYC*, *CTSA*, *EIF6* for iCMS2, and *CNKN2A*, *ATL3*, *CLU* for iCMS3, also showed consistent patterns (Supplementary Fig. 5D). The module scores of iCMS2 and iCMS3 showed a negative correlation and formed 2 distinct malignant groups, in agreement with our epigenetic classification of Group-1 and Group-2 cells (Fig. 3H). Taken together, analysis of chromatin accessibility in malignant cells identified two distinct epigenetic subtypes, which exhibit a remarkable alignment with the iCMS classification. Therefore, we referred to Group-1 as iCMS2 whereas Group-2 as iCMS3 hereafter.

306

### 307 Chromatin dynamics and regulatory elements of iCMS subtypes.

308 Since our unsupervised epigenetic subtypes corresponded to iCMS2 and iCMS3 CRCs  
309 respectively, we could leverage our scATAC-seq data to elucidate the chromatin regulatory  
310 mechanisms underlying iCMS classifications. We identified marker peaks for each iCMS subtype  
311 in comparison to healthy colon epithelial cells. Notably, both subtypes were characterized by a  
312 substantial number of gained peaks, with 27,609 and 23,418 peaks more accessible, and 9,195 and  
313 6,438 peaks less accessible in iCMS2 and iCMS3, respectively (Supplementary Fig. 6A). This  
314 suggested a global increase in chromatin accessibility in malignant cells. Differentially accessible  
315 peaks were predominantly specific to each subtype, with only 8,327 peaks commonly gained by  
316 both iCMS subtypes (Fig. 4A), implying distinct mechanisms of chromatin regulation between  
317 these two subtypes. Subtype-specifically gained peaks were mostly shared by the majority of all  
318 clusters within the subtype, except for several clusters with globally lower chromatin accessibility  
319 (Fig. 4B). These clusters were later categorized as CIMP-High subtype (see below).

320

321 We performed functional enrichment of iCMS2-specific ( $n = 19,241$ ), iCMS3-specific ( $n = 15,018$ ),  
322 and commonly gained ( $n = 8,327$ ) peaks using GREAT (38). The iCMS2-specific peaks were  
323 enriched for Gene Ontology (GO) terms including histone arginine methylation, DNA damage

324 checkpoint, and positive regulation of multiple biological processes associated with mitotic cell  
325 cycle. The iCMS3-specific peaks, in comparison, were related to GO terms including immune cell  
326 activation and negative regulation of programmed cell death, consistent with the higher  
327 inflammation response pathway scores of iCMS3 CRCs (6). Commonly gained peaks were mainly  
328 linked to colon epithelial development and functions, encompassing epithelial differentiation,  
329 morphogenesis, and tube development (Fig. 4C).

330

331 Cis-regulatory elements, especially enhancers, play a vital role in the regulation of cancer-related  
332 genes and tumor development. We further interrogated potential regulatory elements of key  
333 determinants of CRCs. *MYC*, a downstream target of the WNT/β-catenin pathway, plays a crucial  
334 role in CRC (39). Nearly all CRCs have changes in *MYC* downstream targets, either promoted or  
335 inhibited by *MYC* (2). An enhancer near the transcription end site (TES) of the *MYC* gene gained  
336 accessibility in both iCMS2 and iCMS3 tumors, potentially contributing to the consistent  
337 activation of *MYC* in most CRC patients. We also noticed that several enhancers, located in the  
338 third exon of *MYC* gene, were specifically activated in iCMS2 tumors (Fig 4D). This might  
339 explain the relatively higher *MYC* activity observed in iCMS2 tumors. We also identified  
340 iCMS2-specific enhancers for *NKDI1*, a component and regulator of WNT signaling pathway  
341 (40,41), and *EREG*, a ligand of EGFR pathway (42). A similar analysis was also performed for  
342 iCMS3-related genes including *ATP1B1*, *TNRSF10B*, and *CLU* (Supplementary Fig. 6B). These  
343 results underscored that activation of subtype-specific enhancers played a critical role in  
344 regulating iCMS-related genes and signaling pathways to shape iCMS-specific phenotypes.

345

#### 346 **Subtype-specific TF regulators of iCMS classification.**

347 Transcription factors are vital regulators of gene expression and cellular phenotypes. To nominate  
348 TFs driving the progression, differentiation, and subtype features of CRCs, we conducted motif  
349 enrichment analysis for iCMS-specific and commonly gained peaks (Supplementary Fig. 6C and  
350 Supplementary Table 5). All peak sets were most significantly enriched for the AP-1 complex  
351 motifs except iCMS2-specific peaks, where HNF4A was the top hit (Supplementary Fig. 6C),  
352 highlighting the crucial role of HNF4A in iCMS2 tumors. We further compared the enrichment of  
353 all significant motifs and identified TFs with different enrichment levels across these three peak  
354 sets by hierarchical clustering (Supplementary Fig. 6D). iCMS2-specific peaks exhibited higher  
355 enrichment for HNF4A, PPARA, and CDX2 motifs, while iCMS3-specific peaks exhibited greater  
356 enrichment for FOXA3, MAFK, and SOX2 motifs. LEF1 and TCF3 motifs were more  
357 significantly enriched in commonly gained peaks, compared with iCMS-specific peaks (Fig. 4E).  
358 The iCMS-specific or shared activities of these TFs were also validated by TF footprinting  
359 analysis on aggregated pseudo-bulks of each group (Supplementary Fig. 6E).

360

361 In addition to motif enrichment in marker peaks of each subtype, we also compared the single-cell  
362 genome-wide accessibility of motifs associated with these TFs. Since motif analysis was always  
363 confounded by other TFs with similar binding sequences, we incorporated the gene activity score  
364 of each TF as an estimation of their expression levels (Supplementary Fig. 7A). Remarkably,  
365 *FOXA3* and *MAFK* displayed high gene activity scores in both iCMS2 and iCMS3 tumors, while  
366 their motif accessibilities were notably more elevated in iCMS3 cells (Fig. 4F). Similar  
367 phenomenon was also observed in *PPARA*, with higher TF activity in iCMS2 cells. These results

368 underscored the critical role of TF binding activity in gene expression regulation, independent of  
369 the absolute expression levels of the corresponding TFs. *HNF4A*, in comparison, consistently  
370 exhibited higher gene activity and motif accessibility specifically in iCMS2 cells, concordant with  
371 the highest level of HNF4A motif enrichment in iCMS2-specific peaks (Supplementary Fig. 6C).  
372 This underscored the significant role of HNF4A as a pivotal regulator and ideal marker for iCMS2  
373 CRCs. Notably, HNF4A resides upstream of WNT signaling by targeting WNT5A (43) and  
374 regulates the WNT/β-catenin signaling pathway through a double-negative feedback loop (44).  
375 The higher activity of HNF4A might contribute to the well-described aberrant activation of WNT  
376 signaling pathway in iCMS2 CRCs. Similarly, overexpression of FOXA3 could reverse HOXC10  
377 silencing-induced inactivation of the MAPK pathway (45). The higher activity of FOXA3 might  
378 thus contribute to the well-known abnormal activation of MAPK signaling pathway in iCMS3  
379 tumors. In summary, the interplay between iCMS-specific TFs and regulatory elements shaped the  
380 distinct phenotypes of different CRC subtypes.

381

382 **iCMS-specific TF activation contributes to intra-iCMS heterogeneities by regulating  
383 patient-specific targets.**

384 Subtype-specific TFs and regulatory elements were shown to regulate iCMS-specific gene  
385 expression programs. However, despite certain shared molecular and clinical characteristics,  
386 tumors within the same iCMS subtype still possess significant levels of intra-subtype  
387 patient-specific features. To explore how intra-iCMS heterogeneities were shaped at chromatin  
388 accessibility levels, we performed motif enrichment analysis in the gained peaks of each  
389 malignant cell cluster. As expected, we found higher enrichment of PPARA, HNF4A, and CDX2  
390 motifs in iCMS2 clusters, while higher enrichment of FOXA3, MAFK, and SOX motifs in iCMS3  
391 clusters (Fig. 4G). Remarkably, the activity of these TFs displayed significant inter-iCMS  
392 variations yet maintained high intra-iCMS consistency. HNF4A and PPARA, for instance, were  
393 highly enriched in almost every individual iCMS2 cluster but showed significantly lower  
394 enrichment in iCMS3 clusters. Contrary patterns were observed for SOX TFs, which were  
395 strongly enriched in iCMS3 clusters, but not in iCMS2 clusters (Fig. 4G). These findings  
396 underscored the pivotal roles played by key TFs in shaping subtype-related phenotypes.

397

398 Given the consistent activation of iCMS-specific TFs in tumors of the same iCMS subtype, we  
399 further interrogated whether they also regulate shared genomic targets in different tumors. To test  
400 this, we annotated potential target peaks containing specific motifs within the gained peaks across  
401 all malignant clusters. Notably, despite the commonly high enrichment of HNF4A in all iCMS2  
402 clusters, the vast majority of more accessible HNF4A targets were cluster-specific (individual  
403 patient-specific), with only 9.7% targets (1,945 of 20,059) consistently shared across different  
404 iCMS2 clusters (shared by over 80% of the clusters) (Fig. 4H). For iCMS3 tumors, similar  
405 phenomena were also observed for SOX2, with 6.3% targets (379 of 6,048) consistently shared  
406 across iCMS3 clusters (Fig. 4H), PPARA and FOXA3 (Supplementary Fig. 7B). This pattern of  
407 intra-iCMS patient-specific TF targeting closely mirrored the pattern of patient-specific RNA  
408 expression programs, as illustrated by re-analyzing the epithelial cell compartment of previously  
409 published scRNA-seq data of CRC samples(21) (Fig. 4I). This indicates that the strong  
410 inter-patient heterogeneities of RNA expression even within the same iCMS subtype can be  
411 faithfully explained by corresponding strong inter-patient heterogeneous activation of target genes

412 even by the same TF in the same iCMS subtype. Only around one-tenth of the target genes of the  
413 same TF were prevalently activated in the majority of the patients in the same iCMS subtype,  
414 which serves as shared core epigenetic signatures of different CRC patients in the same subtype.

415  
416 To provide a more comprehensive picture of intra-iCMS shared TF activation and patient-specific  
417 TF targeting, we identified all consistently activated TFs within each iCMS in our data, defined as  
418 hyper-activated in more than 80% of the clusters of each iCMS subtype (Supplementary Fig. 7C),  
419 as well as their potential target peaks. For iCMS-specific consistently activated TFs, the  
420 percentage of shared genomic targets varies from 6% to 15%, with a median value of 9.1%  
421 (Supplementary Fig. 7D). On average, each TF's targets were shared by only 2 or 3 clusters within  
422 each subtype. These findings suggested that while master regulatory TFs were prevalently  
423 activated and drove iCMS-specific programs, the majority of their target genes largely varied  
424 across different patients, contributing to the intra-iCMS patient-specific heterogeneities of CRCs.  
425

#### 426 **Epigenetic regulation of intra-tumor heterogeneities.**

427 Since epigenetic regulation contributed to both inter- and intra-iCMS heterogeneities of CRCs, we  
428 sought to explore whether similar mechanisms might also contribute to intra-tumor heterogeneities.  
429 Based on single-cell copy number variation analysis, we identified 7 patients harboring multiple  
430 CNV-defined subclones (Supplementary Fig. 8, Methods). 5 of these 7 patients were classified as  
431 iCMS2 subtype, consistent with the fact that iCMS2 tumors exhibited higher frequencies of CNVs.  
432 Intra-tumor subclones were discernible from each other by at least one arm-level CNV feature.  
433 The number of subclones identified in each patient ranged from 2 to 5 (Fig. 5A), where the  
434 smallest subclone comprised 24 single cells exhibiting identical CNV profiles. This also  
435 underscored the efficacy of single-cell approaches in the analysis of intra-tumor heterogeneities.  
436

437 Among all these 7 patients with distinct subclones, P26 exhibited the highest level of intra-tumor  
438 heterogeneities. Malignant cells in P26 comprised 5 CNV-defined subclones (subclone\_1-5)  
439 distributed across two epithelial clusters (C10, C16) with distinct chromatin accessibility profiles  
440 (Fig. 5B, C). All subclones of P26 were classified as the iCMS2 subtype. Copy-number  
441 phylogenetic analysis of 5 subclones indicated that subclones within the C16 cluster, including  
442 subclone 4 and 5, were derived from common ancestor proximal to diploid healthy epithelial cells,  
443 whereas subclones within C10 cluster, including subclone 1, 2, and 3, emerged later during tumor  
444 evolution in this patient (Fig. 5D, Supplementary Fig. 9A). Consistent with phylogenetic  
445 relationships, subclones within C10 cluster accumulated more CNVs (Fig. 5B), and showed higher  
446 iCMS2 gene module signatures (Fig. 5E). This suggested that iCMS-specific gene expression  
447 programs were gradually established during tumor evolution in patient P26.  
448

449 To further characterize the epigenetic mechanisms of tumor evolution, we identified differentially  
450 accessible peaks in each subclone of P26 (Supplementary Fig. 9B). Subclones of C10 cluster  
451 exhibited largely shared chromatin programs, as did subclones of C16 cluster. Concordant with  
452 phylogenetic relationships, subclones of C16 cluster were characterized by a larger number of  
453 regions with altered chromatin accessibility (Fig. 5F). Motif enrichment in each subclone revealed  
454 that non-iCMS-specific TFs, such as LEF1 and TCF3, showed higher enrichment in early-emerged  
455 subclones of C16 cluster. Conversely, iCMS2-specific TFs, Including HNF4A, PPARA, and TR4,

456 showed increased enrichment in later-emerged subclones of C10 cluster (Fig. 5G, Supplementary  
457 Fig. 9C). The further increased activities of iCMS2-specific TFs in subclones of C10 cluster were  
458 further validated by genome-wide TF activity scores (Fig. 5H). This gradual activation of  
459 iCMS-specific TFs aligns with the expression patterns of iCMS module genes, providing insights  
460 into the epigenetic mechanisms underlying the formation of iCMS phenotypes during tumor  
461 evolution in patient P26. We also identified differentially accessible peaks in the subclones of 6  
462 other patients and observed profound intra-tumor heterogeneities on chromatin accessibility  
463 profiles (Supplementary Fig. 10).

464

#### 465 **Inferring CIMP subtypes with scATAC-seq data.**

466 CRCs have been traditionally categorized into transcriptome-based CMSs(1) and DNA  
467 methylation-based CIMP subtypes (4). Our unsupervised analysis of chromatin accessibilities  
468 predominantly aligned with iCMS classification. Therefore, we figured that we could label CIMP  
469 subtypes in our cohort using a supervised approach, utilizing previously established CIMP  
470 markers. We quantified the chromatin accessibility of promoter-related CGIs associated with a set  
471 of 8 well-established CIMP markers (*CACNA1G*, *CDKN2A*, *CRABP1*, *IGF2*, *MLH1*, *NEUROG1*,  
472 *RUNX3*, and *SOCS1*) (10,46) in each malignant cluster. Notably, hierarchical clustering identified  
473 3 distinct groups of CRCs: one hypo-accessible group with lower chromatin accessibility on  
474 almost all 8 marker CGIs, one hyper-accessible group with higher accessibility on most marker  
475 CGIs, and one intermediate group with a comparable number of more accessible and less  
476 accessible markers (Fig. 6A, Supplementary Fig. 11A, B and Supplementary Table 6). Given the  
477 tight anti-correlation between chromatin accessibility and DNA methylation of promoter-related  
478 CGIs (Fig. 2F, G) (12,47,48), the hypo-accessible group exhibited higher DNA methylation level  
479 at all markers and thus corresponded to the CIMP-High subtype, while the hyper-accessible group  
480 corresponded to the CIMP-Negative subtype. The intermediate group, consequently, corresponded  
481 to the CIMP-Low subtype (Fig. 6A). Besides hierarchical clustering, a straightforward analysis of  
482 the frequency of hypo-accessible marker CGIs was sufficient to clearly identify the CIMP-High  
483 subtype (Supplementary Fig. 11B). Furthermore, since the identification of CIMP subtype was  
484 specifically conducted on malignant cells, we also compared the chromatin accessibility of each  
485 marker in each CIMP group with healthy colon epithelial cells. The CIMP-High group exhibited  
486 lower accessibility at all 8 marker CGIs, while the CIMP-Low group showed higher chromatin  
487 accessibility (Fig. 6B). These results effectively dispelled any concerns of potential sampling bias  
488 in our malignant cohort and confirmed the reliability of CIMP subtype identification.

489

490 Notably, tumors of the different CIMP subtypes were not distinctly separated in the  
491 low-dimensional embeddings of peak accessibility space (Fig. 6C), and were also not well  
492 distinguishable in TF activity space (Supplementary Fig. 11C). This emphasized the importance of  
493 employing supervised approaches in the accurate identification of CIMP subtypes. We further  
494 interrogated the association between CIMP classification and unsupervised iCMS classification in  
495 our data. CIMP-High cells were almost evenly distributed across iCMS categories, while  
496 CIMP-Negative cells were exclusively represented in the iCMS2 category (Fig. 6D,  
497 Supplementary Fig. 11D). This underscored the complex associations between CIMP and iCMS  
498 classifications.

499

500 **Differential chromatin landscape and TF regulators of CIMP subtypes**

501 To elucidate the chromatin characteristics associated with CIMP classification, we conducted  
502 pairwise differential accessibility analysis between each CIMP subtype and healthy colon  
503 epithelial cells. Notably, all 3 groups displayed a substantially larger number of gained peaks  
504 compared to lost peaks, even for the CIMP-High group (Supplementary Fig. 12A). However, the  
505 CIMP-High group exhibited a significantly lower number of gained peaks ( $n = 14,292$ ) in  
506 comparison to CIMP-Low ( $n = 26,635$ ) and CIMP-Negative ( $n = 26,385$ ) groups. This observation  
507 aligned with the large-scale DNA hyper-methylation characteristic of the CIMP-High subtype.  
508 Additionally, in contrast to CIMP-Low and CIMP-Negative tumors where the majority of gained  
509 peaks were specific to their respective groups, gained peaks in the CIMP-High group were largely  
510 non-subtype-specific, being shared across other CIMP subtypes (Fig. 6E). However, we still  
511 identified 289 peaks in promoter regions specifically gained in CIMP-High group compared to  
512 CIMP-Low and CIMP-Negative groups, which extended the marker panels for CIMP-High  
513 subtype (Supplementary Table 7).

514

515 We further performed motif enrichment analysis in differentially accessible peaks of each CIMP  
516 subtype (Supplementary Table 8). Gained peaks in all 3 CIMP subtypes were most significantly  
517 enriched for AP-1 complex motifs (Supplementary Fig. 12B). CIMP-Negative tumors were also  
518 highly enriched for the HNF4A motif, which was previously identified as a marker motif for  
519 iCMS2 CRCs (Fig. 3E). This was anticipated since CIMP-Negative tumors were exclusively  
520 composed of iCMS2 cells. We next focused on the TFs specifically associated with the  
521 CIMP-High group, as this subtype was best characterized and was associated with a significant  
522 reduction in cancer-specific mortality(9). TCF motifs, TEAD motifs, and LEF1 motif displayed  
523 higher enrichment in the gained peaks of CIMP-High group (Fig. 6F, Supplementary Fig. 12C).  
524 Higher enrichment of these TFs in CIMP-High tumors was further confirmed through single-cell  
525 gene activity scores, TF motif activities (Fig. 6G), as well as TF footprinting analysis from  
526 aggregated pseudo-bulks (Fig. 6H). As for lost peaks in CIMP-High group, a strong enrichment  
527 for KLF family motifs was observed (Supplementary Fig. 12D and Supplementary Table 8). The  
528 CpG-rich KLF family motifs have been found to be enriched in hyper-methylated differentially  
529 methylated regions (DMRs) in various cancer types (49). Several KLF family TFs have been  
530 identified to be 5mC readers and bind to methylated CpG sites (28-30). This suggested that losing  
531 of activities of KLF family TFs might lead to losing of chromatin accessibility and gaining of  
532 DNA methylation at their target regions, followed by the silencing of these target genes in  
533 CIMP-High CRCs.

534

535 Intriguingly, LEF1 and TCF family motifs, both of which exhibited higher activity in the  
536 CIMP-High group, were previously identified to be highly enriched in the commonly gained peaks  
537 of iCMS2 and iCMS3 CRCs (Fig. 4E, F). In line with this observation, CIMP-High tumors were  
538 almost evenly distributed across these two iCMS categories. This phenomenon suggested that TFs,  
539 which played pivotal regulatory roles in gene expression and cellular phenotype, could serve as  
540 common mechanisms and bridges connecting transcriptome-based and DNA methylation-based  
541 classifications of CRCs.

542

543 **Network analysis identified TF modules associated with multi-layer CRC classifications**

544 In our analysis, multiple TFs were identified to be associated with the iCMS classification or  
545 CIMP classification. Moreover, the connection between these two distinct classifications was also  
546 well characterized in their respective patterns of TF activities. TFs are master regulators of cellular  
547 phenotypes and have been reported to be linked with multiple hallmarks of cancer, including  
548 stemness (50-52), immortality (53,54), epithelial-mesenchymal transition (EMT) (55,56), drug  
549 resistance (57-59) and immune evasion (60-63). In the gastrointestinal system, SOX and CDX  
550 family TFs can drive lineage-specific developmental programs, both in organ specification and  
551 cancer (64). These insights suggested that TF-driven regulation might be the common mechanism  
552 underlying different phenotypic classifications of CRCs. To delve deeper into the TF modules  
553 related to different aspects of CRC phenotypes, we conducted weighted correlation network  
554 analysis on the genome-wide motif activity score of all TFs. After parameter tuning for  
555 approximate scale-free topology (Supplementary Fig. 13A), we identified 13 distinct TF modules,  
556 each with high intra-module correlation on TF activity (Fig. 7A and Supplementary Table 9). The  
557 identified TF modules aligned poorly with TF family annotation, with an AMI of 0.267  
558 (Supplementary Fig. 13B, C). Most TF families were classified into more than one module  
559 (Supplementary Fig. 13D). This indicated that the correlation network effectively captured the  
560 co-activation of TFs and was not strongly affected by the similarity of motifs from the same TF  
561 family.

562

563 We next interrogated the relationship between TF modules and other phenotypic traits, including  
564 clinical features, iCMS classification, and CIMP classification. The iCMS classification showed a  
565 significant association with the activities of module 5 and module 8 (Fig. 7 B, C). These modules  
566 were also linked to the left/right-sided origin of the tumor, which was anticipated given the tight  
567 correlation between iCMS classification and the location of CRC origin. The CIMP classification  
568 was significantly associated with module 8 and module 11. Notably, module 8 displayed a  
569 significant difference across both iCMS and CIMP subtypes (Fig. 7B, C). This could either result  
570 from the shared regulatory mechanisms underlying these two classifications, or the bias towards  
571 the iCMS2 subtype among CIMP-Negative tumors. We also observed several modules related to  
572 MSI status, particularly with module 9 and module 12 displaying the highest significance (Fig. 7B,  
573 C). These findings suggested that different TF co-activation modules regulated distinct  
574 downstream targets to influence CRC phenotypes in various aspects.

575

576 To further validate TF modules identified in our cohort, we re-analyzed the scATAC-seq data from  
577 the continuum of CRC malignant transformation (12), which included several tumor samples.  
578 After identifying malignant cell clusters, iCMS module signatures were calculated to identify the  
579 iCMS classification of each cluster (Supplementary Fig. 13E). We next calculated the eigengene  
580 score of identified TF modules based on the motif activity of each malignant cell cluster. As  
581 expected, the eigengene scores of module 5 and module 8 were robustly associated with iCMS  
582 classification in the independent published dataset, while other modules showed no significant  
583 correlation (Supplementary Fig. 13F). This confirmed that the TF modules could effectively  
584 distinguish different patient subtypes.

585

586 **Module-trait association and intra-module analysis identified key TFs for CRC phenotypes**

587 To further determine master regulatory TFs associated with each trait, we delved deeper into the  
588 inner structure of each TF module. Module 9 was significantly correlated with MSI status, where  
589 MSI patients showed lower module scores (Fig. 7C). Several members of module 9 had been  
590 previously reported to be associated with the MSI status in CRC (Fig. 7D). Loss of *CDX2*  
591 expression was associated with MSI CRCs (65). Frameshifts in the *TCF7L2* gene were exclusively  
592 observed in MSI cases (66). Copy number amplifications of *HOXA9*, *HOXA11*, and *HOXA13*  
593 genes were prevalent in half of MSS CRCs but were rarely observed in MSI cases (67).  
594 Consistently, motifs of these TFs showed lower enrichment in MSI clusters (Supplementary Fig.  
595 14A). To assess the importance of each TF, we calculated TF membership within the respective  
596 modules, as well as its significance concerning clinical features or subtype classification  
597 (Methods). As anticipated, *CDX2*, *TCF7L2*, and *HOXA* TFs exhibited significant association with  
598 MSI status and high membership score for module 9 (Fig. 7D). Notably, *ZKSCAN1*, a C2H2-type  
599 zinc-finger family TF closely associated with gastric cancer (68) and hepatocellular carcinoma  
600 (69), also displayed high membership in module 9. However, the function of *ZKSCAN1* in CRC  
601 remained largely unexplored.

602

603 The iCMS classification displayed the most significant association with module 8, primarily  
604 comprising SOX and nuclear receptor family TFs. Remarkably, *HNF4A* and *PPARA*, both of  
605 which showed high iCMS2-specific activity in our previous analysis (Fig. 4E-G), emerged as top  
606 hits in module 8 (Fig. 7E). This reinforced the reliability of our module-based analysis. Paralogous  
607 TFs *HNF4G* and *PPARD*, along with several other nuclear receptor family TFs such as *NR2F6*,  
608 *RARB*, and *NR1H2*, displayed a comparable level of association with iCMS classification and  
609 module membership. This highlighted the pivotal role of these TFs in iCMS2 tumors. The activity  
610 of SOX TFs was negatively correlated with that of nuclear receptors in module 8 (Fig. 7E),  
611 consistent with the higher activity of SOX TFs in iCMS3-specific programs (Fig. 4E). Similar  
612 analyses were also conducted on module 5, as well as CIMP classification and related modules  
613 (Supplementary Fig. 14B, C).

614

### 615 **TF modules associated with iCMS-specific expression programs**

616 The iCMS subtypes exhibited subtype-specific expression programs along with TF module  
617 activities. To gain a more comprehensive understanding of TF-regulated subtype phenotypes, we  
618 examined the correlation between TF modules and iCMS-specific gene expression programs.  
619 Several TF modules showed significant associations with iCMS signatures. Notably, different TF  
620 modules were linked to entirely distinct sets of signature genes. Module 5, module 6, and module  
621 8 were all correlated with iCMS2 signatures. Specifically, module 5 and module 6 were solely  
622 associated with the downregulated signature genes of iCMS2 tumors, while module 8 exclusively  
623 correlated with the upregulated signature genes of iCMS2 (Fig. 7F, G). This suggests that module  
624 5 and 6 TFs mainly repress the genes specifically downregulated in iCMS2 patients whereas  
625 module 8 TFs mainly activate the genes specifically upregulated in iCMS2 patients. In  
626 concordance with this, multiple upregulated signature genes of iCMS2 subtype, including *LAMP2*,  
627 *ASXL1*, *FN1*, *TOP1MT*, and *CDK5RAP1*, were direct target genes of module 8 TF *HNF4A* in  
628 intestinal epithelial cells (70). Similarly, module 8 and module 9 were exclusively associated with  
629 the downregulated signatures of iCMS3, whereas module 5 and 6 strongly correlated with the  
630 upregulated signatures of iCMS3 (Fig. 7F, Supplementary Fig. 14D). Several upregulated

631 signature genes of iCMS3 subtype, including *KLK11*, *STIM2*, *C4BPB*, *C9orf3*, were direct target  
632 genes of module 5 TF HOXB7 in breast cancer (71). These results elucidate that different TF  
633 modules regulate distinct downstream targets to shape the iCMS-specific expression programs.

634

### 635 Discussion

636 The molecular classification of CRCs has been extensively characterized from multiple  
637 dimensions. However, most previous work has been performed on a single omics of bulk samples,  
638 which makes it challenging to distinguish the intrinsic characteristics of malignant cells from  
639 microenvironment cell factors, as well as to uncover the connections between different aspects of  
640 tumor classifications. Although single-cell RNA sequencing data remarkably improved the CMS  
641 classification, the regulatory mechanisms underpinning iCMS-related gene expression programs  
642 remained largely unknown. Our high-quality scATAC-seq atlas of CRC epithelial cells provided a  
643 comprehensive regulatory landscape at single-cell resolution, which enabled us to analyze  
644 multi-dimensional, subtype-specific epigenomic regulations, and expand previous work on the  
645 transcriptome and DNA methylome characteristics of CRCs.

646

647 Analysis of chromatin accessibility and TF activity revealed that inter-tumor heterogeneities  
648 gradually increased along CRC development. Malignant tumor cells predominantly exhibited  
649 patient-specific features, and these features started to be established in precancerous adenomas.  
650 Early chromatin dynamics during the transition to adenomas were notably enriched in CGIs and  
651 closely linked to DNA methylation changes in the opposite direction. This multi-dimension  
652 regulatory mechanism impacted many key CRC driver and suppressor genes, offering potential  
653 targets for the early screening of CRCs.

654

655 Unsupervised analysis of malignant cells identified two major cell states, which perfectly matched  
656 the iCMS classification. This confirmed that the iCMS classification is robust and could be  
657 reflected in the epigenetic space. We further identified cis-regulatory elements associated with key  
658 cancer genes, as well as iCMS-specific TFs that regulated distinct signaling pathways (WNT  
659 signaling pathway for iCMS2, MAPK signaling pathway for iCMS3) and shaped phenotypes of  
660 different iCMS subtypes. Notably, tumors within the same subtype exhibited consistent activation  
661 of specific TFs, while their downstream targets varied significantly among patients, potentially  
662 contributing to intra-subtype patient-specific features, with only about one-tenth of the abnormally  
663 activated targets shared in the majority of the patients in the same iCMS subtype. More  
664 importantly, this faithfully mirrored and potentially regulated the diverse patient-specific  
665 expression patterns of tumorigenesis genes, with only around one-tenth of the aberrantly activated  
666 genes pervasively shared in the majority of the patients in the same iCMS subtype. So, both the  
667 inter-patient minor similarities and major diversities of RNA expression signatures of CRCs can  
668 be faithfully explained by corresponding inter-patient minor similarities and major diversities of  
669 chromatin accessibility signatures of TF target genes. More importantly, although diverse target  
670 genes were activated by the same TF in different patients, common TFs were prevalently activated  
671 in different patients. This means that repressing the activities of these pervasively activated TFs  
672 may lead to relatively uniform anti-tumor responses in a wide variety of patients, suggesting that  
673 the prevalently activated TFs in CRCs may serve as potentially promising anti-tumor targets. In  
674 addition, we characterized the regulatory landscape across the CIMP classification, especially for

675 CIMP-High tumors. These results provided profound insights into the epigenetic regulatory  
676 mechanisms of CRC subtypes.

677

678 Another major contribution of our study is that we identified TF co-activation modules associated  
679 with multi-aspect CRC classifications. Several TF modules showed significant subtype-specific  
680 activity. We also illustrated that different TF modules regulated distinct sets of subtype signature  
681 genes to shape iCMS-related gene expression programs. Similar modules and key TFs associated  
682 with CIMP classification and MSI status were also identified. TFs, which served pivotal functions  
683 in regulating gene expression and shaping cellular phenotypes, provided common mechanisms and  
684 connections underlying multi-dimension CRC classifications.

685

686 Accumulating evidence indicates that the aberrant reactivation of developmental programs plays a  
687 critical role in tumorigenesis. The initiation of intestinal tumors is associated with the  
688 de-differentiation of epithelial cells and the acquisition of stem-cell-like properties (72). TFs drive  
689 lineage specification events in development. However, whether TFs are associated with the  
690 abnormal reactivation of developmental programs in cancer remains unknown. The iCMS  
691 classification exhibits a clear connection with lineage specification programs, as iCMS2 tumors  
692 mainly originate from the hindgut-derived left colon, while iCMS3 tumors mainly originate from  
693 the midgut-derived right colon. Many of the iCMS-specific TFs identified in our data, especially  
694 SOX, CDX, and HOX TFs, are also master regulators along the anterior-posterior (A-P) axis of  
695 the gut tube during development (73). Strikingly, the TF activities along the proximal-distal axis  
696 of iCMS closely parallel that along the A-P axis of the gut tube during development. Specifically,  
697 *Sox2* and *Cdx2* are exclusively expressed in the anterior and posterior regions of the primitive gut,  
698 respectively (73). Conditional depletion of *Cdx2* has been shown to cause severe hindgut  
699 abnormalities (74). Corresponding to this, CDX2 showed higher activities in left-colon-enriched  
700 iCMS2 tumors, while SOX TFs exhibited higher enrichment in right-colon-enriched iCMS3  
701 tumors (Fig. 4E, G). Previous work has also illustrated the lineage-specific regulatory role of *Sox2*  
702 and *Cdx2* in gastric and intestinal cancer (64). These findings emphasized that aberrant  
703 reactivation of lineage-specific developmental programs in adult colon epithelium might play  
704 crucial roles in different subtypes of CRCs.

705

706 Dysregulation of TF activity has been observed in numerous cancer types and has been confirmed  
707 to be associated with multiple hallmark properties of cancers. Mutated or dysregulated TFs thus  
708 represent a unique class of drug targets. Various approaches have been developed to target TF  
709 activities, including inhibiting protein-protein interactions with cofactors, disrupting DNA binding  
710 activities, and modulating proteasomal degradation of transcription factors (75). We anticipated  
711 that our single-cell regulatory landscape would provide useful resources for identifying important  
712 regulatory elements and TFs across multi-dimension classifications of CRCs. These findings may  
713 serve as novel targets for CRC prevention, screening, and personalized medicine.

714

## 715 Materials and Methods

### 716 Sample collection

717 Samples in this study were collected from 28 patients pathologically diagnosed with CRC and one  
718 patient with only rectal adenoma (P20) from the Department of General Surgery, Peking

719 University Third Hospital. This study received approval from the Ethics Committee of Peking  
720 University Third Hospital (License No. IRB00006761-M2016170) and was conducted following  
721 ethical guidelines (Declaration of Helsinki). Written informed consent was obtained from all  
722 patients. Primary tumors and matched adjacent normal tissues were obtained from 28 CRC  
723 patients. For 12 of the patients (P02, P03, P05, P08, P10, P15, P18, P22, P23, P24, P28, P29), two  
724 samples from the same tumor mass were collected. For one patient (P26), three samples from the  
725 same tumor mass were collected. Adenoma samples were collected from 7 patients (P04, P11, P12,  
726 P13, P19, P20, P28). For CRC patients, adenoma samples were surgically collected together with  
727 tumor samples from polyps located 5-10 cm from the tumor border. For P20, adenoma sample was  
728 collected from rectal adenoma undergoing transanal local resection.

729

### 730 **Single cell preparation**

731 The mucosa layer of the unaffected tissues was dissociated mechanically and washed with PBS.  
732 Then unaffected tissues, adenoma tissues, and CRC tissues were dissected and digested into  
733 single-cell suspension via 1.5 mg/ml collagenase type II (Gibco, 17101015) and 1.5 mg/ml  
734 collagenase type IV (Gibco, 17104019) treatment for 30 min at 37°C. The cell suspension was  
735 stained with anti-human CD326 (EpCAM) microbeads (Miltenyi Biotec, 130-061-101) and then  
736 enriched for epithelial cells by magnetic-activated cell sorting (MACS). Samples from P09 were  
737 not enriched by MACS. The collected cells were stored at -80°C in QuickFreezing-M cell freezing  
738 medium for further usage.

739

740 CD45 antibody (BioLegend, 368512), CD3 antibody (BioLegend, 300306), CD19 antibody  
741 (Biolegend, 392504), CD90 antibody (BioLegend, 328110), EpCAM antibody (BioLegend,  
742 369810) were used in FACS. CD3+CD45+ T cells, CD19+ CD45+CD3- B cells, CD90+CD45-  
743 fibroblast cells, and EpCAM+CD45- epithelial cells were collected to verify the accuracy of cell  
744 type identification by snATAC-seq.

745

### 746 **Library construction of snATAC-seq**

747 Cells (100,000-200,000 cells before freezing) were treated with Omni-ATAC lysis buffer (10 mM  
748 Tris-HCl (ph7.4), 10 mM NaCl, 3 mM MgCl<sub>2</sub>, 0.1% IGEPAL CA630, 0.1% Tween-20, 0.01%  
749 digitonin) to isolate nuclei. The fragmentation was performed by adding 50 µL of fragmentation mix  
750 (33 mM Tris-acetate (pH 7.8), 66 mM Potassium acetate, 10 mM Magnesium acetate, 16%  
751 Dimethylformamide, 0.01% digitonin, and 10 µL of Tn5 transposome complex (Novoprotein,  
752 N233)) to the nuclei and incubating at 37°C for 30 min, 800 rpm. The reaction was stopped by the  
753 addition of 50 µL of fragmentation stop buffer (10 mM Tris-HCl (ph8.0), 20 mM EDTA). Then  
754 single nucleus was flow-sorted into each well of a 96-well plate containing 4 µL of lysis buffer  
755 (50 mM Tris-HCl (ph8.0), 50 mM NaCl, 20 µg/mL Proteinase K (ZYMO, D3001-2), 0.2% SDS,  
756 2.5 µM S5xx Nextera Index Primer, 2.5 µM N7xx Nextera Index Primer) and then incubated at  
757 65 °C for 15 min to disassociate the transposase enzyme. SDS was quenched with 4 µL of 10%  
758 Tween-20. The libraries were amplified by the addition of 10 µL of 2×KAPA HiFi HotStart  
759 ReadyMix (Roche, KK2601), and the PCR was cycled as follows: 72°C for 10 min, 98°C for 5  
760 min, then 20 cycles of 98°C for 10 s, 63°C for 30 s, 72°C for 20 s and finally 72 °C for 5 min.  
761 DNA fragments of size between 200 bp and 700 bp were selected with agarose gel electrophoresis.  
762 Finally, the libraries were sequenced on the NovaSeq 6000 system.

763

#### 764 scATAC data processing

765 Raw sequencing data were trimmed for adaptors using cutadapt (v3.2) (76) and then aligned to the  
766 GRCh38 (hg38) reference genome using hisat2 (v2.2.1) (77). Duplicated reads were removed  
767 using Picard Toolkit (v2.19.0, <https://github.com/broadinstitute/picard>). The bam files were  
768 converted to fragment files using sinto (v 0.7.2.2, <https://github.com/timoast/sinto>) and then  
769 loaded into R (v 4.1.0) using the createArrowFiles function from the R software package ArchR  
770 (v1.0.2) (78). Gene activity score was also computed as construction of the arrow files. The  
771 downstream analysis was performed predominantly using ArchR. Quality control metrics were  
772 computed for each single cell. Only cells with a TSS enrichment score > 5 and the number of  
773 unique fragments > 2,000 were retained. A cell-by-tile matrix was first computed with 500-bp tiles.  
774 Dimensional reduction using 2 rounds of latent semantic indexing (LSI) was computed on the tile  
775 matrix using the addIterativeLSI function of ArchR. Following LSI, initial clustering was  
776 performed using the addClusters function of ArchR. Pseudo-bulk replicates were generated for  
777 each cluster using the addGroupCovariations function of ArchR and used as input for peak calling.

778

779 For each unsupervised cluster, peak calling was performed using MACS2 (v 2.2.7.1) (79). After  
780 identifying the peak set for each cluster, a union peak set was generated using the iterative overlap  
781 peak merging procedure of ArchR. This step was performed using the addReproduciblePeakSet  
782 function of ArchR. After peak calling, a cell-by-peak matrix was computed, followed by  
783 dimensional reduction using 2 rounds of LSI and unsupervised clustering. UMAP embedding was  
784 then computed in the first 30 dimensions of LSI from the peak matrix. For each unsupervised  
785 cluster from the peak matrix, pseudo-bulk replicates were computed using the  
786 addGroupCovariations function of ArchR. Reads-in-peaks-normalized bigwig files were then  
787 generated using the getGroupBW function, with a tile size of 30 bp and a maximum of 2000 cells  
788 per cluster.

789

#### 790 ChromVar deviation for TF activity

791 Catalog of Inferred Sequence Binding Preferences (CIS-BP) motifs from chromVar motifs  
792 ‘human\_pwms\_v1’ were matched to scATAC peaks using the matchMotif function of the  
793 motifmatchr package. The TF accessibility profiles were computed using the R package chromVar  
794 (v 1.16.0) (20) with the addDeviationsMatrix function of ArchR, which generated a cell-by-motif  
795 matrix with GC bias-corrected chromVar deviation scores utilized as metrics for global TF  
796 activity.

797

#### 798 Inferring copy number variation

799 We utilized a previously published method to identify CNVs from scATAC-seq data (13). The  
800 genome was fired tiled into 10-Mb windows with a step size of 2 Mb using the slidingWindows  
801 function of GenomicRanges. The sliding windows were filtered against ENCODE hg38 blacklist  
802 with the setdiff function of R. Insertions in each window were counted for each cell and generated  
803 a cell-by-window binarized matrix. Insertion per bp was then computed with each filtered window.  
804 The GC content for each cluster was then computed for each filtered window using the BSgenome  
805 of hg38 in R. To identify potential CNVs in each window, we identified 100 other windows with  
806 the highest similarity on GC content and computed the average log2 fold change against the 100

807 nearest neighbors. The log<sub>2</sub> fold change was visualized on heatmaps to identify CNVs. This  
808 pipeline has been validated with matched whole exome sequencing data(13).

809

#### 810 **Differential accessibility analysis**

811 Marker peaks of healthy epithelial, adenoma, and malignant cells were identified using the  
812 getMarkerFeatures function of ArchR in a one-versus-all manner. TSS enrichment and log<sub>10</sub>  
813 number of unique fragments were accounted for differences in data quality amongst the cell  
814 groups. Differential accessibility test was performed using Wilcoxon rank sum test. Significant  
815 peaks were selected with a threshold of false discovery rate (FDR) < 0.001 and the absolute value  
816 of log<sub>2</sub> fold change > 0.5.

817

818 Pair-wise differential test between clusters or cell types was performed with the  
819 getMarkerFeatures function of ArchR. TSS enrichment and log<sub>10</sub> number of unique fragments  
820 were accounted for differences in data quality amongst the cell groups. Pair-wise test was  
821 performed using Wilcoxon rank sum test. Peaks with FDR < 0.05 and the absolute value of log<sub>2</sub>  
822 fold change > 0.5 were identified to be significant.

823

#### 824 **Analysis of TCGA bulk RNA-seq data**

825 The differential gene expression analysis of TCGA bulk data was performed using the online  
826 toolkit of GEPIA2 (80). Briefly, bulk RNA-seq data of patient-paired tumor and normal samples  
827 were extracted from the TCGA COAD cohort. Differential gene expression analysis was  
828 performed using LIMMA (81). Genes with adjusted p value < 0.05 the absolute value of log<sub>2</sub> fold  
829 change > 0.25 were identified to be significant.

830

#### 831 **Analysis of CRC 10X scRNA-seq data**

832 We re-analyzed the 10X scRNA-seq data of CRC samples generated by Lee et al (21). Processed  
833 UMI count matrix and cell metadata were downloaded from NCBI Gene Expression Omnibus  
834 (GSE132465) and loaded into R environment using the software package Seurat (v4.3.0) (82).  
835 Cell types were annotated based on the provided metadata. All non-malignant epithelial cells,  
836 including goblet, intermediate, mature enterocytes type 1 and type 2, and stem-like/TA cells were  
837 collectively categorized as healthy epithelial cells.

838

839 Differential analysis between malignant cells and healthy cells was performed using the  
840 FindMarkers function of Seurat. Differentially expressed genes were identified for each single  
841 patient, respectively. Genes with adjusted p value < 0.05 the absolute value of log<sub>2</sub> fold change >  
842 0.5 were identified to be significant.

843

#### 844 **Analysis of CRC scATAC-seq continuum**

845 We re-analyzed the 10X scATAC-seq data of the continuum of CRC malignant transformation  
846 generated by Becker et al (12). Processed fragment files representing the location of Tn5  
847 insertions were downloaded from NCBI Gene Expression Omnibus (GSE201349) and processed  
848 with similar procedures mentioned above. The scATAC-seq data contained an abundant resource  
849 for adenomas but included only 4 CRC samples from non-FAP donors and one CRC sample from  
850 FAP donor. Adenomas were classified into early and late stages according to the chromatin

851 profiles along the trajectory of malignant transformation. Early adenomas showed minimal  
852 difference compared with healthy colon cells, while a significant number of differential peaks  
853 were identified in late adenomas. Differential peaks showed in supplementary figures were  
854 calculated between late adenomas and healthy colon cells.

855  
856 Differential accessible analysis of 10X scATAC-seq was performed similarly to that in our data  
857 but with a lower threshold due to the lower quality and sequencing depth of 10X data. Peaks with  
858 FDR < 0.05 and the absolute value of log<sub>2</sub> fold change > 0.1 were identified to be significant.

859  
860 **Enrichment analysis for genomic region set overlapping**  
861 We first collected known genome region annotations. The CGI and repeat elements were  
862 downloaded from the UCSC table browser. Gene annotations were loaded from UCSC TxDb gene  
863 annotation of the hg38 genome in R using the GenomicFeatures package. The cis-regulatory  
864 elements annotation (cCRE) was downloaded from ENCODE.

865  
866 The significance of overlapping between different genomic region sets was evaluated using the R  
867 package LOLA (v 1.19.1) (83). Briefly, after universe filtering and counting overlaps between  
868 query sets and reference sets, Fisher's exact test was utilized to evaluate the significance of region  
869 overlapping.

870  
871 **Mapping DNA methylation array data to scATAC-seq peaks**  
872 The processed data (series matrix), as well as probe annotation files for CRC methylation array  
873 profiling, was downloaded from NCBI Gene Expression Omnibus with accession number  
874 GSE48684. The data quality of the array data was evaluated with the R software package Minfi  
875 (v1.40.0) (84). The probe annotation was based on the hg19 genome and was converted to hg38  
876 coordinates using the liftOver function from the rtracklayer package. The probe was loaded as  
877 Grange object and then overlapped with scATAC peaks using the countOverlaps function from the  
878 GenomicRanges package. If more than 1 probe were mapped to the same peak, the methylated  
879 ratio of all probes was averaged as measurements for peak-level methylation ratio.

880  
881 **Non-negative matrix factorization for malignant clusters**  
882 We utilized R package NMF (v0.26) (85) for unsupervised subtype classification on malignant  
883 clusters. The average chromatin accessibility of each malignant cluster was summarized using the  
884 getGroupSE function from ArchR. The top 10,000 peaks with the highest standard deviation  
885 across all clusters were selected as feature input for NMF. We tested the NMF model with a  
886 sequence of different ranks and selected rank equals 2 for the final model. This is based on the  
887 combined consideration of a steep drop-off for the Cophenetic correlation coefficient, as well as  
888 the biological interpretability of the model. 200 NMF runs were calculated independently, and the  
889 consensus clustering results of 200 NMF runs were retained as the final output.

890  
891 **Gene module score for iCMS programs**  
892 Marker genes for iCMS2 and iCMS3 subtypes were provided in previous research (6). We  
893 obtained 4 distinct gene sets: upregulated genes and downregulated genes for iCMS2 and iCMS3,  
894 respectively. For each gene set, the single-cell modules score was computed on the imputed gene

895 activity matrix. Briefly, the genes were binned based on average gene activity scores. A random  
896 control set of genes was selected within the same bins with given genes. Gene module score was  
897 defined as the average gene activity score of a given gene set subtracted by the aggregated activity  
898 of control gene sets. Since both upregulated and downregulated genes were provided, the iCMS  
899 module scores were computed as the module score of upregulated genes subtracted by that of  
900 downregulated genes.

901

## 902 **GREAT functional enrichment of genomic regions**

903 The functional enrichment of the peaks set was performed with the online tool GREAT (v4.0.4,  
904 <https://great.stanford.edu/public/html/index.php>) (38). Bed files were generated and uploaded to  
905 the GREAT web server. The whole genome was used as background regions. GREAT analysis was  
906 performed in the basal plus extension mode with 5 kb upstream and 1 kb downstream, plus distal  
907 region up to 1000 kb. Since the peaks hit a large fraction of the genes in the genome, the results  
908 were demonstrated with the significant-by-region-based binomial view of GREAT. Only  
909 enrichment results for GO biological processes were downloaded. Subtype-related GO terms were  
910 manually selected for further visualization.

911

## 912 **Motif enrichment from peak set**

913 We use Homer (v4.11.1) (86) to identify enriched motifs in specific genome positions. After  
914 exporting peak sets as bed files, motif enrichment analysis was performed with  
915 findMotifsGenome.pl script with the built-in hg38 (v6.4) motif annotation of homer. The output  
916 results for known motifs were loaded into R for visualization.

917

## 918 **TF footprinting**

919 The CIS-BP motifs were matched to scATAC peaks using the matchMotif function of the  
920 motifmatchr package. After annotating peaks for motifs, peaks with specific TF motifs were  
921 extracted. The aggregated insertion frequency was calculated using pre-calculated pseudo-bulk  
922 replicates. To account for Tn5 bias, the k-mer ( $k = 6$ ) sequence surrounding each Tn5 insertion site  
923 was identified from each pseudo-bulk and converted into a k-mer frequency table using the  
924 oligonucleotide frequency function from the Biostrings package. The genome-wide expected  
925 k-mers were calculated from hg38 BSgenome using the same function. For each motif, the k-mer  
926 frequency matrix represented all possible k-mers across 250 bp upstream and downstream of the  
927 motif center and was generated by iterating over each motif site. A genome-wide k-mer frequency  
928 matrix is also generated using a similar approach. The expected Tn5 insertions were estimated by  
929 multiplexing the k-mer position frequency table by the observed/expected Tn5 k-mer frequency.  
930 The Tn5 bias was subtracted from the footprinting signal. The footprinting analysis was performed  
931 using the getFootprints function of ArchR.

932

## 933 **CNV-based subclone identification and phylogenetic analysis**

934 Tumor subclones were identified based on single-cell CNV profile. All cells from the same patient  
935 were clustered in single-cell CNV space using hierarchical clustering. The number of subclones  
936 was manually defined to ensure that cells within the same subclone shared the same CNV profile,  
937 while different subclones had distinct CNV patterns. Only subclones with more than 20 cells were  
938 retained. 7 out of the 28 patients were identified with more than one CNV-based subclone.

939  
940 After the identification of subclones, single-cell CNV signals from the same subclone were  
941 aggregated into subclone-level CNV profiles. Only chromosome arm-level CNVs were retained  
942 for downstream analysis. The subclone-by-arm copy number matrix was utilized to infer the  
943 evolutionary tree of a given tumor. The phylogenetic tree was constructed using MEDICC2  
944 (v1.0.2) (87) in the total-copy-number mode. The evolutionary order of each subclone was defined  
945 by the relative branch distance from healthy diploid ancestors.  
946

#### 947 **Identify CIMP subtypes**

948 The CIMP subtype was identified using a previously published set of well-established CIMP  
949 markers (*CACNA1G*, *CDKN2A*, *CRABP1*, *IGF2*, *MLH1*, *NEUROG1*, *RUNX3*, and *SOCS1*) (10,46).  
950 We manually overlapped the CGI region and promoters of 8 genes and obtained 8 marker CGIs.  
951 The chromatin accessibility at each marker CGI was summarized with the getGroupSE function of  
952 ArchR. We calculated the z-score for each CGI across 25 malignant clusters and performed  
953 hierarchical clustering on all malignant clusters. Notably, a cluster of samples exhibited lower  
954 chromatin accessibility in all 8 markers. This cluster was labeled as CIMP-High. Similarly, a  
955 cluster with higher chromatin accessibility was labeled as CIMP-Negative. The intermediate  
956 cluster was then labeled as CIMP-Low.

957

#### 958 **Weighted correlation network analysis of TF activity**

959 The weighted correlation network analysis of TF activity was performed using the R package  
960 WGCNA (v1.72-1) (88). The TF activity of each cluster was summarized using the getGroupSE  
961 function of ArchR from the motif deviation matrix generated by chromeVar. The TF-by-cluster  
962 activity matrix was used as input for WGCNA. After testing different parameters, we selected a  
963 soft threshold (power) of 7 to obtain a scale-free topology of the correlation network. The  
964 topological overlap matrix (TOM) was then computed with signed correlation. Based on TOM, the  
965 correlation network was constructed in unsigned mode. TF modules were detected based on the  
966 hierarchical clustering dendrogram of the network with a minimal module size of 20. 13 distinct  
967 TF modules were identified. The module eigengene, defined as the first principal component of a  
968 given module, was utilized as the module score for each cluster. The network was visualized with  
969 R package igraph (v1.4.2).

970

#### 971 **Module-trait association, TF significance, and module membership.**

972 For categorical traits, including MSI status, side, and iCMS category, the variables were converted  
973 to binarize 0 and 1. Next, the Pearson correlation, as well as correlation test, was computed  
974 between each module eigengene and trait. Associations with  $p < 0.05$  were considered to be  
975 significant. We also further performed Wilcoxon rank sum test for MSI status, and iCMS group,  
976 as well as One-Way ANONA for CIMP subtypes, to validate the difference of module eigengene  
977 between different categorical traits.

978

979 The significance of each TF associated with a specific trait was measured by the Pearson  
980 correlation between the TF activity score and the binarized trait variable. Similarly, the module  
981 membership of specific TF was computed as the Pearson correlation between TF activity and  
982 module eigengene.

983

984 **Validation of TF modules by independent dataset.**

985 To validate the robustness of detected TF modules, we projected the defined TF modules onto an  
986 independent scATAC-seq dataset (12). For the new dataset, the iCMS classification of each malignant  
987 cluster was defined using gene module scores of known iCMS signatures. After summarizing the TF  
988 activity for each malignant cluster, the module eigengenes of previously established TF modules were  
989 calculated using the moduleEigengenes function of the WGCNA package. The eigengenes of  
990 iCMS-related TF modules were compared between iCMS subtypes in the independent dataset.

991

992 **Data availability**

993 All sequencing data of scATAC-seq generated in this study have been deposited in the Genome  
994 Sequence Archive for Human (GSA-Human, <https://ngdc.cncb.ac.cn/gsa-human/>) under accession  
995 number HRA000992. Processed fragments files of scATAC-seq have been deposited in the Open  
996 Archive for Miscellaneous Data (OMIX, <https://ngdc.cncb.ac.cn/omix/>) under accession number  
997 OMIX005759. The methylation array data were downloaded from the Gene Expression Omnibus  
998 (GEO) with accession number GSE48684 (ref (31)). The 10X scRNA-seq data of CRCs were  
999 downloaded from GEO with accession number GSE132465 (ref (21)). The scATAC-seq  
1000 continuum of CRC malignant transformation was downloaded from GEO with accession number  
1001 GSE201349 (ref (12)). TCGA bulk ATAC-seq data and RNA-seq were downloaded from the GDC  
1002 data portal (<https://portal.gdc.cancer.gov/>).

1003

1004 **Code availability**

1005 All custom codes including scATAC-seq data processing, subtype identification, and TF module  
1006 analysis are available on GitHub ([https://github.com/liuzhenyu-yyy/CRC\\_Epi\\_scATAC](https://github.com/liuzhenyu-yyy/CRC_Epi_scATAC)) or Code  
1007 Ocean (<https://codeocean.com/capsule/1194257>). The software used in this study is all published  
1008 and cited in the main text or Methods.

1009

1010 **Acknowledgments**

1011 This work was supported by grants from the National Natural Science Foundation of China  
1012 (81972702) and the Beijing Nova Program (2022029). We thank the Flow Cytometry Core at the  
1013 National Center for Protein Sciences at Peking University, particularly Y. Guo and J. Luo, as well  
1014 as the Flow Cytometry Core at Biomedical Pioneering Innovation Center (Beijing), especially Y.  
1015 Chen, for technical help. Part of the bioinformatics analyses were performed on the  
1016 High-Performance Computing Platform at the Center for Life Sciences (Peking University).

1017

1018 **Author contributions**

1019 F.T., X.Z., and W.F. conceptualized the study. Y.H. and K.C. performed the sample preparation and  
1020 library construction experiments. X.Z. acquired human CRC samples and performed pathology  
1021 reviews. Z.L. and H.X. performed the bioinformatics analyses. Z.L. and Y.H. wrote the method  
1022 section. Z.L. and F.T. wrote the manuscript with the help of all authors.

1023

1024 **References**

- 1025 1. Guinney J, Dienstmann R, Wang X, de Reynies A, Schlicker A, Soneson C, *et al.* The  
1026 consensus molecular subtypes of colorectal cancer. Nat Med **2015**;21(11):1350-6 doi

- 1027 10.1038/nm.3967.
- 1028 2. Cancer Genome Atlas N. Comprehensive molecular characterization of human colon and  
1029 rectal cancer. *Nature* **2012**;487(7407):330-7 doi 10.1038/nature11252.
- 1030 3. Issa JP. CpG island methylator phenotype in cancer. *Nat Rev Cancer* **2004**;4(12):988-93  
1031 doi 10.1038/nrc1507.
- 1032 4. Toyota M, Ahuja N, Ohe-Toyota M, Herman JG, Baylin SB, Issa JP. CpG island  
1033 methylator phenotype in colorectal cancer. *Proc Natl Acad Sci U S A* **1999**;96(15):8681-6  
1034 doi 10.1073/pnas.96.15.8681.
- 1035 5. Marusyk A, Janiszewska M, Polyak K. Intratumor Heterogeneity: The Rosetta Stone of  
1036 Therapy Resistance. *Cancer Cell* **2020**;37(4):471-84 doi 10.1016/j.ccr.2020.03.007.
- 1037 6. Joanito I, Wirapati P, Zhao N, Nawaz Z, Yeo G, Lee F, *et al.* Single-cell and bulk  
1038 transcriptome sequencing identifies two epithelial tumor cell states and refines the  
1039 consensus molecular classification of colorectal cancer. *Nat Genet* **2022**;54(7):963-75 doi  
1040 10.1038/s41588-022-01100-4.
- 1041 7. Flavahan WA, Gaskell E, Bernstein BE. Epigenetic plasticity and the hallmarks of cancer.  
1042 *Science* **2017**;357(6348) doi 10.1126/science.aal2380.
- 1043 8. Mazor T, Pankov A, Song JS, Costello JF. Intratumoral Heterogeneity of the Epigenome.  
1044 *Cancer Cell* **2016**;29(4):440-51 doi 10.1016/j.ccr.2016.03.009.
- 1045 9. Ogino S, Noshio K, Kirkner GJ, Kawasaki T, Meyerhardt JA, Loda M, *et al.* CpG island  
1046 methylator phenotype, microsatellite instability, BRAF mutation and clinical outcome in  
1047 colon cancer. *Gut* **2009**;58(1):90-6 doi 10.1136/gut.2008.155473.
- 1048 10. Weisenberger DJ, Siegmund KD, Campan M, Young J, Long TI, Faasse MA, *et al.* CpG  
1049 island methylator phenotype underlies sporadic microsatellite instability and is tightly  
1050 associated with BRAF mutation in colorectal cancer. *Nat Genet* **2006**;38(7):787-93 doi  
1051 10.1038/ng1834.
- 1052 11. Samowitz WS, Albertsen H, Herrick J, Levin TR, Sweeney C, Murtaugh MA, *et al.*  
1053 Evaluation of a large, population-based sample supports a CpG island methylator  
1054 phenotype in colon cancer. *Gastroenterology* **2005**;129(3):837-45 doi  
1055 10.1053/j.gastro.2005.06.020.
- 1056 12. Becker WR, Nevins SA, Chen DC, Chiu R, Horning AM, Guha TK, *et al.* Single-cell  
1057 analyses define a continuum of cell state and composition changes in the malignant  
1058 transformation of polyps to colorectal cancer. *Nat Genet* **2022**;54(7):985-95 doi  
1059 10.1038/s41588-022-01088-x.
- 1060 13. Satpathy AT, Granja JM, Yost KE, Qi Y, Meschi F, McDermott GP, *et al.* Massively  
1061 parallel single-cell chromatin landscapes of human immune cell development and  
1062 intratumoral T cell exhaustion. *Nat Biotechnol* **2019**;37(8):925-36 doi  
1063 10.1038/s41587-019-0206-z.
- 1064 14. Armaghany T, Wilson JD, Chu Q, Mills G. Genetic alterations in colorectal cancer.  
1065 *Gastrointest Cancer Res* **2012**;5(1):19-27.
- 1066 15. Hermsen M, Postma C, Baak J, Weiss M, Rapallo A, Scitto A, *et al.* Colorectal adenoma  
1067 to carcinoma progression follows multiple pathways of chromosomal instability.  
1068 *Gastroenterology* **2002**;123(4):1109-19 doi 10.1053/gast.2002.36051.
- 1069 16. Ding Q, Chang CJ, Xie X, Xia W, Yang JY, Wang SC, *et al.* APOBEC3G promotes liver  
1070 metastasis in an orthotopic mouse model of colorectal cancer and predicts human hepatic

- 1071 metastasis. *J Clin Invest* **2011**;121(11):4526-36 doi 10.1172/JCI45008.
- 1072 17. Milicic A, Harrison LA, Goodlad RA, Hardy RG, Nicholson AM, Presz M, *et al.* Ectopic  
1073 expression of P-cadherin correlates with promoter hypomethylation early in colorectal  
1074 carcinogenesis and enhanced intestinal crypt fission in vivo. *Cancer Res*  
1075 **2008**;68(19):7760-8 doi 10.1158/0008-5472.CAN-08-0020.
- 1076 18. Imai K, Hirata S, Irie A, Senju S, Ikuta Y, Yokomine K, *et al.* Identification of a novel  
1077 tumor-associated antigen, cadherin 3/P-cadherin, as a possible target for immunotherapy  
1078 of pancreatic, gastric, and colorectal cancers. *Clin Cancer Res* **2008**;14(20):6487-95 doi  
1079 10.1158/1078-0432.CCR-08-1086.
- 1080 19. Corces MR, Granja JM, Shams S, Louie BH, Seoane JA, Zhou W, *et al.* The chromatin  
1081 accessibility landscape of primary human cancers. *Science* **2018**;362(6413) doi  
1082 10.1126/science.aav1898.
- 1083 20. Schep AN, Wu B, Buenrostro JD, Greenleaf WJ. chromVAR: inferring  
1084 transcription-factor-associated accessibility from single-cell epigenomic data. *Nat  
1085 Methods* **2017**;14(10):975-8 doi 10.1038/nmeth.4401.
- 1086 21. Lee HO, Hong Y, Etioglu HE, Cho YB, Pomella V, Van den Bosch B, *et al.*  
1087 Lineage-dependent gene expression programs influence the immune landscape of  
1088 colorectal cancer. *Nat Genet* **2020**;52(6):594-603 doi 10.1038/s41588-020-0636-z.
- 1089 22. Trusolino L, Comoglio PM. Scatter-factor and semaphorin receptors: cell signalling for  
1090 invasive growth. *Nat Rev Cancer* **2002**;2(4):289-300 doi 10.1038/nrc779.
- 1091 23. Boccaccio C, Comoglio PM. Invasive growth: a MET-driven genetic programme for  
1092 cancer and stem cells. *Nat Rev Cancer* **2006**;6(8):637-45 doi 10.1038/nrc1912.
- 1093 24. Wang X, Luo X, Chen C, Tang Y, Li L, Mo B, *et al.* The Ap-2alpha/Elk-1 axis regulates  
1094 Sirpalpha-dependent tumor phagocytosis by tumor-associated macrophages in colorectal  
1095 cancer. *Signal Transduct Target Ther* **2020**;5(1):35 doi 10.1038/s41392-020-0124-z.
- 1096 25. Ma Y, Yang Y, Wang F, Moyer MP, Wei Q, Zhang P, *et al.* Long non-coding RNA CCAL  
1097 regulates colorectal cancer progression by activating Wnt/beta-catenin signalling pathway  
1098 via suppression of activator protein 2alpha. *Gut* **2016**;65(9):1494-504 doi  
1099 10.1136/gutjnl-2014-308392.
- 1100 26. Singh AB, Sharma A, Smith JJ, Krishnan M, Chen X, Eschrich S, *et al.* Claudin-1  
1101 up-regulates the repressor ZEB-1 to inhibit E-cadherin expression in colon cancer cells.  
1102 *Gastroenterology* **2011**;141(6):2140-53 doi 10.1053/j.gastro.2011.08.038.
- 1103 27. Li D, Bai Y, Feng Z, Li W, Yang C, Guo Y, *et al.* Study of Promoter Methylation Patterns  
1104 of HOXA2, HOXA5, and HOXA6 and Its Clinicopathological Characteristics in  
1105 Colorectal Cancer. *Front Oncol* **2019**;9:394 doi 10.3389/fonc.2019.00394.
- 1106 28. Liu Y, Olanrewaju YO, Zheng Y, Hashimoto H, Blumenthal RM, Zhang X, *et al.*  
1107 Structural basis for Klf4 recognition of methylated DNA. *Nucleic Acids Res*  
1108 **2014**;42(8):4859-67 doi 10.1093/nar/gku134.
- 1109 29. Hu S, Wan J, Su Y, Song Q, Zeng Y, Nguyen HN, *et al.* DNA methylation presents  
1110 distinct binding sites for human transcription factors. *eLife* **2013**;2:e00726 doi  
1111 10.7554/eLife.00726.
- 1112 30. Spruijt CG, Gnerlich F, Smits AH, Pfaffeneder T, Jansen PW, Bauer C, *et al.* Dynamic  
1113 readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell*  
1114 **2013**;152(5):1146-59 doi 10.1016/j.cell.2013.02.004.

- 1115 31. Luo Y, Wong CJ, Kaz AM, Dzieciatkowski S, Carter KT, Morris SM, *et al.* Differences in  
1116 DNA methylation signatures reveal multiple pathways of progression from adenoma to  
1117 colorectal cancer. *Gastroenterology* **2014**;147(2):418-29 e8 doi  
1118 10.1053/j.gastro.2014.04.039.
- 1119 32. Zhai J, Chen H, Wong CC, Peng Y, Gou H, Zhang J, *et al.* ALKBH5 Drives Immune  
1120 Suppression Via Targeting AXIN2 to Promote Colorectal Cancer and Is a Target for  
1121 Boosting Immunotherapy. *Gastroenterology* **2023**;165(2):445-62 doi  
1122 10.1053/j.gastro.2023.04.032.
- 1123 33. Liu W, Dong X, Mai M, Selan RS, Taniguchi K, Krishnadath KK, *et al.* Mutations in  
1124 AXIN2 cause colorectal cancer with defective mismatch repair by activating  
1125 beta-catenin/TCF signalling. *Nat Genet* **2000**;26(2):146-7 doi 10.1038/79859.
- 1126 34. Li C, Liu X, Liu Y, Liu X, Wang R, Liao J, *et al.* Keratin 80 promotes migration and  
1127 invasion of colorectal carcinoma by interacting with PRKDC via activating the AKT  
1128 pathway. *Cell death & disease* **2018**;9(10):1009 doi 10.1038/s41419-018-1030-y.
- 1129 35. Tian J, Chang J, Gong J, Lou J, Fu M, Li J, *et al.* Systematic Functional Interrogation of  
1130 Genes in GWAS Loci Identified ATF1 as a Key Driver in Colorectal Cancer Modulated  
1131 by a Promoter-Enhancer Interaction. *Am J Hum Genet* **2019**;105(1):29-47 doi  
1132 10.1016/j.ajhg.2019.05.004.
- 1133 36. Baharudin R, Ishak M, Muhamad Yusof A, Saidin S, Syafruddin SE, Wan Mohamad  
1134 Nazarie WF, *et al.* Epigenome-Wide DNA Methylation Profiling in Colorectal Cancer  
1135 and Normal Adjacent Colon Using Infinium Human Methylation 450K. *Diagnostics*  
1136 (Basel) **2022**;12(1) doi 10.3390/diagnostics12010198.
- 1137 37. Xie H, Mahoney DW, Foote PH, Burger KN, Doering KA, Taylor WR, *et al.* Novel  
1138 Methylated DNA Markers in the Surveillance of Colorectal Cancer Recurrence. *Clin  
1139 Cancer Res* **2021**;27(1):141-9 doi 10.1158/1078-0432.CCR-20-2589.
- 1140 38. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, *et al.* GREAT  
1141 improves functional interpretation of cis-regulatory regions. *Nat Biotechnol*  
1142 **2010**;28(5):495-501 doi 10.1038/nbt.1630.
- 1143 39. Liu J, Xiao Q, Xiao J, Niu C, Li Y, Zhang X, *et al.* Wnt/beta-catenin signalling: function,  
1144 biological mechanisms, and therapeutic opportunities. *Signal Transduct Target Ther*  
1145 **2022**;7(1):3 doi 10.1038/s41392-021-00762-6.
- 1146 40. Katoh M, Katoh M. WNT signaling pathway and stem cell signaling network. *Clin  
1147 Cancer Res* **2007**;13(14):4042-5 doi 10.1158/1078-0432.CCR-06-2316.
- 1148 41. Twaroski K, Mallanna SK, Jing R, DiFurio F, Urick A, Duncan SA. FGF2 mediates  
1149 hepatic progenitor cell formation during human pluripotent stem cell differentiation by  
1150 inducing the WNT antagonist NKD1. *Genes Dev* **2015**;29(23):2463-74 doi  
1151 10.1101/gad.268961.115.
- 1152 42. Cheng WL, Feng PH, Lee KY, Chen KY, Sun WL, Van Hiep N, *et al.* The Role of  
1153 EREG/EGFR Pathway in Tumor Progression. *Int J Mol Sci* **2021**;22(23) doi  
1154 10.3390/ijms222312828.
- 1155 43. Chang HR, Nam S, Kook MC, Kim KT, Liu XP, Yao H, *et al.* HNF4 $\alpha$  is a therapeutic  
1156 target that links AMPK to WNT signalling in early-stage gastric cancer. *Gut*  
1157 **2016**;65(1):19-+ doi 10.1136/gutjnl-2014-307918.
- 1158 44. Lv DD, Zhou LY, Tang H. Hepatocyte nuclear factor 4 $\alpha$  and cancer-related cell signaling

- pathways: a promising insight into cancer treatment. *Experimental and Molecular Medicine* **2021**;53(1):8-18 doi 10.1038/s12276-020-00551-1.
- He X, Wang H, Wang R, Li Y, Li S, Cao X, *et al.* HOXC10 promotes esophageal squamous cell carcinoma progression by targeting FOXA3 and indicates poor survival outcome. *Heliyon* **2023**;9(10):e21056 doi 10.1016/j.heliyon.2023.e21056.
- Ogino S, Kawasaki T, Kirkner GJ, Kraft P, Loda M, Fuchs CS. Evaluation of markers for CpG island methylator phenotype (CIMP) in colorectal cancer by a large population-based sample. *J Mol Diagn* **2007**;9(3):305-14 doi 10.2353/jmoldx.2007.060170.
- Pott S. Simultaneous measurement of chromatin accessibility, DNA methylation, and nucleosome phasing in single cells. *Elife* **2017**;6 doi 10.7554/eLife.23203.
- Clark SJ, Argelaguet R, Kapourani CA, Stubbs TM, Lee HJ, Alda-Catalinas C, *et al.* scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. *Nat Commun* **2018**;9(1):781 doi 10.1038/s41467-018-03149-4.
- Detilleux D, Spill YG, Balaramane D, Weber M, Bardet AF. Pan-cancer predictions of transcription factors mediating aberrant DNA methylation. *Epigenetics Chromatin* **2022**;15(1):10 doi 10.1186/s13072-022-00443-w.
- Kuo YH, Landrette SF, Heilman SA, Perrat PN, Garrett L, Liu PP, *et al.* Cbf beta-SMMHC induces distinct abnormal myeloid progenitors able to develop acute myeloid leukemia. *Cancer Cell* **2006**;9(1):57-68 doi 10.1016/j.ccr.2005.12.014.
- Krivtsov AV, Twomey D, Feng Z, Stubbs MC, Wang Y, Faber J, *et al.* Transformation from committed progenitor to leukaemia stem cell initiated by MLL-AF9. *Nature* **2006**;442(7104):818-22 doi 10.1038/nature04980.
- Barabe F, Kennedy JA, Hope KJ, Dick JE. Modeling the initiation and progression of human acute leukemia in mice. *Science* **2007**;316(5824):600-4 doi 10.1126/science.1139851.
- Ptasinska A, Assi SA, Mannari D, James SR, Williamson D, Dunne J, *et al.* Depletion of RUNX1/ETO in t(8;21) AML cells leads to genome-wide changes in chromatin structure and transcription factor binding. *Leukemia* **2012**;26(8):1829-41 doi 10.1038/leu.2012.49.
- Bell RJ, Rube HT, Kreig A, Mancini A, Fouse SD, Nagarajan RP, *et al.* Cancer. The transcription factor GABP selectively binds and activates the mutant TERT promoter in cancer. *Science* **2015**;348(6238):1036-9 doi 10.1126/science.aab0015.
- Little GH, Baniwal SK, Adisetiyo H, Groshen S, Chimge NO, Kim SY, *et al.* Differential effects of RUNX2 on the androgen receptor in prostate cancer: synergistic stimulation of a gene set exemplified by SNAI2 and subsequent invasiveness. *Cancer Res* **2014**;74(10):2857-68 doi 10.1158/0008-5472.CAN-13-2003.
- Pratap J, Javed A, Languino LR, van Wijnen AJ, Stein JL, Stein GS, *et al.* The Runx2 osteogenic transcription factor regulates matrix metalloproteinase 9 in bone metastatic cancer cells and controls cell invasion. *Mol Cell Biol* **2005**;25(19):8581-91 doi 10.1128/MCB.25.19.8581-8591.2005.
- Boregowda RK, Medina DJ, Markert E, Bryan MA, Chen W, Chen S, *et al.* The transcription factor RUNX2 regulates receptor tyrosine kinase expression in melanoma. *Oncotarget* **2016**;7(20):29689-707 doi 10.18632/oncotarget.8822.

- 1203 58. Cauchy P, James SR, Zacarias-Cabeza J, Ptasińska A, Imperato MR, Assi SA, *et al.*  
1204 Chronic FLT3-ITD Signaling in Acute Myeloid Leukemia Is Connected to a Specific  
1205 Chromatin Signature. *Cell Rep* **2015**;12(5):821-36 doi 10.1016/j.celrep.2015.06.069.
- 1206 59. Matkar S, Sharma P, Gao S, Gurung B, Katona BW, Liao J, *et al.* An Epigenetic Pathway  
1207 Regulates Sensitivity of Breast Cancer Cells to HER2 Inhibition via FOXO/c-Myc Axis.  
1208 *Cancer Cell* **2015**;28(4):472-85 doi 10.1016/j.ccr.2015.09.005.
- 1209 60. Elias S, Yamin R, Golomb L, Tsukerman P, Stanietsky-Kaynan N, Ben-Yehuda D, *et al.*  
1210 Immune evasion by oncogenic proteins of acute myeloid leukemia. *Blood*  
1211 **2014**;123(10):1535-43 doi 10.1182/blood-2013-09-526590.
- 1212 61. Cerezo M, Guemiri R, Druillennec S, Girault I, Malka-Mahieu H, Shen S, *et al.*  
1213 Translational control of tumor immune escape via the eIF4F-STAT1-PD-L1 axis in  
1214 melanoma. *Nat Med* **2018**;24(12):1877-86 doi 10.1038/s41591-018-0217-1.
- 1215 62. Xu Y, Poggio M, Jin HY, Shi Z, Forester CM, Wang Y, *et al.* Translation control of the  
1216 immune checkpoint in cancer and its therapeutic targeting. *Nat Med* **2019**;25(2):301-11  
1217 doi 10.1038/s41591-018-0321-2.
- 1218 63. Rakhra K, Bachireddy P, Zabuawala T, Zeiser R, Xu L, Kopelman A, *et al.* CD4(+) T  
1219 cells contribute to the remodeling of the microenvironment required for sustained tumor  
1220 regression upon oncogene inactivation. *Cancer Cell* **2010**;18(5):485-98 doi  
1221 10.1016/j.ccr.2010.10.002.
- 1222 64. Francis R, Guo H, Streutker C, Ahmed M, Yung T, Dirks PB, *et al.* Gastrointestinal  
1223 transcription factors drive lineage-specific developmental programs in organ specification  
1224 and cancer. *Sci Adv* **2019**;5(12):eaax8898 doi 10.1126/sciadv.aax8898.
- 1225 65. Konukiewitz B, Schmitt M, Silva M, Pohl J, Lang C, Steiger K, *et al.* Loss of CDX2 in  
1226 colorectal cancer is associated with histopathologic subtypes and microsatellite instability  
1227 but is prognostically inferior to hematoxylin-eosin-based morphologic parameters from  
1228 the WHO classification. *Br J Cancer* **2021**;125(12):1632-46 doi  
1229 10.1038/s41416-021-01553-0.
- 1230 66. Zheng K, Wan H, Zhang J, Shan G, Chai N, Li D, *et al.* A novel NGS-based microsatellite  
1231 instability (MSI) status classifier with 9 loci for colorectal cancer patients. *J Transl Med*  
1232 **2020**;18(1):215 doi 10.1186/s12967-020-02373-1.
- 1233 67. Lin CH, Lin JK, Chang SC, Chang YH, Chang HM, Liu JH, *et al.* Molecular profile and  
1234 copy number analysis of sporadic colorectal cancer in Taiwan. *J Biomed Sci*  
1235 **2011**;18(1):36 doi 10.1186/1423-0127-18-36.
- 1236 68. Li Y, Zhao Q, Fan LQ, Wang LL, Tan BB, Leng YL, *et al.* Zinc finger protein 139  
1237 expression in gastric cancer and its clinical significance. *World J Gastroenterol*  
1238 **2014**;20(48):18346-53 doi 10.3748/wjg.v20.i48.18346.
- 1239 69. Yao Z, Luo J, Hu K, Lin J, Huang H, Wang Q, *et al.* ZKSCAN1 gene and its related  
1240 circular RNA (circZKSCAN1) both inhibit hepatocellular carcinoma cell growth,  
1241 migration, and invasion but through different signaling pathways. *Mol Oncol*  
1242 **2017**;11(4):422-37 doi 10.1002/1878-0261.12045.
- 1243 70. Boyd M, Bressendorff S, Moller J, Olsen J, Troelsen JT. Mapping of HNF4alpha target  
1244 genes in intestinal epithelial cells. *BMC Gastroenterol* **2009**;9:68 doi  
1245 10.1186/1471-230X-9-68.
- 1246 71. Heinonen H, Lepikhova T, Sahu B, Pehkonen H, Pihlajamaa P, Louhimo R, *et al.*

- 1247 Identification of several potential chromatin binding sites of HOXB7 and its downstream  
1248 target genes in breast cancer. *Int J Cancer* **2015**;137(10):2374-83 doi 10.1002/ijc.29616.
- 1249 72. Schwitalla S, Fingerle AA, Cammareri P, Nebelsiek T, Göktuna SI, Ziegler PK, *et al.*  
1250 Intestinal Tumorigenesis Initiated by Dedifferentiation and Acquisition of Stem-Cell-like  
1251 Properties. *Cell* **2013**;152(1-2):25-38 doi 10.1016/j.cell.2012.12.012.
- 1252 73. Zorn AM, Wells JM. Vertebrate endoderm development and organ formation. *Annu Rev  
1253 Cell Dev Biol* **2009**;25:221-51 doi 10.1146/annurev.cellbio.042308.113344.
- 1254 74. Gao N, White P, Kaestner KH. Establishment of intestinal identity and  
1255 epithelial-mesenchymal signaling by Cdx2. *Dev Cell* **2009**;16(4):588-99 doi  
1256 10.1016/j.devcel.2009.02.010.
- 1257 75. Bushweller JH. Targeting transcription factors in cancer - from undruggable to reality. *Nat  
1258 Rev Cancer* **2019**;19(11):611-24 doi 10.1038/s41568-019-0196-7.
- 1259 76. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads.  
1260 *EMBnetjournal* **2011**;17(1):3 doi 10.14806/ej.17.1.200.
- 1261 77. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and  
1262 genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **2019**;37(8):907-15 doi  
1263 10.1038/s41587-019-0201-4.
- 1264 78. Granja JM, Corces MR, Pierce SE, Bagdatli ST, Choudhry H, Chang HY, *et al.* ArchR is  
1265 a scalable software package for integrative single-cell chromatin accessibility analysis.  
1266 *Nat Genet* **2021**;53(3):403-11 doi 10.1038/s41588-021-00790-6.
- 1267 79. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, *et al.* Model-based  
1268 analysis of ChIP-Seq (MACS). *Genome Biol* **2008**;9(9):R137 doi  
1269 10.1186/gb-2008-9-9-r137.
- 1270 80. Tang Z, Kang B, Li C, Chen T, Zhang Z. GEPIA2: an enhanced web server for large-scale  
1271 expression profiling and interactive analysis. *Nucleic Acids Res*  
1272 **2019**;47(W1):W556-W60 doi 10.1093/nar/gkz430.
- 1273 81. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, *et al.* limma powers differential  
1274 expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*  
1275 **2015**;43(7):e47 doi 10.1093/nar/gkv007.
- 1276 82. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, 3rd, Zheng S, Butler A, *et al.* Integrated  
1277 analysis of multimodal single-cell data. *Cell* **2021**;184(13):3573-87 e29 doi  
1278 10.1016/j.cell.2021.04.048.
- 1279 83. Sheffield NC, Bock C. LOLA: enrichment analysis for genomic region sets and  
1280 regulatory elements in R and Bioconductor. *Bioinformatics* **2016**;32(4):587-9 doi  
1281 10.1093/bioinformatics/btv612.
- 1282 84. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, *et al.*  
1283 Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium  
1284 DNA methylation microarrays. *Bioinformatics* **2014**;30(10):1363-9 doi  
1285 10.1093/bioinformatics/btu049.
- 1286 85. Gaujoux R, Seoighe C. A flexible R package for nonnegative matrix factorization. *BMC  
1287 Bioinformatics* **2010**;11:367 doi 10.1186/1471-2105-11-367.
- 1288 86. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, *et al.* Simple combinations of  
1289 lineage-determining transcription factors prime cis-regulatory elements required for  
1290 macrophage and B cell identities. *Mol Cell* **2010**;38(4):576-89 doi

- 1291 10.1016/j.molcel.2010.05.004.
- 1292 87. Kaufmann TL, Petkovic M, Watkins TBK, Colliver EC, Laskina S, Thapa N, *et al.*  
1293 MEDICC2: whole-genome doubling aware copy-number phylogenies for cancer  
1294 evolution. *Genome Biol* **2022**;23(1):241 doi 10.1186/s13059-022-02794-9.
- 1295 88. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network  
1296 analysis. *BMC Bioinformatics* **2008**;9:559 doi 10.1186/1471-2105-9-559.
- 1297

1298 **Figure 1.** Single-cell ATAC-seq atlas of epithelial cells in colorectal cancer. **A**, Schematic  
1299 representation of the sample collection, sequencing experiment, and bioinformatic analysis of the  
1300 study. **B**, UMAP visualization of all 16,778 single-cells, colored by cell type. **C**, UMAP  
1301 visualization of all epithelial cells, colored by cell type (left), unsupervised clusters (middle), and  
1302 patient (right). **D**, Pseudo-bulk chromatin accessibility tracks of each epithelial cluster,  
1303 non-epithelial cells, and selected TCGA COAD samples on known marker genes of CRCs.  
1304 CRC-specific signals were highlighted in shadows. **E**, Row-scaled peak accessibility heatmap of  
1305 cell-type specific marker peaks in patient-specific pseudo-bulks of epithelial cells. **F**, Distance  
1306 heatmap representing the similarity between healthy epithelial cells, adenoma cells, and malignant  
1307 cells from each patient. **G**, Row-scaled single-cell TF activity heatmap of cell-type marker TFs in  
1308 epithelial cells.

1309 **Figure 2.** Early chromatin regulation in adenoma is tightly associated with DNA methylation. **A**,  
1310 Volcano plot representing the identification of gained (red) and lost (blue) peaks during the  
1311 transition from healthy colon epithelial cells to adenoma cells. **B**, Significance of genomic region  
1312 overlapping between gained peaks (red), lost peaks (blue), and ENCODE cCREs, CGIs, and gene  
1313 annotation. PLS, promoter-like signature; pELS, proximal enhancer-like signature; dELS, distal  
1314 enhancer-like signature; CGI, CpG island. **C**, Venn diagram representing the overlap between lost  
1315 peaks (top), promoters (bottom), and CGIs. **D**, Row-scaled cluster pseudo-bulk accessibility  
1316 heatmaps demonstrating differential peaks between adenoma and healthy colon epithelial cells in.  
1317 **E**, The numbers of gained and lost peaks in each malignant cluster compared with healthy colon  
1318 cells. Peaks presented in adenomas were labeled with different colors. Note that the numbers of  
1319 lost peaks in adenoma were minimal (nearly invisible in the figure). **F**, Heatmap of DNA  
1320 methylation ratio in each gained and lost peak across different sample locations. **G**, Aggregated  
1321 DNA methylation ratio (red) and chromatin accessibility (green) of all lost peaks (top) and gained  
1322 peaks (bottom) along CRC during the transition to adenoma. P values were calculated with  
1323 Wilcoxon test. **H**, DNA methylation ratio at known CRC-related genes associated with gained  
1324 peaks. **I**, Gene activity score at known CRC-related genes associated with gained peaks in each  
1325 patient with paired adenoma and malignant samples. **J**, DNA methylation ratio at known  
1326 CRC-related genes associated with lost peaks. **K**, Gene activity score at known CRC-related genes  
1327 associated with lost peaks in each patient with paired adenoma and malignant samples. P values  
1328 were calculated with Wilcoxon test.

1329 **Figure 3.** Unsupervised analysis identified 2 malignant groups concordant with the iCMS  
1330 classification. **A**, Heatmap representing the consensus clustering result of 200 independent NMF  
1331 runs. The colors of the heatmap represented the frequency that 2 clusters were categorized into the  
1332 same group in 200 independent runs. **B**, UMAP visualization of all epithelial cells in chromatin  
1333 accessibility space, colored by cell type and malignant subtype. **C**, UMAP embedding of  
1334 malignant cells in TF activity space, colored by malignant subtype. **D**, Hierarchically clustered  
1335 heatmap of estimated copy number scores in malignant cells. CNV-defined clusters were  
1336 significantly associated with iCMS classification ( $p < 2.2 \times 10^{-16}$ , Chi-square test). **E**, Breakdown  
1337 of Group-1 and Group-2 tumors by MSI status (top), anatomical side (middle), and gender  
1338 (bottom). Unavailable data were labeled as NA. P values were calculated with Chi-square test. **F**,  
1339 UMAP visualization of gene module score for iCMS2 (left) and iCMS3 (right) program. **G**,  
1340 Comparison of iCMS module scores between healthy cells, Group-1 and Group-2 CRCs  
1341 (Wilcoxon test). **H**, Scatter plot indicating the distribution of iCMS2 and iCMS3 module scores in  
1342 all healthy and malignant clusters (Pearson correlation). The ellipses indicated 90% confidence  
1343 level for multivariate t-distributions.

1344 **Figure 4.** Chromatin and TF regulators of iCMS subtypes. **A**, Upset plot indicating the overlap  
1345 between gained and lost peaks in iCMS2 and iCMS3 tumors. Gained peaks were colored by  
1346 comparison between iCMS2 and iCMS3. **B**, Row-scaled cluster pseudo-bulk accessibility heatmap  
1347 of gained peaks in iCMS2 and iCMS3 tumors. **C**, GREAT functional enrichment for GO  
1348 biological processes of iCMS2-specific (top), iCMS3-specific (middle), and iCMS2&3 common  
1349 (bottom) gained peaks. **D**, Pseudo-bulk chromatin accessibility tracks of all epithelial clusters near  
1350 iCMS2 specific genes. iCMS2-specific regulatory elements were labeled in shadow. **E**, Dot plot  
1351 representing the enrichment of selected TF motifs in 3 sets of peaks (left), as well as expression in  
1352 different subtypes (right). F, UMAP visualization of gene activity score (top) and TF activity score  
1353 (bottom) for hyper-activated TFs in iCMS2-specific peaks (HNF4A, PPARA), common peaks  
1354 (TCF3, LEF1), and iCMS3-specific peaks (FOXA3, MAFK). **G**, Dot plot indicating the  
1355 enrichment of iCMS-related TF motifs in the gained peaks of each malignant cluster. **H**,  
1356 Hierarchically clustered heatmaps for the target peaks of HNF4A (top) and SOX2 (bottom) gained  
1357 in each malignant cluster. **I**, Column-scaled gene expression heatmap showing the upregulated  
1358 genes of iCMS2 (top) and iCMS3 (bottom) in all patients.

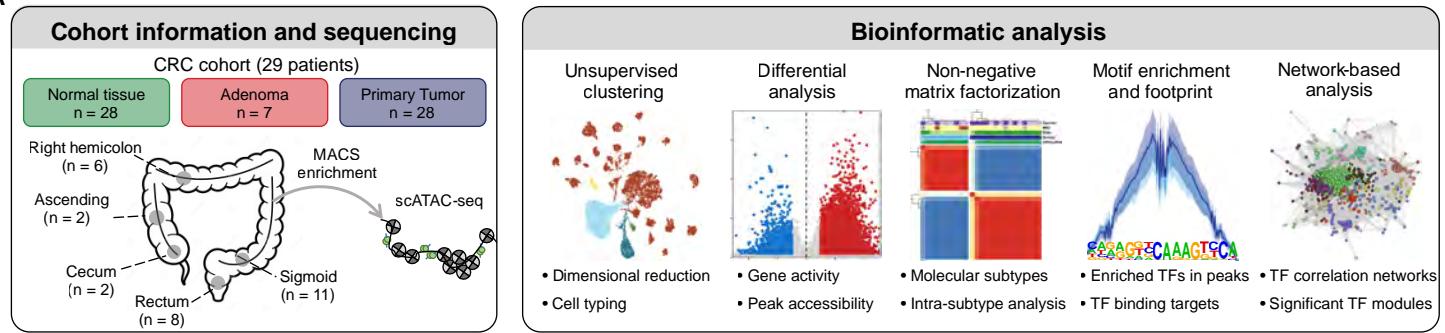
1359 **Figure 5.** Epigenetic regulation of tumor evolution. **A**, The number of CNV-defined subclones in  
1360 each patient. **B**, Hierarchically clustered heatmap of estimated copy number scores in malignant  
1361 cells of patient P26. **C**, UMAP visualization of malignant cells from P26 in the embedding of all  
1362 epithelial cells. **D**, Phylogenetic tree inferred based on the CNV profile of 5 subclones. Subclones  
1363 of the same cluster were labeled with boxes. **E**, Violin plot of iCMS2 module score across  
1364 different subclones of 2 malignant clusters. P value was calculated by Wilcoxon test between 2  
1365 clusters. **F**, Column-scaled subclone chromatin accessibility heatmaps of gained peaks in 5  
1366 subclones. **G**, Dot plot indicating the enrichment of specific TF motifs in the gained peaks of each  
1367 subclone. Note that subclones with no significantly gained peaks (subclone\_2) were not shown. **H**,  
1368 Violin plot showing TF activity score of iCMS2-related TFs across 5 subclones (Pearson  
1369 correlation; ns, not significant; \*, p < 0.05, \*\*; p < 0.01; \*\*\*, p < 0.001; \*\*\*\*, p < 0.001).

1370 **Figure 6.** Identification and epigenetic characterization of CIMP subtypes. **A**, Hierarchically  
1371 clustered heatmaps of chromatin accessibility on marker CGIs across all malignant clusters. CGIs  
1372 with lower and higher chromatin accessibility were colored in blue and red, respectively. **B**,  
1373 Comparison of chromatin accessibility on marker CGIs between each CIMP cluster and healthy  
1374 colon epithelial cells (paired Wilcoxon test). **C**, UMAP visualization of all epithelial cells, colored  
1375 by CIMP subtypes. **D**, Breakdown of cells in each CIMP group by iCMS classification. **E**,  
1376 Column-scaled cluster heatmap of gained peaks in all CIMP subtypes. **F**, Dot plot representing the  
1377 enrichment of CIMP-High-specific TF motifs in gained peaks of all 3 CIMP subtypes. **G**, UMAP  
1378 visualization of gene activity score (top) and TF activity score (bottom) for hyper-activated TFs in  
1379 CIMP-High-specific peaks. **H**, Tn5 bias-subtracted TF footprinting for hyper-activated TFs in  
1380 CIMP-High-specific peaks by scATAC-seq CIMP subtypes.

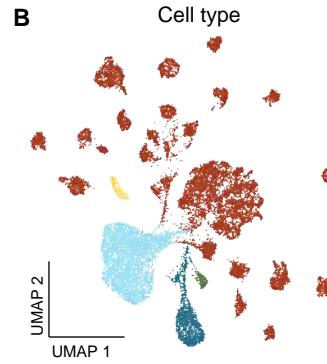
1381 **Figure 7.** Network analysis identified co-activated TF modules related to multi-layer CRC  
1382 heterogeneities. **A**, Network correlation heatmap of all TFs. Branches in the dendrogram  
1383 correspond to modules. Modules correspond to blocks of highly interconnected TFs. TFs with  
1384 high correlation of activities were colored in red. Unclassified TFs were labeled in gray. **B**,  
1385 Heatmap representing the correlation between clinical traits, subtype classification, and module  
1386 eigengene (MEs) (Pearson correlation; \*, p < 0.05, \*\*; p < 0.01; \*\*\*, p < 0.001). **C**, Comparison  
1387 of trait-related module eigengene scores for different MSI status (left, Wilcoxon test), iCMS group  
1388 (middle, Wilcoxon test), and CIMP group (right, One-Way ANOVA). **D,E**, Intra-module structure  
1389 and module-trait relationship between module 9 and MSI status (**D**), and module 8 and iCMS  
1390 group (**E**). Large graph layout (LGL) network visualization of the correlation network of each  
1391 module (left). Dots represented TFs and edges represented correlation of TF activities. The scatter  
1392 plot showed the distribution of TF module membership and TF significance for the trait (light)  
1393 (Pearson correlation). For both plots, the size of the dot reflected the gene activity score of the TF  
1394 in malignant cells. The color of the dot represented the family of the TF. **F**, Heatmap representing  
1395 the correlation between TF modules and iCMS gene signatures (Pearson correlation; \*, p < 0.05,  
1396 \*\*; p < 0.01; \*\*\*, p < 0.001). The iCMS2/3 signatures were defined as the module score of  
1397 upregulated genes subtracted by the module score of downregulated genes. **G**, Scatter showing the  
1398 correlation between selected module eigengenes and iCMS gene signatures (Pearson correlation).

Figure 1

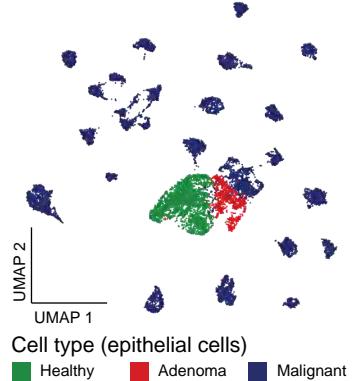
A



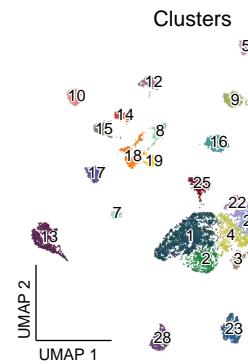
All cells



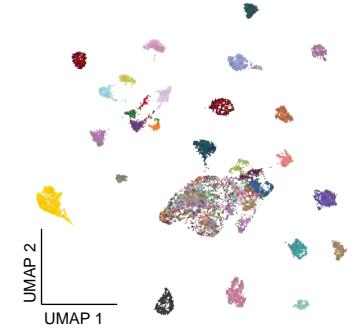
Cell type



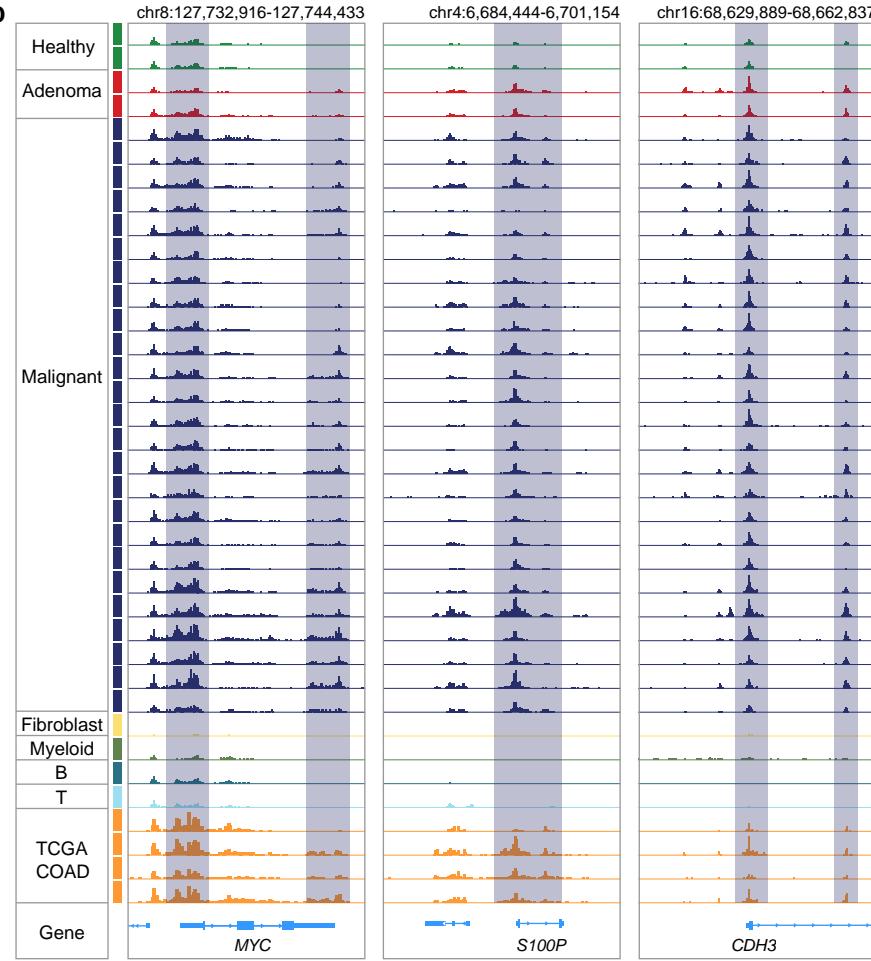
Epithelial cells



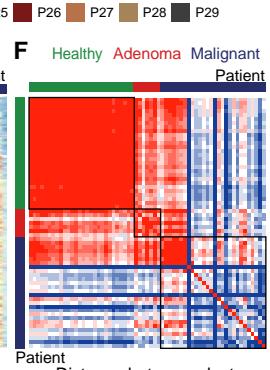
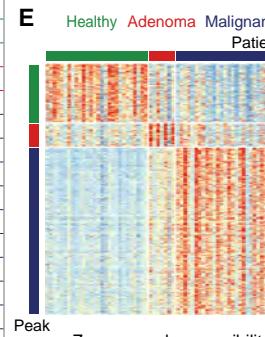
Patient



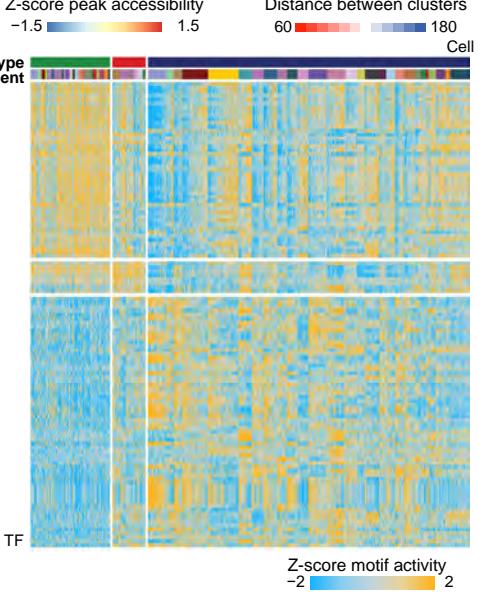
D



E



F



G

**Figure 2**

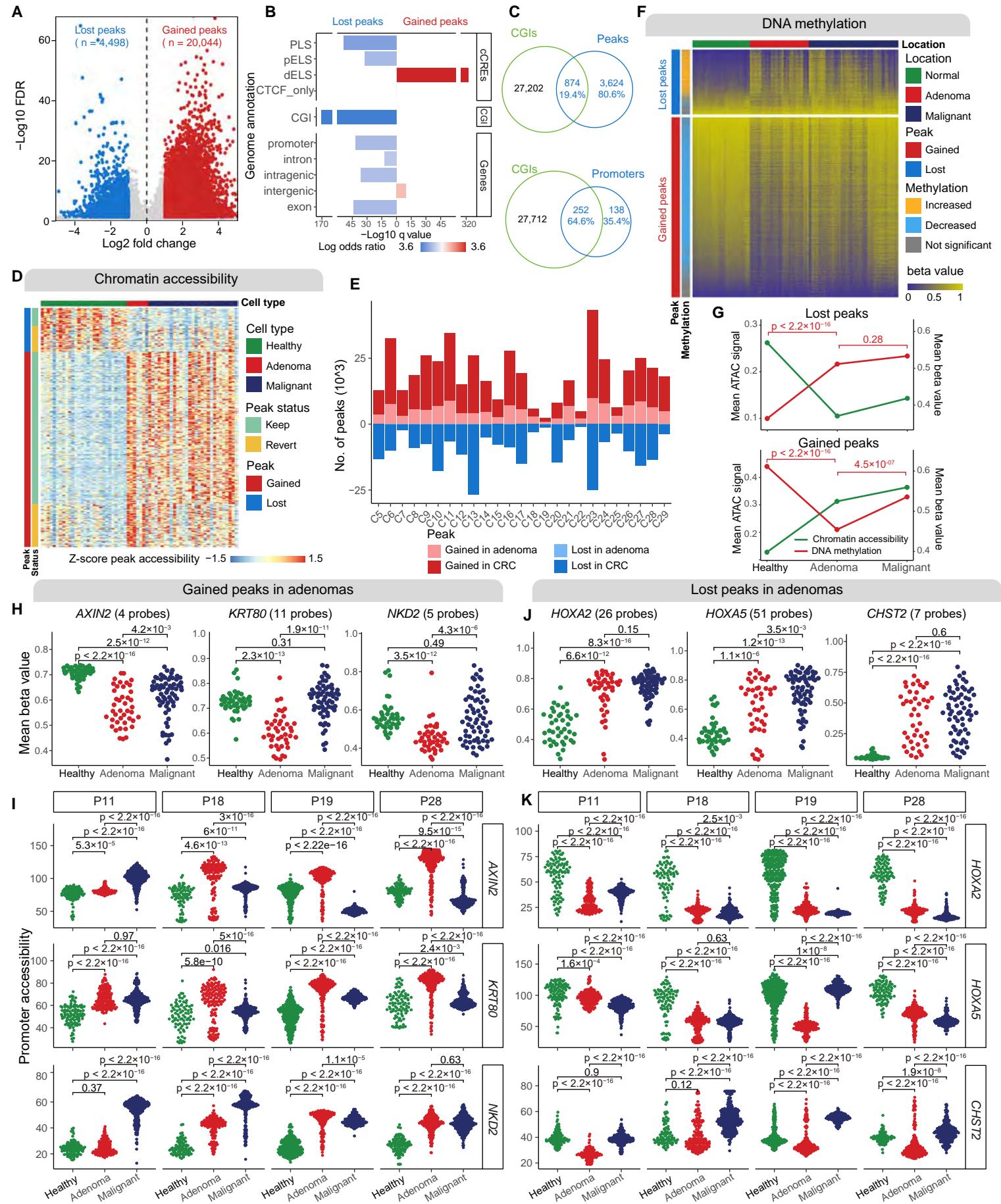


Figure 3

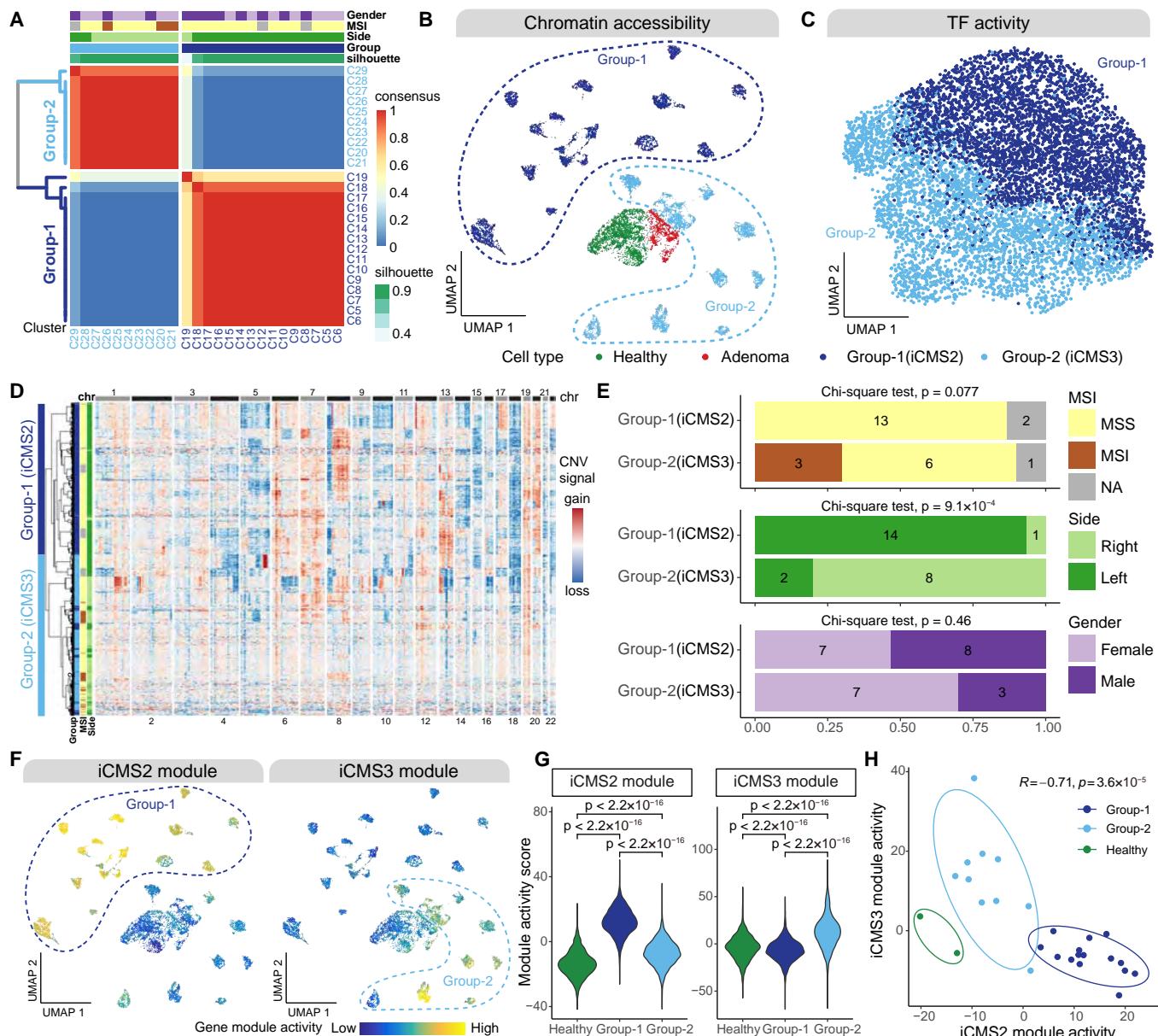


Figure 4

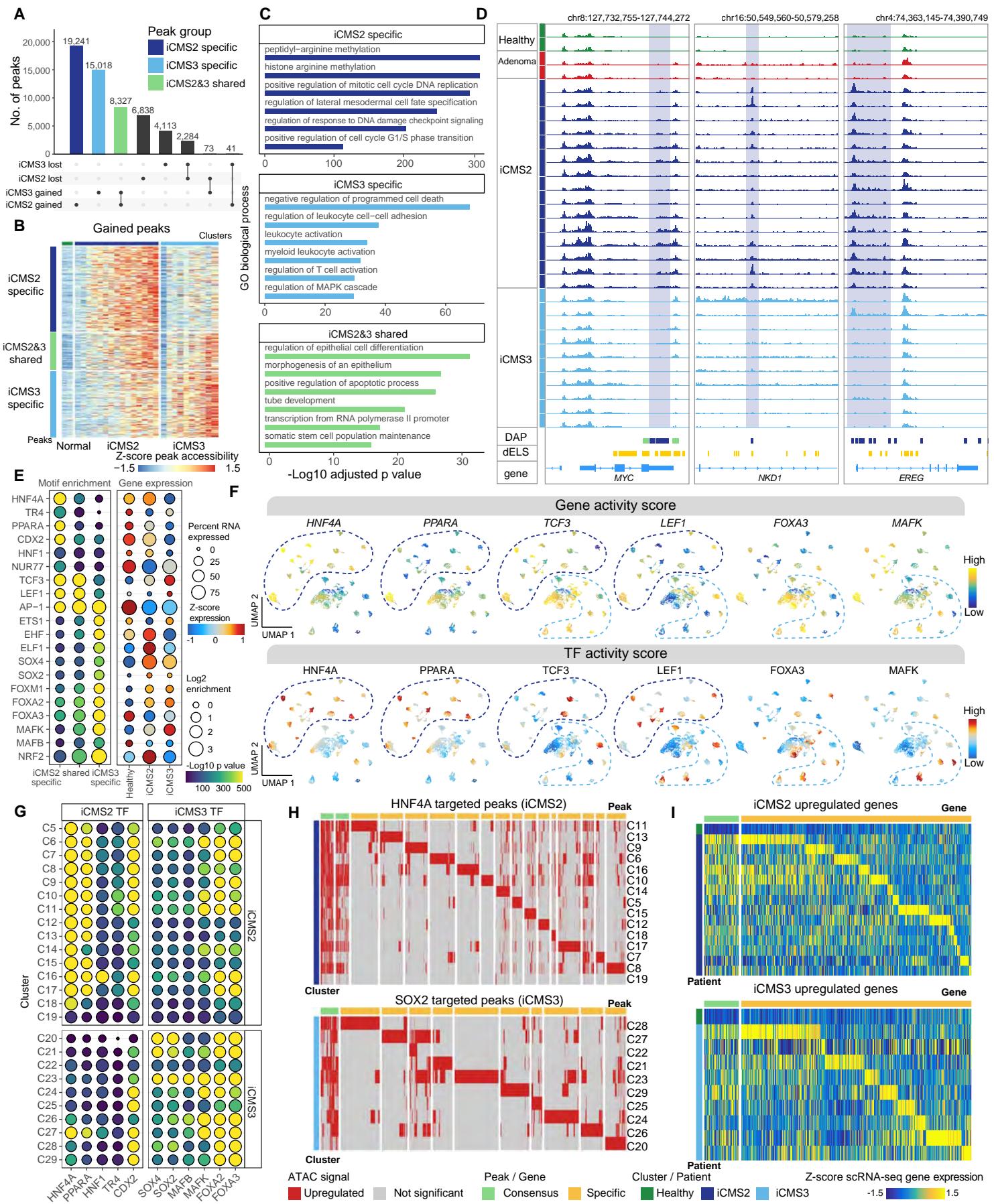


Figure 5

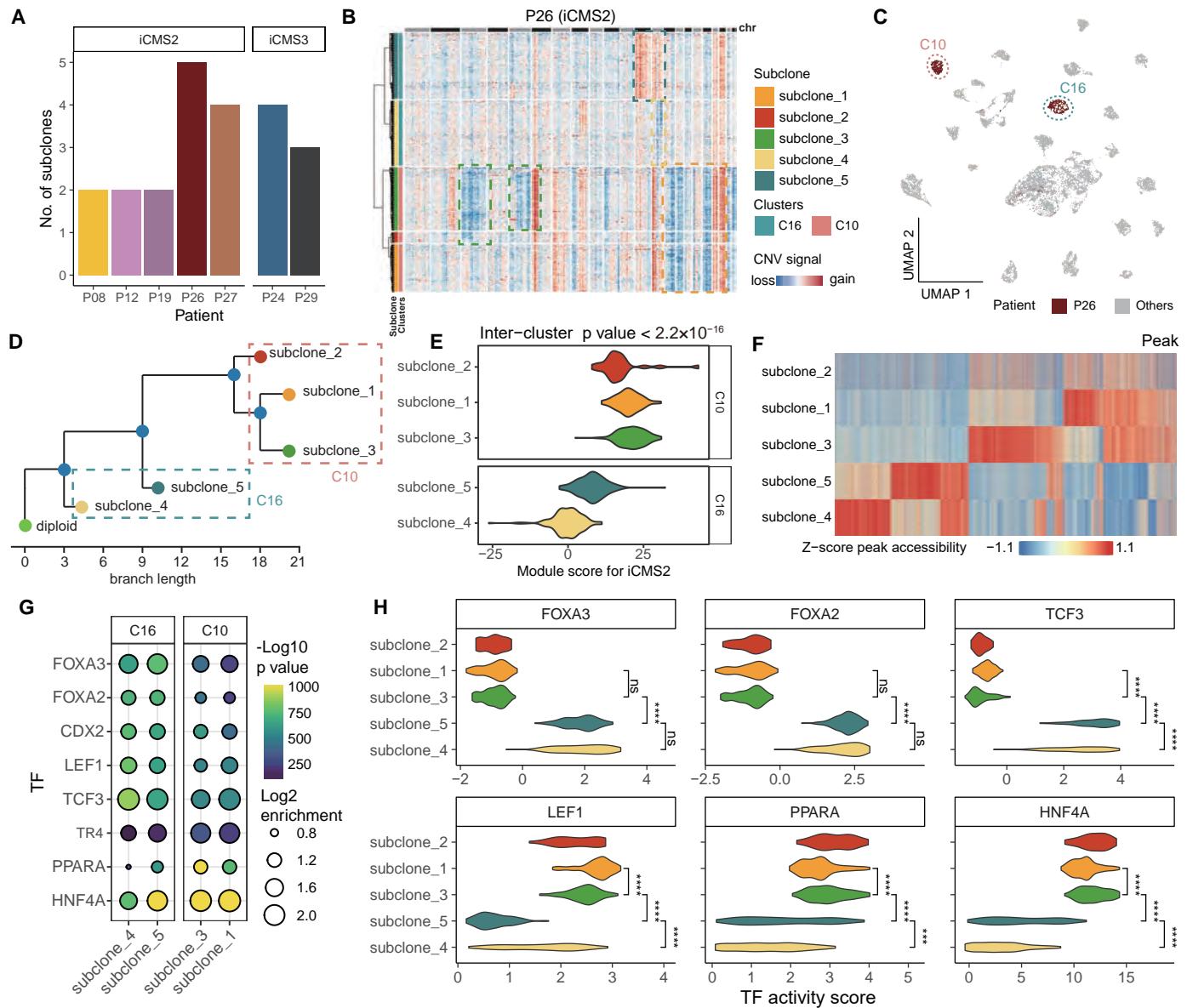


Figure 6

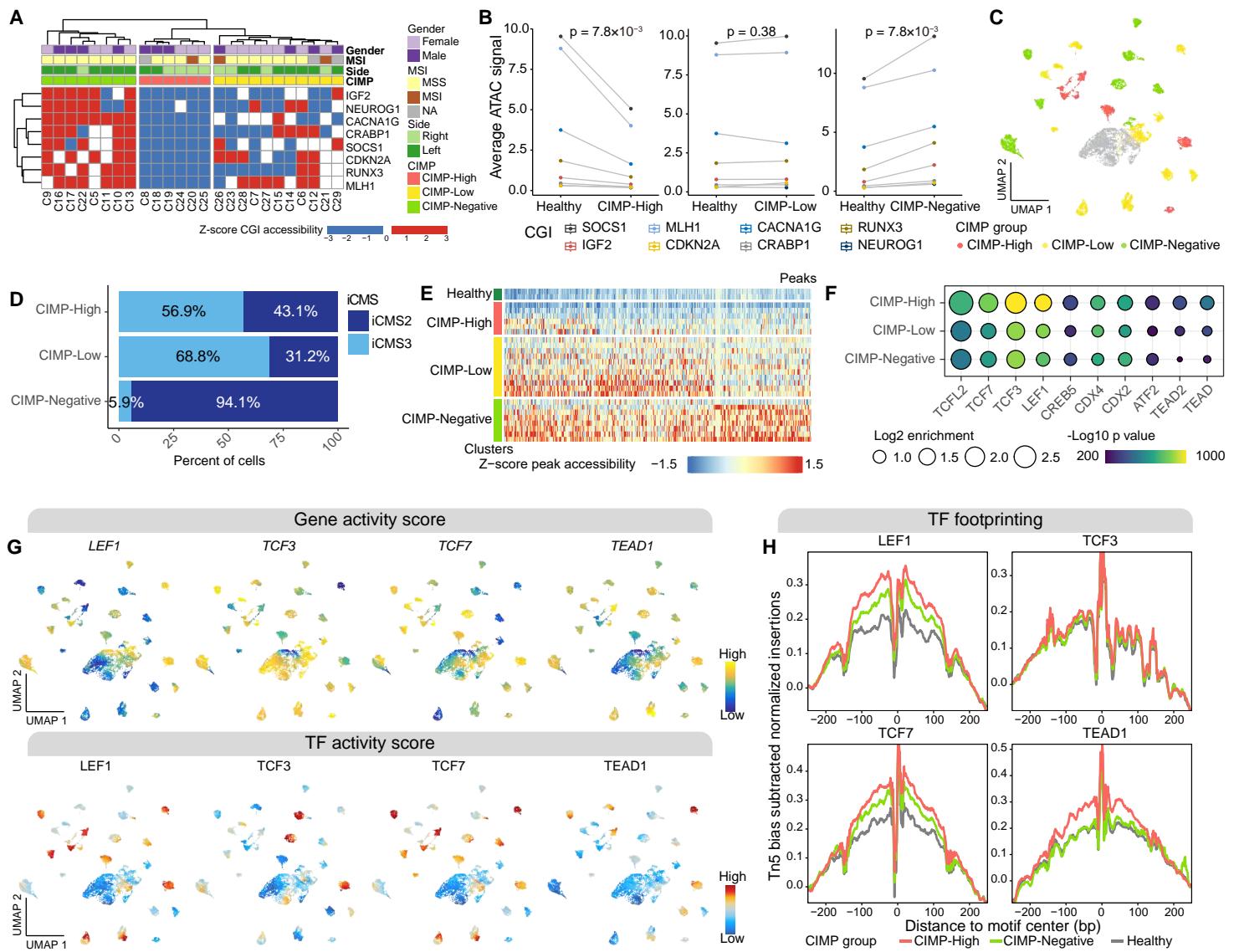
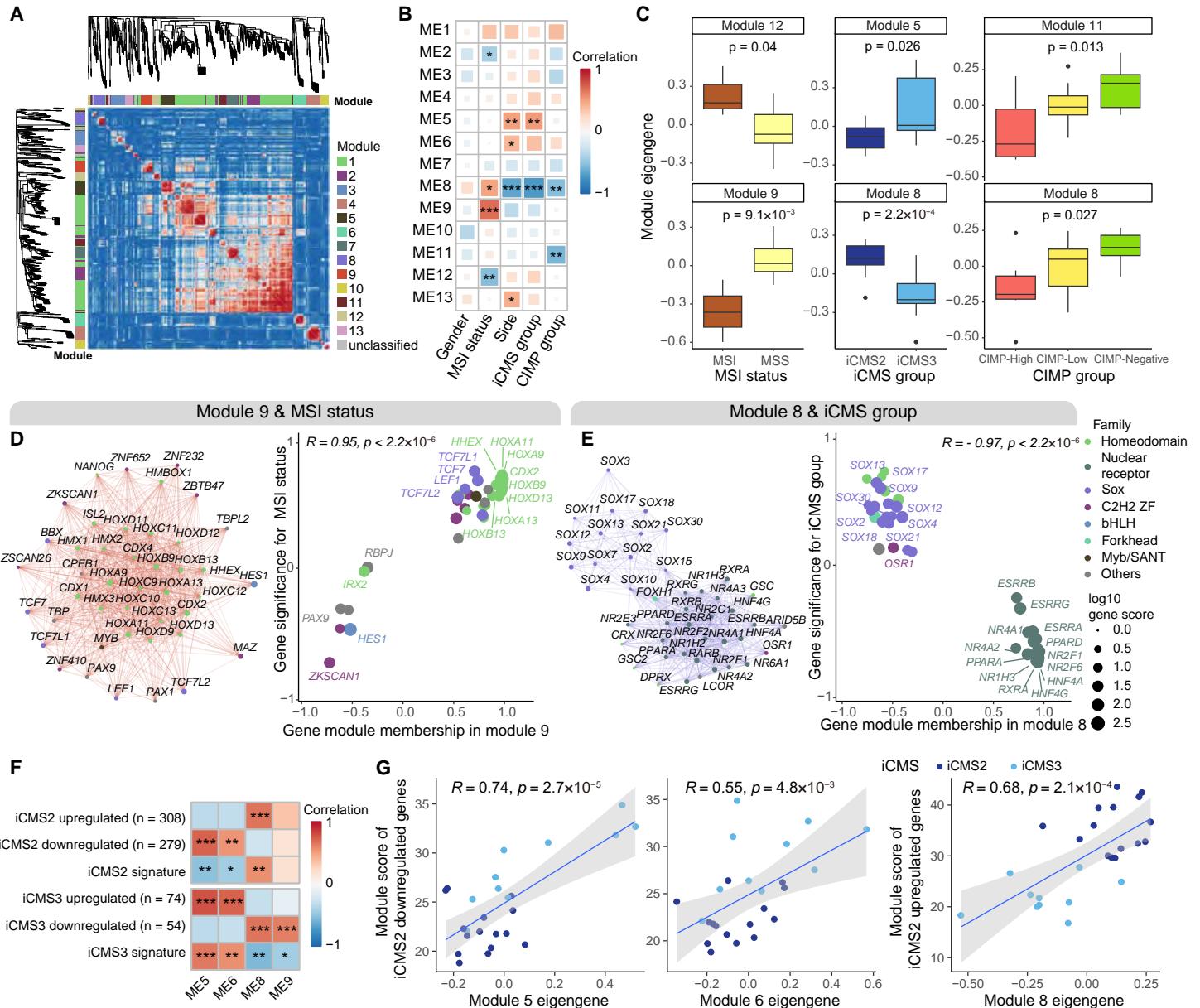


Figure 7



## Parsed Citations

**1. Guinney J, Dienstmann R, Wang X, de Reynies 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: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**2. Cancer Genome Atlas N. Comprehensive molecular characterization of human colon and rectal cancer.** *Nature* 2012;487(7407):330-7 doi 10.1038/nature11252.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**3. Issa JP. CpG island methylator phenotype in cancer.** *Nat Rev Cancer* 2004;4(12):988-93 doi 10.1038/nrc1507.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**4. Toyota M, Ahuja N, Ohe-Toyota M, Herman JG, Baylin SB, Issa JP. CpG island methylator phenotype in colorectal cancer.** *Proc Natl Acad Sci U S A* 1999;96(15):8681-6 doi 10.1073/pnas.96.15.8681.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**5. Marusyk A, Janiszewska M, Polyak K. Intratumor Heterogeneity: The Rosetta Stone of Therapy Resistance.** *Cancer Cell* 2020;37(4):471-84 doi 10.1016/j.ccr.2020.03.007.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**6. Joanito I, Wirapati P, Zhao N, Nawaz Z, Yeo G, Lee F, et al. Single-cell and bulk transcriptome sequencing identifies two epithelial tumor cell states and refines the consensus molecular classification of colorectal cancer.** *Nat Genet* 2022;54(7):963-75 doi 10.1038/s41588-022-01100-4.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**7. Flavahan WA, Gaskell E, Bernstein BE. Epigenetic plasticity and the hallmarks of cancer.** *Science* 2017;357(6348) doi 10.1126/science.aal2380.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**8. Mazor T, Pankov A, Song JS, Costello JF. Intratumoral Heterogeneity of the Epigenome.** *Cancer Cell* 2016;29(4):440-51 doi 10.1016/j.ccr.2016.03.009.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**9. Ogino S, Noshio K, Kirkner GJ, Kawasaki T, Meyerhardt JA, Loda M, et al. CpG island methylator phenotype, microsatellite instability, BRAF mutation and clinical outcome in colon cancer.** *Gut* 2009;58(1):90-6 doi 10.1136/gut.2008.155473.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**10. Weisenberger DJ, Siegmund KD, Campan M, Young J, Long TI, Faasse MA, et al. CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer.** *Nat Genet* 2006;38(7):787-93 doi 10.1038/ng1834.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**11. Samowitz WS, Albertsen H, Herrick J, Levin TR, Sweeney C, Murtaugh MA, et al. Evaluation of a large, population-based sample supports a CpG island methylator phenotype in colon cancer.** *Gastroenterology* 2005;129(3):837-45 doi 10.1053/j.gastro.2005.06.020.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**12. Becker WR, Nevins SA, Chen DC, Chiu R, Horning AM, Guha TK, et al. Single-cell analyses define a continuum of cell state and composition changes in the malignant transformation of polyps to colorectal cancer. Nat Genet 2022;54(7):985-95 doi 10.1038/s41588-022-01088-x.**

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**13. Satpathy AT, Granja JM, Yost KE, Qi Y, Meschi F, McDermott GP, et al. Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. Nat Biotechnol 2019;37(8):925-36 doi 10.1038/s41587-019-0206-z.**

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**14. Armaghany T, Wilson JD, Chu Q, Mills G. Genetic alterations in colorectal cancer. Gastrointest Cancer Res 2012;5(1):19-27.**

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**15. Hermsen M, Postma C, Baak J, Weiss M, Rapallo A, Sciuotto A, et al. Colorectal adenoma to carcinoma progression follows multiple pathways of chromosomal instability. Gastroenterology 2002;123(4):1109-19 doi 10.1053/gast.2002.36051.**

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**16. Ding Q, Chang CJ, Xie X, Xia W, Yang JY, Wang SC, et al. APOBEC3G promotes liver metastasis in an orthotopic mouse model of colorectal cancer and predicts human hepatic metastasis. J Clin Invest 2011;121(11):4526-36 doi 10.1172/JCI45008.**

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**17. Milicic A, Harrison LA, Goodlad RA, Hardy RG, Nicholson AM, Presz M, et al. Ectopic expression of P-cadherin correlates with promoter hypomethylation early in colorectal carcinogenesis and enhanced intestinal crypt fission in vivo. Cancer Res 2008;68(19):7760-8 doi 10.1158/0008-5472.CAN-08-0020.**

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**18. Imai K, Hirata S, Irie A, Senju S, Ikuta Y, Yokomine K, et al. Identification of a novel tumor-associated antigen, cadherin 3/P-cadherin, as a possible target for immunotherapy of pancreatic, gastric, and colorectal cancers. Clin Cancer Res 2008;14(20):6487-95 doi 10.1158/1078-0432.CCR-08-1086.**

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**19. Corces MR, Granja JM, Shams S, Louie BH, Seoane JA, Zhou W, et al. The chromatin accessibility landscape of primary human cancers. Science 2018;362(6413) doi 10.1126/science.aav1898.**

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**20. Schep AN, Wu B, Buenrostro JD, Greenleaf WJ. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. Nat Methods 2017;14(10):975-8 doi 10.1038/nmeth.4401.**

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**21. Lee HO, Hong Y, Etlioglu HE, Cho YB, Pomella V, Van den Bosch B, et al. Lineage-dependent gene expression programs influence the immune landscape of colorectal cancer. Nat Genet 2020;52(6):594-603 doi 10.1038/s41588-020-0636-z.**

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**22. Trusolino L, Comoglio PM. Scatter-factor and semaphorin receptors: cell signalling for invasive growth. Nat Rev Cancer 2002;2(4):289-300 doi 10.1038/nrc779.**

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**23. Boccaccio C, Comoglio PM. Invasive growth: a MET-driven genetic programme for cancer and stem cells.** *Nat Rev Cancer* 2006;6(8):637-45 doi 10.1038/nrc1912.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**24. Wang X, Luo X, Chen C, Tang Y, Li L, Mo B, et al. The Ap-2alpha/Elk-1 axis regulates Sirpalpha-dependent tumor phagocytosis by tumor-associated macrophages in colorectal cancer.** *Signal Transduct Target Ther* 2020;5(1):35 doi 10.1038/s41392-020-0124-z.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**25. Ma Y, Yang Y, Wang F, Moyer MP, Wei Q, Zhang P, et al. Long non-coding RNA CCAL regulates colorectal cancer progression by activating Wnt/beta-catenin signalling pathway via suppression of activator protein 2alpha.** *Gut* 2016;65(9):1494-504 doi 10.1136/gutjnl-2014-308392.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**26. Singh AB, Sharma A, Smith JJ, Krishnan M, Chen X, Eschrich S, et al. Claudin-1 up-regulates the repressor ZEB-1 to inhibit E-cadherin expression in colon cancer cells.** *Gastroenterology* 2011;141(6):2140-53 doi 10.1053/j.gastro.2011.08.038.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**27. Li D, Bai Y, Feng Z, Li W, Yang C, Guo Y, et al. Study of Promoter Methylation Patterns of HOXA2, HOXA5, and HOXA6 and Its Clinicopathological Characteristics in Colorectal Cancer.** *Front Oncol* 2019;9:394 doi 10.3389/fonc.2019.00394.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**28. Liu Y, Olanrewaju YO, Zheng Y, Hashimoto H, Blumenthal RM, Zhang X, et al. Structural basis for Klf4 recognition of methylated DNA.** *Nucleic Acids Res* 2014;42(8):4859-67 doi 10.1093/nar/gku134.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**29. Hu S, Wan J, Su Y, Song Q, Zeng Y, Nguyen HN, et al. DNA methylation presents distinct binding sites for human transcription factors.** *Elife* 2013;2:e00726 doi 10.7554/eLife.00726.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**30. Spruijt CG, Gnerlich F, Smits AH, Pfaffeneder T, Jansen PW, Bauer C, et al. Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives.** *Cell* 2013;152(5):1146-59 doi 10.1016/j.cell.2013.02.004.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**31. Luo Y, Wong CJ, Kaz AM, Dzieciatkowski S, Carter KT, Morris SM, et al. Differences in DNA methylation signatures reveal multiple pathways of progression from adenoma to colorectal cancer.** *Gastroenterology* 2014;147(2):418-29 e8 doi 10.1053/j.gastro.2014.04.039.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**32. Zhai J, Chen H, Wong CC, Peng Y, Gou H, Zhang J, et al. ALKBH5 Drives Immune Suppression Via Targeting AXIN2 to Promote Colorectal Cancer and Is a Target for Boosting Immunotherapy.** *Gastroenterology* 2023;165(2):445-62 doi 10.1053/j.gastro.2023.04.032.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**33. Liu W, Dong X, Mai M, Seelan RS, Taniguchi K, Krishnadath KK, et al. Mutations in AXIN2 cause colorectal cancer with defective mismatch repair by activating beta-catenin/TCF signalling.** *Nat Genet* 2000;26(2):146-7 doi 10.1038/79859.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**34. Li C, Liu X, Liu Y, Liu X, Wang R, Liao J, et al. Keratin 80 promotes migration and invasion of colorectal carcinoma by interacting with PRKDC via activating the AKT pathway. Cell death & disease 2018;9(10):1009 doi 10.1038/s41419-018-1030-y.**

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**35. Tian J, Chang J, Gong J, Lou J, Fu M, Li J, et al. Systematic Functional Interrogation of Genes in GWAS Loci Identified ATF1 as a Key Driver in Colorectal Cancer Modulated by a Promoter-Enhancer Interaction. Am J Hum Genet 2019;105(1):29-47 doi 10.1016/j.ajhg.2019.05.004.**

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**36. Baharudin R, Ishak M, Muhamad Yusof A, Saidin S, Syafruddin SE, Wan Mohamad Nazarie WF, et al. Epigenome-Wide DNA Methylation Profiling in Colorectal Cancer and Normal Adjacent Colon Using Infinium Human Methylation 450K. Diagnostics (Basel) 2022;12(1) doi 10.3390/diagnostics12010198.**

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**37. Xie H, Mahoney DW, Foote PH, Burger KN, Doering KA, Taylor WR, et al. Novel Methylated DNA Markers in the Surveillance of Colorectal Cancer Recurrence. Clin Cancer Res 2021;27(1):141-9 doi 10.1158/1078-0432.CCR-20-2589.**

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**38. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, et al. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol 2010;28(5):495-501 doi 10.1038/nbt.1630.**

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**39. Liu J, Xiao Q, Xiao J, Niu C, Li Y, Zhang X, et al. Wnt/beta-catenin signalling: function, biological mechanisms, and therapeutic opportunities. Signal Transduct Target Ther 2022;7(1):3 doi 10.1038/s41392-021-00762-6.**

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**40. Katoh M, Katoh M. WNT signaling pathway and stem cell signaling network. Clin Cancer Res 2007;13(14):4042-5 doi 10.1158/1078-0432.CCR-06-2316.**

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**41. Twaroski K, Mallanna SK, Jing R, DiFurio F, Urick A, Duncan SA. FGF2 mediates hepatic progenitor cell formation during human pluripotent stem cell differentiation by inducing the WNT antagonist NKD1. Genes Dev 2015;29(23):2463-74 doi 10.1101/gad.268961.115.**

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**42. Cheng WL, Feng PH, Lee KY, Chen KY, Sun WL, Van Hiep N, et al. The Role of EREG/EGFR Pathway in Tumor Progression. Int J Mol Sci 2021;22(23) doi 10.3390/ijms222312828.**

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**43. Chang HR, Nam S, Kook MC, Kim KT, Liu XP, Yao H, et al. HNF4 $\alpha$  is a therapeutic target that links AMPK to WNT signalling in early-stage gastric cancer. Gut 2016;65(1):19-+ doi 10.1136/gutjnl-2014-307918.**

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**44. Lv DD, Zhou LY, Tang H. Hepatocyte nuclear factor 4 $\alpha$  and cancer-related cell signaling pathways: a promising insight into cancer treatment. Experimental and Molecular Medicine 2021;53(1):8-18 doi 10.1038/s12276-020-00551-1.**

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**45. He X, Wang H, Wang R, Li Y, Li S, Cao X, et al. HOXC10 promotes esophageal squamous cell carcinoma progression by targeting FOXA3 and indicates poor survival outcome. *Heliyon* 2023;9(10):e21056 doi 10.1016/j.heliyon.2023.e21056.**

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**46. Ogino S, Kawasaki T, Kirkner GJ, Kraft P, Loda M, Fuchs CS. Evaluation of markers for CpG island methylator phenotype (CIMP) in colorectal cancer by a large population-based sample. *J Mol Diagn* 2007;9(3):305-14 doi 10.2353/jmoldx.2007.060170.**

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**47. Pott S. Simultaneous measurement of chromatin accessibility, DNA methylation, and nucleosome phasing in single cells. *eLife* 2017;6 doi 10.7554/eLife.23203.**

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**48. Clark SJ, Argelaguet R, Kapourani CA, Stubbs TM, Lee HJ, Alda-Catalinas C, et al. scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. *Nat Commun* 2018;9(1):781 doi 10.1038/s41467-018-03149-4.**

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**49. Detilleux D, Spill YG, Balaramane D, Weber M, Bardet AF. Pan-cancer predictions of transcription factors mediating aberrant DNA methylation. *Epigenetics Chromatin* 2022;15(1):10 doi 10.1186/s13072-022-00443-w.**

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**50. Kuo YH, Landrette SF, Heilman SA, Perrat PN, Garrett L, Liu PP, et al. Cbf beta-SMMHC induces distinct abnormal myeloid progenitors able to develop acute myeloid leukemia. *Cancer Cell* 2006;9(1):57-68 doi 10.1016/j.ccr.2005.12.014.**

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**51. Krivtsov AV, Twomey D, Feng Z, Stubbs MC, Wang Y, Faber J, et al. Transformation from committed progenitor to leukaemia stem cell initiated by MLL-AF9. *Nature* 2006;442(7104):818-22 doi 10.1038/nature04980.**

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**52. Barabe F, Kennedy JA, Hope KJ, Dick JE. Modeling the initiation and progression of human acute leukemia in mice. *Science* 2007;316(5824):600-4 doi 10.1126/science.1139851.**

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**53. Ptasińska A, Assi SA, Mannari D, James SR, Williamson D, Dunne J, et al. Depletion of RUNX1/ETO in t(8;21) AML cells leads to genome-wide changes in chromatin structure and transcription factor binding. *Leukemia* 2012;26(8):1829-41 doi 10.1038/leu.2012.49.**

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**54. Bell RJ, Rube HT, Kreig A, Mancini A, Fouse SD, Nagarajan RP, et al. Cancer. The transcription factor GABP selectively binds and activates the mutant TERT promoter in cancer. *Science* 2015;348(6238):1036-9 doi 10.1126/science.aab0015.**

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**55. Little GH, Baniwal SK, Adisetiyo H, Groshen S, Chimge NO, Kim SY, et al. Differential effects of RUNX2 on the androgen receptor in prostate cancer: synergistic stimulation of a gene set exemplified by SNAI2 and subsequent invasiveness. *Cancer Res* 2014;74(10):2857-68 doi 10.1158/0008-5472.CAN-13-2003.**

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**56. Pratap J, Javed A, Languino LR, van Wijnen AJ, Stein JL, Stein GS, et al.** The Runx2 osteogenic transcription factor regulates matrix metalloproteinase 9 in bone metastatic cancer cells and controls cell invasion. *Mol Cell Biol* 2005;25(19):8581-91 doi 10.1128/MCB.25.19.8581-8591.2005.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**57. Boregowda RK, Medina DJ, Markert E, Bryan MA, Chen W, Chen S, et al.** The transcription factor RUNX2 regulates receptor tyrosine kinase expression in melanoma. *Oncotarget* 2016;7(20):29689-707 doi 10.18632/oncotarget.8822.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**58. Cauchy P, James SR, Zacarias-Cabeza J, Ptasinska A, Imperato MR, Assi SA, et al.** Chronic FLT3-ITD Signaling in Acute Myeloid Leukemia Is Connected to a Specific Chromatin Signature. *Cell Rep* 2015;12(5):821-36 doi 10.1016/j.celrep.2015.06.069.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**59. Matkar S, Sharma P, Gao S, Gurung B, Katona BW, Liao J, et al.** An Epigenetic Pathway Regulates Sensitivity of Breast Cancer Cells to HER2 Inhibition via FOXO/c-Myc Axis. *Cancer Cell* 2015;28(4):472-85 doi 10.1016/j.ccr.2015.09.005.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**60. Elias S, Yamin R, Golomb L, Tsukerman P, Stanietsky-Kaynan N, Ben-Yehuda D, et al.** Immune evasion by oncogenic proteins of acute myeloid leukemia. *Blood* 2014;123(10):1535-43 doi 10.1182/blood-2013-09-526590.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**61. Cerezo M, Guemiri R, Druilennec S, Girault I, Malka-Mahieu H, Shen S, et al.** Translational control of tumor immune escape via the eIF4F-STAT1-PD-L1 axis in melanoma. *Nat Med* 2018;24(12):1877-86 doi 10.1038/s41591-018-0217-1.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**62. Xu Y, Poggio M, Jin HY, Shi Z, Forester CM, Wang Y, et al.** Translation control of the immune checkpoint in cancer and its therapeutic targeting. *Nat Med* 2019;25(2):301-11 doi 10.1038/s41591-018-0321-2.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**63. Rakhra K, Bachireddy P, Zabuawala T, Zeiser R, Xu L, Kopelman A, et al.** CD4(+) T cells contribute to the remodeling of the microenvironment required for sustained tumor regression upon oncogene inactivation. *Cancer Cell* 2010;18(5):485-98 doi 10.1016/j.ccr.2010.10.002.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**64. Francis R, Guo H, Streutker C, Ahmed M, Yung T, Dirks PB, et al.** Gastrointestinal transcription factors drive lineage-specific developmental programs in organ specification and cancer. *Sci Adv* 2019;5(12):eaax8898 doi 10.1126/sciadv.aax8898.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**65. Konukiewitz B, Schmitt M, Silva M, Pohl J, Lang C, Steiger K, et al.** Loss of CDX2 in colorectal cancer is associated with histopathologic subtypes and microsatellite instability but is prognostically inferior to hematoxylin-eosin-based morphologic parameters from the WHO classification. *Br J Cancer* 2021;125(12):1632-46 doi 10.1038/s41416-021-01553-0.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**66. Zheng K, Wan H, Zhang J, Shan G, Chai N, Li D, et al.** A novel NGS-based microsatellite instability (MSI) status classifier with 9 loci for colorectal cancer patients. *J Transl Med* 2020;18(1):215 doi 10.1186/s12967-020-02373-1.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**67. Lin CH, Lin JK, Chang SC, Chang YH, Chang HM, Liu JH, et al.** Molecular profile and copy number analysis of sporadic colorectal cancer in Taiwan. *J Biomed Sci* 2011;18(1):36 doi 10.1186/1423-0127-18-36.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**68. Li Y, Zhao Q, Fan LQ, Wang LL, Tan BB, Leng YL, et al.** Zinc finger protein 139 expression in gastric cancer and its clinical significance. *World J Gastroenterol* 2014;20(48):18346-53 doi 10.3748/wjg.v20.i48.18346.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**69. Yao Z, Luo J, Hu K, Lin J, Huang H, Wang Q, et al.** ZKSCAN1 gene and its related circular RNA (circZKSCAN1) both inhibit hepatocellular carcinoma cell growth, migration, and invasion but through different signaling pathways. *Mol Oncol* 2017;11(4):422-37 doi 10.1002/1878-0261.12045.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**70. Boyd M, Bressendorff S, Moller J, Olsen J, Troelsen JT.** Mapping of HNF4alpha target genes in intestinal epithelial cells. *BMC Gastroenterol* 2009;9:68 doi 10.1186/1471-230X-9-68.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**71. Heinonen H, Lepikhova T, Sahu B, Pehkonen H, Pihlajamaa P, Louhimo R, et al.** Identification of several potential chromatin binding sites of HOXB7 and its downstream target genes in breast cancer. *Int J Cancer* 2015;137(10):2374-83 doi 10.1002/ijc.29616.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**72. Schwitalla S, Fingerle AA, Cammareri P, Nebelsiek T, Göktuna SI, Ziegler PK, et al.** Intestinal Tumorigenesis Initiated by Dedifferentiation and Acquisition of Stem-Cell-like Properties. *Cell* 2013;152(1-2):25-38 doi 10.1016/j.cell.2012.12.012.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**73. Zorn AM, Wells JM.** Vertebrate endoderm development and organ formation. *Annu Rev Cell Dev Biol* 2009;25:221-51 doi 10.1146/annurev.cellbio.042308.113344.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**74. Gao N, White P, Kaestner KH.** Establishment of intestinal identity and epithelial-mesenchymal signaling by Cdx2. *Dev Cell* 2009;16(4):588-99 doi 10.1016/j.devcel.2009.02.010.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**75. Bushweller JH.** Targeting transcription factors in cancer - from undruggable to reality. *Nat Rev Cancer* 2019;19(11):611-24 doi 10.1038/s41568-019-0196-7.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**76. Martin M.** Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal* 2011;17(1):3 doi 10.14806/ej.17.1.200.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**77. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL.** Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* 2019;37(8):907-15 doi 10.1038/s41587-019-0201-4.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**78. Granja JM, Corces MR, Pierce SE, Bagdatli ST, Choudhry H, Chang HY, et al.** ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat Genet* 2021;53(3):403-11 doi 10.1038/s41588-021-00790-6.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**79. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al.** Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 2008;9(9):R137 doi 10.1186/gb-2008-9-9-r137.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**80. 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-W60 doi 10.1093/nar/gkz430.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**81. 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: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**82. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, 3rd, Zheng S, Butler A, et al.** Integrated analysis of multimodal single-cell data. *Cell* 2021;184(13):3573-87 e29 doi 10.1016/j.cell.2021.04.048.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**83. Sheffield NC, Bock C.** LOLA: enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor. *Bioinformatics* 2016;32(4):587-9 doi 10.1093/bioinformatics/btv612.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**84. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al.** Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* 2014;30(10):1363-9 doi 10.1093/bioinformatics/btu049.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**85. Gaujoux R, Seoighe C.** A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics* 2010;11:367 doi 10.1186/1471-2105-11-367.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**86. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al.** Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 2010;38(4):576-89 doi 10.1016/j.molcel.2010.05.004.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**87. Kaufmann TL, Petkovic M, Watkins TBK, Colliver EC, Laskina S, Thapa N, et al.** MEDICC2: whole-genome doubling aware copy-number phylogenies for cancer evolution. *Genome Biol* 2022;23(1):241 doi 10.1186/s13059-022-02794-9.

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)

**88. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 2008;9:559 doi 10.1186/1471-2105-9-559.**

Pubmed: [Author and Title](#)

Google Scholar: [Google Scholar Search](#)