



EDITORIAL

Epigenetic regulators as the foundation for molecular classification of colorectal cancer

Zhenyu Liu^{1,2}, Xin Zhou^{1,3}, Fuchou Tang^{1,2,4}

¹School of Life Sciences, Biomedical Pioneering Innovation Center, Department of General Surgery, Third Hospital, Peking University, Beijing 100871, China; ²Beijing Advanced Innovation Center for Genomics (ICG), Ministry of Education Key Laboratory of Cell Proliferation and Differentiation, Beijing 100871, China; ³Peking University Third Hospital Cancer Center, Beijing 100191, China; ⁴Academy for Advanced Interdisciplinary Studies, Peking University, Beijing 100871, China

Colorectal cancer (CRC) accounts for approximately 10% of newly diagnosed cancer cases and cancer-related deaths worldwide¹. The identification of molecular subtypes of CRCs has significantly advanced treatment strategies, including targeted therapy and immunotherapy. CRC forms a highly heterogeneous set of tumors with well-classified subtypes based on the genome, DNA methylome, and transcriptome signatures²⁻⁴, among which genetic factors and transcriptomic profiles have been more extensively studied and are better understood. However, epigenetic regulatory mechanisms also exert profound influence on tumor cell phenotype⁵, while epigenetic subtyping and regulation of gene expression under different CRC subtypes remain poorly understood. This gap in knowledge underscores the need to understand the epigenetic determinants of CRC phenotypic diversity.

The (intrinsic) consensus molecular subtype (CMS) of CRCs

Cancer subtyping based on transcriptomic profiles has been widely accepted as an important source of disease stratification because gene expression is closely associated with cellular phenotypes and tumor behaviors. In 2015 an international

consortium integrated six independently identified transcriptome-based subtyping systems and established the CMS system for CRCs. Most tumors were classified into one of four CMSs (CMS1-4)³, which were characterized by high immune infiltration, canonical WNT and MYC activation, metabolic disorders, and high mesenchymal infiltration, respectively. The fibrotic CMS4 had the worse overall relapse-free survival³.

The CMS groups were identified from bulk gene expression profiles of heterogeneous tissues, which obscured the composition of cell types and microenvironmental interactions within the tumor. Subsequent research using single-cell RNA sequencing (scRNA-seq) provided global cellular landscapes of CRCs. Joanito et al.⁶ performed subtyping analysis exclusively on malignant epithelial cells, which identified two cancer cell states and established the intrinsic CMS (iCMS) classification system. Cancer cells were classified into iCMS2 and iCMS3, in which iCMS2 cells exhibited higher WNT and MYC activities and adenoma gene signatures, while iCMS3 cells exhibited higher MAPK pathway activity and sessile serrated lesion gene signatures.

The (i)CMS systems were initially defined based on transcriptomic profiles, while a deeper correlative analysis with well-defined genetic and microenvironmental features further intensified the understanding of each CMS (**Figure 1**). First, iCMS2 cells exhibited high frequencies of somatic copy number alterations (SCNAs) and mutations in *APC* and *TP53*, whereas iCMS3 cells had few SCNAs and were enriched for mutations in *KRAS*, *PIK3CA*, and *BRAF*. Next, all iCMS2 cells were microsatellite stable (MSS), while iCMS3 cells were further classified into MSS and microsatellite unstable (MSI). The iCMS3-MSI tumors featured elevated immune response signatures and higher immune cell infiltration⁶. Indeed, iCMS2, iCMS3-MSS, and iCMS3-MSI corresponded to CMS2, CMS3

Correspondence to: Xin Zhou and Fuchou Tang
E-mail: zhouxinasd@sina.cn and tangfuchou@pku.edu.cn
ORCID ID: <https://orcid.org/0000-0002-4048-4017> and
<https://orcid.org/0000-0002-8625-7717>

Received May 9, 2024; accepted June 24, 2024;
published online July 16, 2024.

Available at www.cancerbiomed.org

©2024 The Authors. Creative Commons Attribution-NonCommercial 4.0 International License

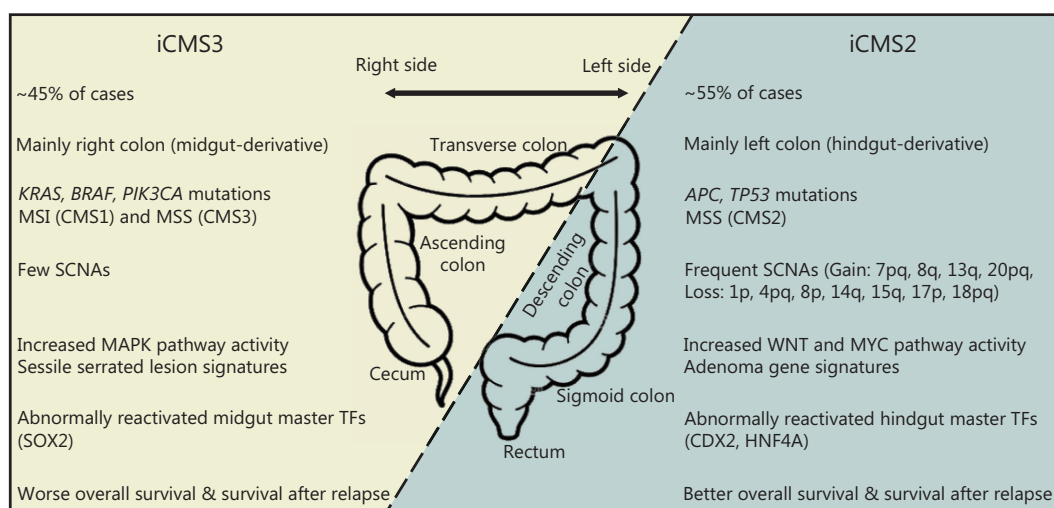


Figure 1 Schematic illustrating the iCMS classification of CRCs. iCMS subtypes have distinct characteristics on driver mutations, genome instability, gene expression signatures, TF regulators, and clinical features. MSS, microsatellite stable; MSI, microsatellite unstable; SCNA, somatic copy number alteration.

and CMS1 CRCs, respectively. The fibrotic CMS4 subtype was not included in the iCMS system, suggesting that microenvironmental mesenchymal features are decoupled from the intrinsic characteristics of tumor cells.

Epigenetic classification of CRCs

Non-mutational epigenetic reprogramming is a hallmark of cancer^{5,7}. Previous research involving CRC subtyping has primarily focused on DNA methylation due to the higher technical feasibility of DNA methylation assays compared with chromatin-related assays. In 1999, Toyota et al.⁴ first reported that a subset of CRCs exhibited abnormally increased DNA methylation at cancer-specific methylated regions (hypermethylation of the promoter regions of eight marker genes), forming a distinct subtype referred to as CpG-island methylator phenotype (CIMP)⁴. Subsequent studies expanded the methylated marker panel of CIMP tumors and further classified all CRCs into CIMP-high, CIMP-low, and CIMP-negative groups based on the degree of genome-wide hypermethylation^{8,9}. The CIMP-High subtype is closely associated with the prognosis of CRC patients and encompasses nearly all cases with *BRAF* mutations^{8,9}.

Unlike DNA methylation, chromatin-related epigenetic regulatory profiles, including histone modification, chromatin accessibility, and transcription factor (TF) binding, have

not been thoroughly studied in CRC. Liu and colleagues¹⁰ recently reported in *Cancer Discovery* a systematic analysis of the CRC epigenetic regulatory landscape using single-cell assay for transposase-accessible chromatin using sequencing (scATAC-seq). Based on the genome-wide chromatin accessibility profile, Liu et al.¹⁰ identified two distinct epigenetic CRC subgroups [group-1 and group-2 (Figure 3A in Liu et al.¹⁰)]. Group-1 cells are characterized by frequent SCNAs, MSS, and predominant origination from the left colon, whereas group-2 cells encompass all MSI cases and primarily originate from the right colon. The genetic alterations, clinical features, and more importantly, gene expression signatures of epigenetically identified group-1 and group-2 CRC subgroups perfectly matched the scRNA-seq-defined iCMS2 and iCMS3 CRCs (Figure 3D-F of Liu et al.¹⁰), respectively. This finding confirmed for the first time that the well-characterized (i)CMS classification system is robust at the epigenetic level, which further expands the scope of iCMS systems to the epigenetic regulation level.

TF activities shape both inter- and intra-subtype heterogeneities

The (i)CMS system provides a multi-dimensional phenotypic classification of CRCs, encompassing genomic, transcriptomic,

microenvironmental, and clinical features. However, the subtype-specific gene regulation mechanisms underlying this phenotypic classification have not been established. Using the scATAC-seq atlas, Liu and colleagues¹⁰ systematically identified epigenetic regulators with subtype-specific activities, including cis-regulatory elements (Figure 4B, D in Liu et al.¹⁰) and transcription factors. The activities of several TFs show significant differences between iCMS subtypes, including HNF4A, PPARA, and CDX2 for iCMS2 tumors, and FOXA3 and MAFK for iCMS3 tumors (Figure 4E, F in Liu et al.¹⁰). Notably, a subsequent patient-level analysis confirmed that the activities of these TFs exhibit strong inter-subtype discrepancies and high intra-subtype consistencies. HNF4A, for example, was hyper-activated in the tumor cells of every iCMS2 subtype patient, but not in the iCMS3 subtype (Figure 4G, H in Liu et al.¹⁰). This finding indicated that hyper-activation of

subtype-specific TFs shape the phenotypic characteristics of iCMS subtypes.

Despite the shared TF activity profiles of tumors within the same subtype, strong intra-subtype heterogeneities have been noted for both gene expression and chromatin accessibility. Among patients of the same subtype, only a small fraction (approximately 10%) of genes and regulatory elements were consistently activated across all patients, while the vast majority of aberrantly activated genes were still individual-patient-specific (Figure 4H, I in Liu et al.¹⁰). Notably, this intra-subtype minor similarity and major diversity for gene expression and chromatin accessibility could be accurately explained by TF binding profiles. TFs, which serve as master regulators of gene expression and cellular phenotypes, shape both inter- and intra-subtype heterogeneities by subtype-specific activation and patient-specific binding and activation of downstream target sites (Figure 2).

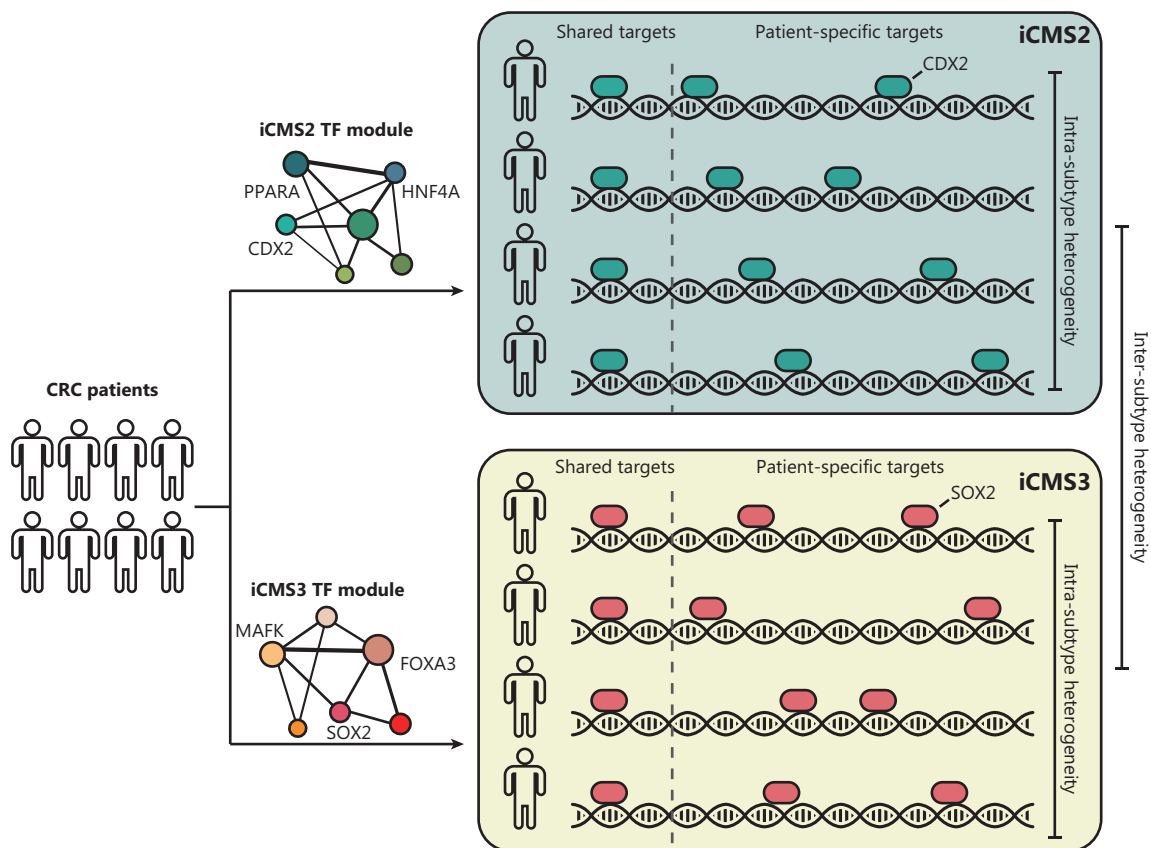


Figure 2 TF activities shape both inter- and intra-subtype heterogeneities. Consistent hyperactivation of iCMS-specific TF modules shapes inter-subtype heterogeneities, including HNF4A and PPARA for iCMS2 tumors, and FOXA3 and MAFK for iCMS3 tumors. Among tumors of the same subtype, the majority of downstream targets of aberrantly activated TFs are patient-specific, whereas only a small fraction of downstream targets is shared by most patients, thus contributing to intra-subtype heterogeneities.

Synergistic TF modules regulate multi-dimensional CRC phenotypes

The molecular subtyping of CRCs encompasses multiple aspects of tumor cell phenotypes, including clinical features, driver mutations, genome instabilities, transcriptome signatures, and DNA methylation profiles¹¹. Different aspects of cancer cell characteristics are modulated and interconnected by complicated gene regulation networks. Using weighted correlation network analysis, Liu and colleagues¹⁰ identified several TF modules related to multi-dimensional CRC heterogeneities. TFs within the same module exhibit high correlations in terms of TF activities (Figure 7A in Liu et al.¹⁰), thus forming distinct collaborative regulatory units to shape cellular phenotypes. TF modules associated with iCMS classifications, including ME5 and ME8, are also significantly associated with the left/right side origins of CRC tumors, confirming shared gene regulation programs between iCMS subtypes and the location of tumor origin (Figure 7B, C in Liu et al.¹⁰). Several TF modules are independently related to microsatellite instability or CIMP subtypes, while iCMS-related ME8 also contribute to these features. These results indicate that the complicated phenotypic subtypes are deconvoluted into distinct regulatory modules, while the synergistic activities of different TFs within the same modules collaboratively shape different cellular phenotypes of tumor cells.

In addition to modulating multi-dimensional CRC phenotypes, regulation of a single CRC subtype also requires the collaborative efforts of different TF modules. Specifically, three distinct TF modules (ME5, ME6, and ME8) are all associated with the gene expression signatures of iCMS2 CRCs. However, ME5 and ME6 are exclusively associated with the downregulated iCMS2 CRC genes compared to healthy colon epithelial cells, while ME8 is only associated with the upregulated iCMS2 genes¹⁰ (Figure 7F, G in Liu et al.¹⁰). Consistent with this observation, several upregulated genes in iCMS2 were confirmed as direct targets of module 8 TFs. Taken together, the synergistic activities of different TF modules collaboratively regulate distinct downstream genes and shape multi-dimensional heterogeneities of CRCs.

Epigenetic regulation of intra-tumor heterogeneities (ITHs)

As a tumor evolves, cancer cells gradually form genetically heterogeneous subclones, which form complicated ITHs.

ITHs can arise from genetic, epigenetic, or microenvironmental inputs and provide fuel for therapy resistance during tumor evolution¹². ITH analysis requires multi-regional sampling and high-resolution methods, which is difficult using traditional methods. Recent progress in single-cell multi-omics sequencing technologies has provided powerful tools for studying the dynamics of multi-dimensional ITH along cancer evolution. A milestone study published in *Science* by Bian and colleagues¹³ used single-cell multi-omics sequencing to analyze primary tumor and lymphatic and distant metastases of CRCs, which simultaneously acquired genome (CNVs), DNA methylome, and transcriptome profiles from each single cell. Different subclones within each tumor, as well as the phylogenetic relationships between subclones, can be accurately identified using SCNA profiles. Further correlative analysis between genetic lineages and DNA methylation profiles revealed that genome-wide DNA methylation levels are tightly associated with genetic lineages, confirming that the ITH of DNA methylation is also shaped by genetic ITH. Associations between DNA methylation and gene expression levels have also been demonstrated, in which the DNA methylation ratios of gene promoters are negatively associated with the levels of gene expression¹³. These results demonstrated the feasibility of single-cell multi-omics sequencing in reconstructing genetic lineages and tracing epigenomic and transcriptomic dynamics during tumor evolution.

The milestone work of Bian et al.¹³ provided comprehensive insights into the genomic, transcriptomic, and DNA methylome evolution of CRCs. However, little is known about the dynamics of chromatin-related epigenetic regulators, including histone modification, chromatin accessibility, and TF binding. Liu and colleagues¹⁰ adopted a similar strategy in the scATAC-seq atlas, in which genetic subclones and corresponding phylogenetic relationships could be inferred from SCNA profiles, thus making it possible to investigate the dynamics of chromatin accessibility and TF binding during cancer evolution. Profound ITH on chromatin accessibility profiles has been demonstrated. Subsequently emerged subclones were shown to have higher iCMS-specific gene expression signatures (Figure 5E in Liu et al.¹⁰). The activities of iCMS-specific TFs were gradually established along the phylogenetic evolutionary trajectory (Figure 5G, H in Liu et al.¹⁰). This stepwise activation of iCMS-specific TFs aligned with the expression patterns of iCMS signature genes, providing insights into the epigenetic programs regulating the formation of iCMS phenotypes during tumor evolution.

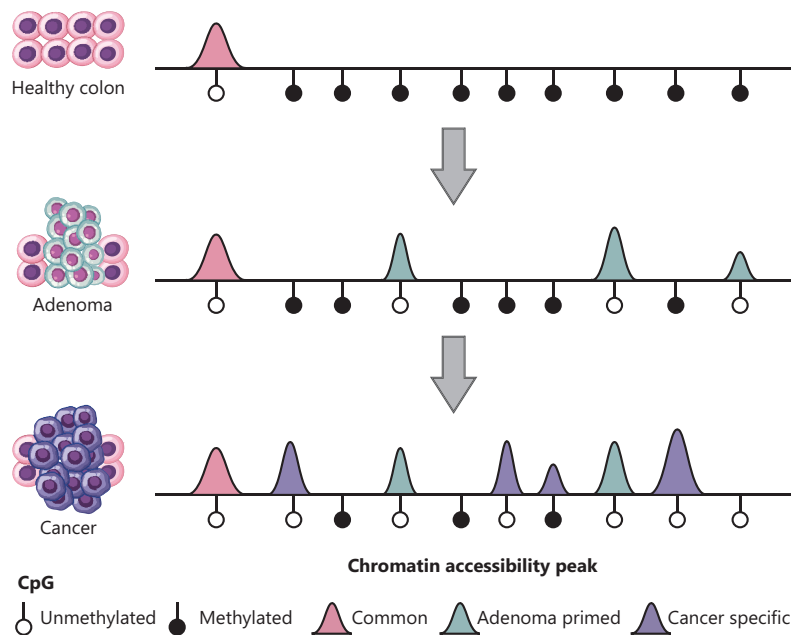


Figure 3 Epigenetic regulation during generation of early adenomas. Gained peaks during the generation of precancerous benign adenomas are mostly maintained in malignant tumor cells. Dynamics of chromatin accessibility is tightly accompanied by opposite changes in DNA methylation.

Early epigenetic dynamics before malignant transformation

Most CRCs are derived from adenomatous polyps. Therefore, illustrating the regulatory mechanisms during the generation of adenomas could provide targets for early screening and diagnostics of CRCs. Liu and colleagues¹⁰ systematically identified the chromatin state signatures for precancerous polyps and observed a strong anti-correlation between chromatin accessibility and DNA methylation levels (Figure 2D, F in Liu et al.¹⁰), which collaboratively regulate the expression of several CRC drivers (**Figure 3**). Notably, abnormal chromatin states acquired in adenomas are predominantly maintained during subsequent malignant transformation, both for chromatin accessibility and DNA methylation (**Figure 3**). This epigenetic co-regulation of chromatin accessibility and DNA methylation provides promising candidates for CRC screening (Figure 2H-K in Liu et al.¹⁰). For example, *HOXA* genes exhibit extremely high frequencies of hypermethylation in CRCs and have excellent diagnostic value¹⁴. However, the chronologic order and causal relationships between the dynamic of chromatin accessibility and DNA methylation are unclear. In fact, it is more likely that an epithelial cell first loses

DNA methylation due to metabolic abnormalities¹⁵. Then, loss of DNA methylation on many regulatory elements, such as promoters and enhancers, leads to increases in chromatin accessibilities of these elements, which leads to aberrant activation of target gene transcription. Delicate efforts are still needed to further illustrate how these different epigenetic factors collaboratively regulate gene expression and contribute to disease initiation and progression.

Perspective

The (i)CMS classification system, initially defined from gene expression profiles, provides comprehensive molecular stratification for CRCs. Different CRC subtypes have been characterized based on distinct clinical, genomic, transcriptomic, and microenvironmental profiles, which have significantly revolutionized the paradigms of precision medicine in CRC¹¹. Liu et al.¹⁰ first extended the scope of (i)CMS from phenotypic traits to gene regulation programs, linking the well-established subtypes to epigenetic underpinnings, which provided novel insights and potential targets for CRC therapy.

TFs are master regulators of cellular phenotypes. Dysregulated TF activities have been frequently observed in

various cancer types and are associated with multiple cancer hallmark properties. Thus, dysregulated TFs could serve as a unique class of therapeutic targets. Several strategies have been established to design TF-targeting drugs, including disrupting DNA binding activities, inhibiting interactions with coactivators, or modulating proteasomal degradations¹⁶. Several TFs are widely activated in different patients, suggesting that suppressing these TFs may lead to relatively uniform antitumor responses in a wide variety of patients. Abnormally activated TFs in CRCs, both subtype-specific or shared, might serve as novel targets for CRC treatment.

Grant support

This work was supported by grants from the National Natural Science Foundation of China (Grant No. 81972702) and the Beijing Nova Program (Grant No. 2022029).

Conflict of interest statement

No potential conflicts of interest are disclosed.

Author contributions

Conceived and designed the analysis: Xin Zhou, Fuchou Tang.

Performed the analysis: Zhenyu Liu.

Wrote the paper: Zhenyu Liu.

References

- Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2024; 74: 229-63.
- The Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature.* 2012; 487: 330-7.
- Guinney J, Dienstmann R, Wang X, de Reyniès A, Schlicker A, Soneson C, et al. The consensus molecular subtypes of colorectal cancer. *Nat Med.* 2015; 21: 1350-6.
- Toyota M, Ahuja N, Ohe-Toyota M, Herman JG, Baylin SB, Issa J-PJ. CpG island methylator phenotype in colorectal cancer. *Proc Natl Acad Sci U S A.* 1999; 96: 8681-6.
- Flavahan WA, Gaskell E, Bernstein BE. Epigenetic plasticity and the hallmarks of cancer. *Science.* 2017; 357: eaal2380.
- 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: 963-75.
- Hanahan D. Hallmarks of cancer: new dimensions. *Cancer Discov.* 2022; 12: 31-46.
- 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: 787-93.
- Ogino S, Nosho 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: 90-6.
- Liu Z, Hu Y, Xie H, Chen K, Wen L, Fu W, et al. Single-cell chromatin accessibility analysis reveals the epigenetic basis and signature transcription factors for the molecular subtypes of colorectal cancers. *Cancer Discov.* 2024; 14: 1082-105.
- Dienstmann R, Vermeulen L, Guinney J, Kopetz S, Tejpar S, Tabernero J. Consensus molecular subtypes and the evolution of precision medicine in colorectal cancer. *Nat Rev Cancer.* 2017; 17: 79-92.
- Marusyk A, Janiszewska M, Polyak K. Intratumor heterogeneity: the rosetta stone of therapy resistance. *Cancer Cell.* 2020; 37: 471-84.
- Bian S, Hou Y, Zhou X, Li X, Yong J, Wang Y, et al. Single-cell multiomics sequencing and analyses of human colorectal cancer. *Science.* 2018; 362: 1060-3.
- 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.
- Gaudet F, Hodgson JG, Eden A, Jackson-Grusby L, Dausman J, Gray JW, et al. Induction of tumors in mice by genomic hypomethylation. *Science.* 2003; 300: 489-92.
- Bushweller JH. Targeting transcription factors in cancer — from undruggable to reality. *Nat Rev Cancer.* 2019; 19: 611-24.

Cite this article as: Liu Z, Zhou X, Tang F. Epigenetic regulators as the foundation for molecular classification of colorectal cancer. *Cancer Biol Med.* 2024; x: xx-xx. doi: 10.20892/j.issn.2095-3941.2024.0176