

# Comprehensive characterization of tumor microenvironment in colorectal cancer via molecular analysis

Xiangkun Wu<sup>1,2†</sup>, Hong Yan<sup>1,2,3†</sup>, Mingxing Qiu<sup>1,2†</sup>, Xiaoping Qu<sup>4,5†</sup>, Jing Wang<sup>1,2</sup>, Shaowan Xu<sup>1,2</sup>, Yiran Zheng<sup>4,5</sup>, Minghui Ge<sup>4,5</sup>, Linlin Yan<sup>4,5</sup>, Li Liang<sup>1,2,6\*</sup>

<sup>1</sup>Department of Pathology, Nanfang Hospital/School of Basic Medical Sciences, Southern Medical University, Guangzhou, China; <sup>2</sup>Department of Pathology and Guangdong Province Key Laboratory of Molecular Tumor Pathology, School of Basic Medical Sciences, Southern Medical University, Guangzhou, China; <sup>3</sup>Department of Pathology, The First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, China; <sup>4</sup>Nanjing Simcere Medical Laboratory Science Co., Ltd, Nanjing, China; <sup>5</sup>State Key Laboratory of Translational Medicine and Innovative Drug Development, Jiangsu Simcere Diagnostics Co., Ltd, Nanjing, China; <sup>6</sup>Jinfeng Laboratory, Chongqing, China

**Abstract** Colorectal cancer (CRC) remains a challenging and deadly disease with high tumor microenvironment (TME) heterogeneity. Using an integrative multi-omics analysis and artificial intelligence-enabled spatial analysis of whole-slide images, we performed a comprehensive characterization of TME in colorectal cancer (CCCRC). CRC samples were classified into four CCCRC subtypes with distinct TME features, namely, C1 as the proliferative subtype with low immunogenicity; C2 as the immunosuppressed subtype with the terminally exhausted immune characteristics; C3 as the immune-excluded subtype with the distinct upregulation of stromal components and a lack of T cell infiltration in the tumor core; and C4 as the immunomodulatory subtype with the remarkable upregulation of anti-tumor immune components. The four CCCRC subtypes had distinct histopathologic and molecular characteristics, therapeutic efficacy, and prognosis. We found that the C1 subtype may be suitable for chemotherapy and cetuximab, the C2 subtype may benefit from a combination of chemotherapy and bevacizumab, the C3 subtype has increased sensitivity to the WNT pathway inhibitor WIK14, and the C4 subtype is a potential candidate for immune checkpoint blockade treatment. Importantly, we established a simple gene classifier for accurate identification of each CCCRC subtype. Collectively our integrative analysis ultimately established a holistic framework to thoroughly dissect the TME of CRC, and the CCCRC classification system with high biological interpretability may contribute to biomarker discovery and future clinical trial design.

## Editor's evaluation

This study represents a valuable body of work in which the authors assemble a molecular description of colorectal cancer and a classification into subtypes. Overall, the evidence supporting the findings is solid, and consensus over a diverse range of data from publicly available sources is convincing. When added to existing knowledge this work may contribute to future biomarker discoveries for colorectal cancer.

\*For correspondence:  
lli@smu.edu.cn

†These authors contributed  
equally to this work

Competing interest: See page  
29

Funding: See page 29

Received: 08 January 2023

Preprinted: 20 February 2023

Accepted: 10 May 2023

Published: 02 June 2023

Reviewing Editor: Paul R Barber,  
King's College London, United  
Kingdom

© Copyright Wu, Yan, Qiu  
et al. This article is distributed  
under the terms of the [Creative  
Commons Attribution License](#),  
which permits unrestricted use  
and redistribution provided that  
the original author and source  
are credited.

## Introduction

Colorectal cancer (CRC) is the third most deadly malignancy worldwide (*Siegel et al., 2023*), and the incidence of early-onset CRC is steadily increasing (*Archambault et al., 2021*). CRC at early and localized stages is primarily a preventable and curable disease, but up to 50% of patients with locally advanced disease eventually develop mCRC (*Andrei et al., 2022; Ciardiello et al., 2022*). Therefore, the clinical systematic management of CRC patients is still an unmet medical challenge (*Ciardiello et al., 2022*).

With the development of high-throughput technologies and bioinformatics strategies, multi-omics data are used to identify and characterize the molecular subtypes of CRC, such as genomics (*Zhao et al., 2022*), transcriptomics (*Budinska et al., 2013; De Sousa E Melo et al., 2013; Marisa et al., 2013; Roepman et al., 2014; Sadanandam et al., 2013; Schlicker et al., 2012*), and proteomics (*Li et al., 2020*). The consensus molecular subtype (CMS) integrates six independent classification systems based on transcriptomics; however, it is still not explicitly used to guide clinical treatment (*Guinney et al., 2015*). The Cancer Genome Atlas Program (TCGA) and Clinical Proteomic Tumor Analysis Consortium (CPTAC) colorectal studies have dissected the molecular heterogeneity of CRC by integrating multi-omics data (*Cancer Genome Atlas Network, 2012; Vasaikar et al., 2019*). Nevertheless, multi-omics data are complex and highly dimensional, and extracting valuable information from these data to guide clinical treatment is still a tremendous challenge (*Leng et al., 2022*). By reviewing the biological characteristics of the tumor, useful information can be screened for identifying molecular subtypes.

The tumor cells can interact with cellular or non-cellular components, triggering dramatic molecular, cellular, and physical changes in the tumor microenvironment (TME) to build a self-sustainable tumor ecosystem (*Anderson and Simon, 2020; Chen and Song, 2022*). Simultaneously, TME profoundly affects tumor biology, responses to therapy, and clinical outcomes, which is a dynamic network mainly comprised of immune components and stromal components (*Hirata and Sahai, 2017; Jia et al., 2022; Zhang et al., 2022*). Furthermore, TME can adversely affect the metabolic activities of tumor, immune and stromal cells, and form diverse metabolic phenotypes (*Elia and Haigis, 2021; Kaymak et al., 2021*). Identifying the components of the TME and their functions, as well as the crosstalk between tumor cells and TME contributes to our understanding of the clinical heterogeneity of CRC, thereby bringing about new advances in precision medicine. Previous studies have used immune or stromal components of the TME, or a combination of both, to study the TME (*Bagaev et al., 2021; He et al., 2018*), but they are insufficient to completely reconstruct the heterogeneity of the TME.

In this study, we considered the tumor cells and its TME as a whole and performed a comprehensive characterization of TME in colorectal cancer (CCRC), including the functional states of the tumor cells, immune and stromal signatures, and metabolic reprogramming features. We successfully identified the four CCRC subtypes based on 61 TME-related signatures. Integrated analyses determined that the CCRC subtypes had distinct histopathologic and molecular characteristics, therapeutic efficacy, and prognosis.

## Results

### Establishment of the TME panel

The molecular and clinical features of a tumor are characterized by the functional states of tumor cells, as well as other TME-related signatures, including immune and stromal components, and metabolic reprogramming signatures. In brief, 15 signatures (including angiogenesis, apoptosis, cell cycle, differentiation, DNA damage, DNA repair, endothelial-to-mesenchymal transition (EMT), hypoxia, inflammation, invasion, metastasis, proliferation, quiescence, stemness, and cancer stem cells) were used to describe the functional states of tumor cells. As for the immune signatures, we focused on eight categories of immune cells (T cells, natural killer cells, dendritic cells, macrophages, myeloid-derived suppressor cells, B cells, mast cells, and neutrophils) and their subpopulations, as well as the other immune-related signatures. In addition to the signatures of endothelial cells, mesenchymal cells, and mesenchymal stem cells, we included signatures of extracellular matrix, matrix remodeling and interactions of cells with the extracellular matrix to characterize the stromal compartments. A total of seven major metabolic pathways (amino acid, nucleotide, vitamin cofactor, carbohydrate, TCA cycle, energy, and lipid metabolism) were used to reveal the metabolic reprogramming of the TME. According to

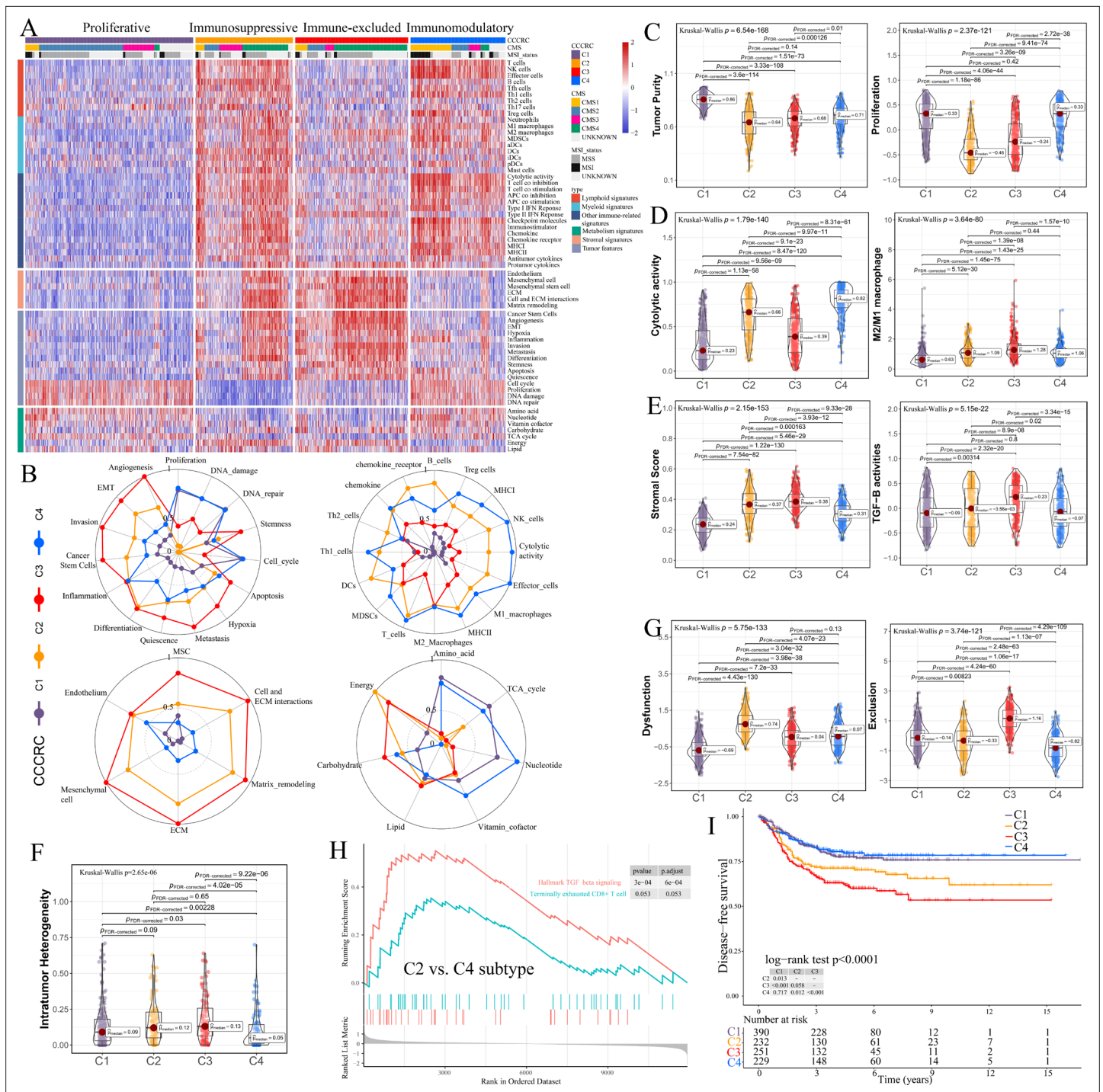
the above biological framework, a total of 61 TME-related signatures were collected to form the TME panel (**Supplementary file 1a, Supplementary file 1b**), which ultimately established a holistic approach to thoroughly dissect the TME of CRC.

Principal component analysis (PCA) indicated that after using the 'ComBat' function to remove batch effects, there was no significant batch effect in the merged cohorts of eight microarray datasets (**Figure 1—figure supplement 1A, B**) and two RNA sequencing datasets (**Figure 1—figure supplement 1C, D, Supplementary file 1c**). We used Gene set variation analysis (GSVA) to calculate the TME-related signature scores for each sample in each cohort. Principal coordinate analysis (PCOA) revealed that the CRC samples could be distinguished from normal samples by the TME-related signatures in the GSE39582 and TCGA cohorts (**Figure 1—figure supplement 1E**). We further focused on the signatures of the functional states of tumor cells, which could classify CRC and normal samples (**Figure 1—figure supplement 1F**). The p-values for intercomparisons of the 'Euclidean' distances between normal and CRC samples were all  $<0.05$  using permutational multivariate analysis of variance (PERMANOVA) test (**Zhu et al., 2021**). Most immune signatures had higher GSVA scores in the normal samples compared with the CRC samples, while stromal signatures and the signatures of the functional states of tumor cells had higher GSVA scores in CRC tissues (**Figure 1—figure supplement 1G, H**). As expected, amino acid, carbohydrate, and nucleotide metabolic processes were more prominent in CRC samples, which was consistent with the hallmark of infinite proliferation of tumor cells (**Figure 1—figure supplement 1G, H**).

Spearman's correlation analysis of the TME-related signatures revealed three major patterns bound by positive correlations in the CRC-AFFY cohort (**Figure 1—figure supplement 1I**). One pattern defining the proliferation of tumor cells consisted of cell cycle and metabolic reprogramming signatures. The second was mainly comprised of immune components, such as T cells, natural killer cells (NK cells), myeloid-derived suppressor cells (MDSCs), and M2 macrophages. The third pattern was associated with stromal components such as angiogenesis and extracellular matrix, as well mesenchymal cells and cancer stem cells. In addition, our findings indicated a strong positive correlation between lymphocytic and stromal signatures and MCP-counter algorithm-derived signatures, thereby emphasizing the robustness of the employed methods (**Figure 1—figure supplement 1J**). Meanwhile, we found that 15 signatures associated with the functional states of tumor cells were positively correlated with the activity of 10 classical oncogenic pathways (**Figure 1—figure supplement 1K**). Finally, we used the Kaplan–Meier method and univariate Cox proportional hazard regression analysis to evaluate the prognosis of the TME-related signatures. The stromal and tumor components significantly were correlated with decreased survival, particularly in the case of mesenchymal cells, endothelial cells, metastasis, differentiation, and EMT signatures (**Figure 1—figure supplement 1L, Supplementary file 1d**). Lymphoid-associated cells generally tended to be associated with a better prognosis, while myeloid-associated cells generally tended to be associated with a poor prognosis. Among the metabolism-related signatures, energy metabolism and carbohydrate metabolism were significantly related to poor prognosis, while nucleotide metabolism, amino acid metabolism, and TCA metabolism were strongly predictive of a favorable prognosis. Collectively, our findings demonstrated that the TME heterogeneity, including unique differences in immune, stromal, and metabolic reprogramming, played a crucial role in tumor development, and that the TME panel could be used to comprehensively characterize CRC.

## Identification and validation of CCCRC classification

With the increasing application of immunotherapy and tumor vaccines, there is growing evidence highlighting the importance of the TME in tumorigenesis and development (**Bejarano et al., 2021; Saxena et al., 2021**). To reveal the TME heterogeneity of CRC using the curated TME panel, consensus clustering analysis was performed based on the 61 TME-related signature scores in the CRC-AFFY cohort, and the optimal cluster number was determined to be four using the consensus matrices heatmap, the empirical cumulative distribution function (CDF) plot, and delta area plot (**Wilkerson and Hayes, 2010**) (Materials and methods, **Figure 1—figure supplement 2A–C**). Subsequently, the CRC samples in the CRC-AFFY cohort were classified into the four CCCRCs with distinct TME components (**Figure 1A, B, Figure 1—figure supplement 2D**). To evaluate the reproducibility of the CCCRC subtypes, we utilized the PAM (Prediction Analysis of Microarrays) algorithm (**Tibshirani et al., 2002**) to extract 61 TME-related signatures that best represent each subtype, using a threshold of 0.566



**Figure 1.** Comprehensive characterization of colorectal cancer (CCCRC). **(A)** Heatmap of 1471 colorectal cancer (CRC) patients in the CRC-AFFY cohort classified into four distinct tumor microenvironment (TME) subtypes based on the 61 TME-related signatures. CMS: consensus molecular subtypes; MSI: microsatellite instability; MSS: microsatellite stability. **(B)** Radars display the characteristic TME-related signatures, including tumor, immune, stroma, and metabolism signatures, of each CCCRC subtype in the CRC-AFFY cohort. MSC: mesenchymal stem cells; ECM: extracellular matrix. Box plots show differences in tumor **(C)**, immune **(D)**, and stroma **(E)** signatures in the CRC-AFFY cohort. Tumor purity and stroma scores were obtained from the ESTIMATE algorithm. Proliferative activity (proliferation), cytolysis score, M1 and M2 macrophage proportions, and TGFB activity were calculated by Gene set variation analysis (GSVA). **(F)** Differences in intratumor heterogeneity between four CCCRC subtypes in CRC-RNAseq cohort. **(G)** Differences in T cell dysfunction and T cell exclusion scores between four CCCRC subtypes were analyzed based on the gene expression profiles in CRC-AFFY cohort. **(H)** Gene set enrichment analysis (GSEA) of the terminally exhausted CD8+ T cell signature and the TGF-beta signaling signature between C2 and C4

Figure 1 continued on next page

Figure 1 continued

subtypes in the CRC-AFFY cohort. (I) Kaplan–Meier method with log-rank test of disease-free survival among the four CCCRC subtypes in the CRC-AFFY cohort.

The online version of this article includes the following figure supplement(s) for figure 1:

**Figure supplement 1.** Establishment of tumor microenvironment (TME) panel.

**Figure supplement 2.** Comprehensive characterization of colorectal cancer (CCCRC).

**Figure supplement 3.** Differences in the tumor microenvironment (TME) components obtained from other algorithm among the comprehensive characterization of colorectal cancer (CCCRC) subtypes.

**Figure supplement 4.** Limiting the consensus clustering analysis to only immune-related or immune- and stroma-related signatures in the CRC-AFFY cohort.

**Figure supplement 5.** Overlap of the comprehensive characterization of colorectal cancer (CCCRC) subtypes with published colorectal cancer (CRC) molecular subtypes in the CRC-AFFY and CRC-RNAseq cohorts.

**Figure supplement 6.** Associations between comprehensive characterization of colorectal cancer (CCCRC) subtypes and consensus molecular subtypes (CMS).

**Figure supplement 7.** Survival analyses of the comprehensive characterization of colorectal cancer (CCCRC) subtypes and consensus molecular subtypes (CMS).

(Figure 1—figure supplement 2E). These signatures were then used to construct a PAMR classifier with superior predictive capability, exhibiting an overall error rate of 15%. The PAMR classifier based on the PAM algorithm is publicly available at [https://github.com/XiangkunWu/PAMR\\_classifier](https://github.com/XiangkunWu/PAMR_classifier) (copy archived at Wu, 2023a). We used the established PAMR classifier to predict the CCCRC subtypes on the CRC-RNAseq cohort and the same four CCCRC subtypes were revealed, with similar patterns of differences in the TME components (Figure 1—figure supplement 2F, G). PCOA showed that the four CCCRC subtypes were distinctly separated and the p-values for intercomparisons of the ‘Euclidean’ distances between them were all <0.05 using PERMANOVA test (Figure 1—figure supplement 2H) in the CRC-RNAseq and CRC-AFFY cohorts. PCOA also demonstrated highly similar TME compartments in the same subtype between the CRC-RNAseq and CRC-AFFY cohorts. Differences in the TME components between the CCCRC subtypes were also observed in the analysis of previously reported immune and stromal signatures obtained by the microenvironment cell populations (MCP)-counter, CIBERSORT, and ESTIMATE algorithm, and 10 classical oncogenic pathway activities and 86 metabolic pathway enrichment scores calculated by GSVA (Figure 1—figure supplement 3A–E, Supplementary file 1e). Notably, limiting the consensus clustering analysis to only immune-related or immune- and stroma-related signatures, as done in previous studies (Bagaev et al., 2021; He et al., 2018), did not allow reliable identification of all four CCCRC subtypes (Figure 1—figure supplement 4). These sensitivity analyses underscored the necessity of our well-designed TME panel to achieve this identification.

C1 (35% of all tumors), hereafter designated as the proliferative subtype, was characterized by the relative upregulation of tumor proliferative activity, tumor purity, and minimal or complete lack of lymphocyte and stromal infiltration, which was highly similar to the cold tumor phenotype (Figure 1B–E). The MYC, cell cycle, TP53, and PI3K pathways associated with tumor proliferation had the highest GSVA scores in the C1 subtype (Figure 1—figure supplement 3E). C2 (21% of all tumors), hereafter designated as the immunosuppressed subtype, was characterized by the relative upregulation of immune and stromal components, such as T cells, M2 macrophages, and cancer-associated fibroblasts (CAFs) (Figure 1B–E, Figure 1—figure supplement 3A–D). However, the extent of infiltration of effector cells, as well as the cytolytic score, was much lower than that of the C4 subtype. C3 (24% of all tumors), hereafter designated as the immune-excluded subtype, was characterized by the distinct upregulation of stromal components, such as CAFs, and cancer stem cells, as well as angiogenesis and hypoxia signatures (Figure 1B–E, Figure 1—figure supplement 3A–D). During tumor progression, TGF-beta secreted by CAFs is leveraged by tumor cells to suppress and exclude the anti-tumor immune components (Liu et al., 2019). We observed that the TGF-beta pathway, as well as WNT, NOTCH, and RAS pathways, and the ratio of M2/M1 macrophages, were distinctly upregulated in C2 and C3 subtypes (Figure 1D, E, Figure 1—figure supplement 3E). The scores of 5/10 oncogenic pathways were the highest in the C3 subtype (Figure 1—figure supplement 3E), suggesting that the activation of oncogenic pathways could lead to the formation of immune-excluded phenotypes which

was consistent with the previous theory (*Galon and Bruni, 2019*). The level of intratumor heterogeneity (ITH) was significantly linked to poor prognosis and drug resistance (*Caswell and Swanton, 2017*). As expected, the ITH of the C2 and C3 subtypes was higher than that of the other subtypes (*Figure 1F*). C4 (20% of all tumors), hereafter designated as the immunomodulatory subtype, was characterized by the remarkable upregulation of anti-tumor immune components, such as effector T cells, NK cells, and Th1 cells. The C4 subtype also had the highest cytolytic score compared with the other subtypes and lacked stromal components and the other immunosuppressed components, which indicated an immunomodulatory microenvironment (*Figure 1B–E*).

To further explore the immune escape mechanism of each CCCRC subtype, the differences in T cell dysfunction and T cell exclusion scores between the four CCCRC subtypes were analyzed based on the gene expression profiles (GEP), which reflected the T cell function of the global tumor (*Jiang et al., 2018*). Strikingly, the C2 subtype had highest T cell dysfunction score, indicating that T cell exhaustion in the C2 subtype was at the late stage (*Figure 1G, Figure 1—figure supplement 3F*). Using gene set enrichment analysis (GSEA) with all genes ranked according to the fold change (FC) between C2 and C4 subtypes, we found that terminally exhausted CD8+ T cell and TGF-beta signaling signatures were upregulated in the C2 subtype in the CRC-AFFY (*Figure 1H*) and CRC-RNAseq (*Figure 1—figure supplement 3G*) cohorts, which might reveal that CD8+ T cell infiltration within the tumor bed was suppressed by the stroma and was in a late state of exhaustion. The C3 subtype had the highest T cell exclusion score (*Figure 1G, Figure 1—figure supplement 3F*), demonstrating that the low T cell infiltration into the tumor bed was due to the increased abundance of CAFs and M2 macrophages, thereby leading to the exclusion of T cells from the tumor bed. Metabolic reprogramming also differed significantly among the four CCCRC subtypes (*Figure 1B, Figure 1—figure supplement 3H*). We analyzed the 86 metabolic pathways obtained from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (*Supplementary file 1e*) and observed that the number of upregulated metabolic pathways of the C3 subtype was the lowest. We also found that glycan metabolism was distinctly upregulated in C2 and C3 subtypes, which indicated that glycan metabolism was significantly associated with the stroma.

## Associations between CCCRC subtypes and other molecular subtypes and clinical characteristics

Previous studies have identified several molecular subtypes of CRC based on GEP (*Budinska et al., 2013; De Sousa E Melo et al., 2013; Guinney et al., 2015; Roepman et al., 2014; Sadanandam et al., 2013*). We investigated their associations with the CCCRC subtypes in the CRC-AFFY and CRC-RNAseq cohorts (*Figure 1—figure supplement 5*). The C1 subtype was primarily comprised of the CMS2 and lower crypt-like subtypes, and it contained the highest frequencies of the CCS1, B-type, and TA subtypes. The C2 subtype mainly consisted of the CMS4, stem-like, surface crypt-like, CCS3, and C-type subtypes, and included the highest frequency of the enterocyte subtype. The C3 subtype contained the highest frequencies of CMS4, stem-like, CCS3, and C-type subtypes and was mainly comprised of the mesenchymal and TA subtypes. The C4 subtype included the highest frequencies of microsatellite instability (MSI) and the CMS1, CIMP-H-like, A-type, and inflammatory subtypes, and was mainly comprised of the CCS2 subtype.

We also focused on the differences in the TME components between the CCCRC and the CMS subtypes. Compared with the CMS1 subtype, the C4 subtype showed upregulated anti-tumor immune components and lacked immunosuppressive components (*Figure 1—figure supplement 6A*). CRC patients with MSI were sensitive to immune checkpoint blockade (ICB) therapy (*Jin and Sinicrope, 2022*), and C4 and CMS1 subtypes containing approximately 47 and 75% of MSI cases, respectively. The C4 subtype with MSI showed upregulated scores of effector cells and cytolytic activity and down-regulated scores of extracellular matrix and matrix remodeling compared with the CMS1 subtype with MSI (*Figure 1—figure supplement 6B*). Moreover, we observed that the C4 subtype with MSI and the C4 subtype with MSS had higher scores of anti-tumor immune signatures and lower scores of stromal components than the other subtype with MSI (*Figure 1—figure supplement 6C*). We also observed that CMS2 subtype contained more C4 subtypes in addition to mainly C1 subtypes. Therefore, we analyzed the differences in the TME components between C1 and CMS2 subtypes and found that CMS2 subtypes indeed had higher immune-related components than C1 subtypes, such as MCH-I, MCH-II, and inflammatory signature, and also contained more stromal components, such

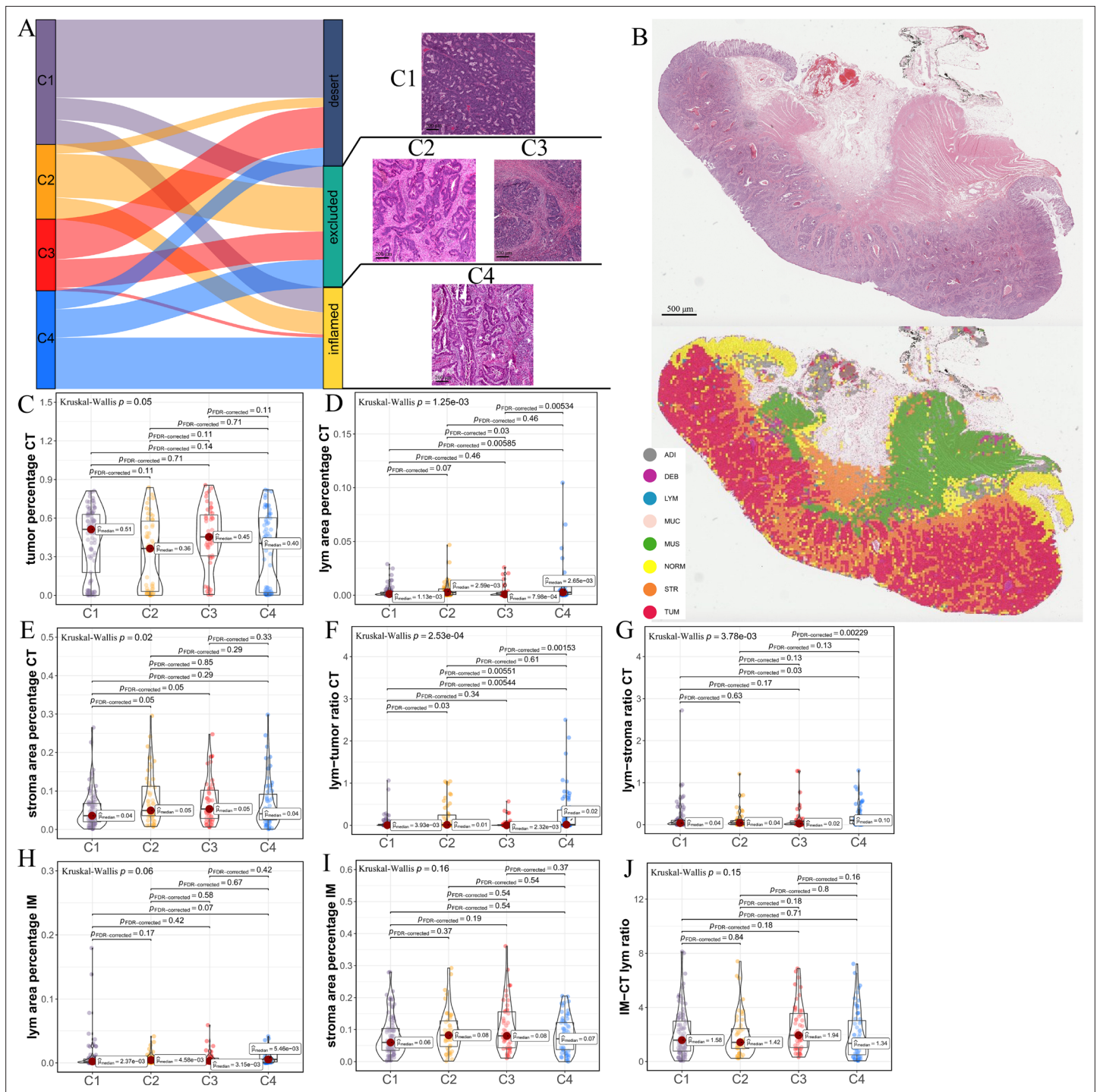
as extracellular matrix, than C1 subtypes (**Figure 1—figure supplement 6D**). This suggested that the C1 subtype was less immunogenic than CMS2 and more closely resembles cold tumor characteristics. Specifically, we found that CMS4 contained mainly C2 and C3 subtypes. Our findings indicated that the C2 subtypes within CMS4 exhibited a higher abundance of immune components, such as T cells and NK cells, compared to the C3 subtype within CMS4. However, the differences in stromal components between these subtypes were not statistically significant (**Figure 1—figure supplement 6E**).

We further analyzed the association of CCCRC subtypes with clinicopathological characteristics (**Supplementary file 1f, Supplementary file 1g**). We found that the C4 subtype was mostly diagnosed in right-sided CRC lesions and in females, which was consistent with the CMS1 subtype. The C1 and C3 subtypes were mainly observed in left-sided CRC lesions and in males, consistent with the CMS2 and CMS4 subtypes. The C3 subtype was strongly associated with more advanced tumor stages, which was the similarity to the CMS4 subtype, while the C4 subtype was associated with higher histopathologic grade, which was the similarity to the CMS1 subtype. Furthermore, our analysis using the Kaplan–Meier method demonstrated that patients with the C4 subtype had significantly higher disease-free survival (DFS) and overall survival (OS) compared to those with the C2 and C3 subtypes in the CRC-AFFY (**Figure 11, Figure 1—figure supplement 7A**) and CRC-RNAseq cohorts (**Figure 1—figure supplement 7B, C**). Multivariate Cox proportional hazard regression analysis showed that the C4 subtype was an independent predictor of the best OS and DFS, whereas the C3 subtype was an independent predictor of the worst OS and DFS after adjustment for age, gender, tumor site, TNM stage, grade, adjuvant chemotherapy or not, MSI status, BRAF and KRAS mutations, and the CMS classification system in the combined cohort (the CRC-AFFY and CRC-RNAseq cohorts) (**Supplementary file 1h**). Considering that the C1, C2/C3, and C4 subtypes partially overlap with the CMS2, CMS4, and CMS1 subtypes, respectively, we also analyzed the prognostic differences between them in the combined cohort. We found that the DFS/OS of patients with the C1 subtype was worse than those with the CMS2 subtype (**Figure 1—figure supplement 7D, E**), the DFS/OS of patients with the C2 subtype was better than those with the CMS4 subtype (**Figure 1—figure supplement 7F, G**), the DFS/OS of patients with the C3 subtype was not significantly different from those with the CMS4 subtype (**Figure 1—figure supplement 7F, G**), and the DFS/OS of patients with the C4 subtype was significantly better than those with the CMS1 subtype (**Figure 1—figure supplement 7H, I**). Notably, the C2 subtype within the CMS4 subtype also had a better prognosis than the C3 subtype within the CMS4 subtype (**Figure 1—figure supplement 7J, K**). The above analysis demonstrated that the CCCRC classification system was closely associated with clinicopathological characteristics, were able to refine the CMS classification system and MSI status, as well as contributed to the understanding of the mechanisms underlying the different clinical phenotypes resulting from TME heterogeneity.

## Differences in histopathologic characteristics between CCCRC subtypes

To further explore the biological differences between CCCRC subtypes, we investigated the histopathologic phenotypes by evaluating the whole-slide images (WSIs) of the TCGA-CRC dataset. We compared our CCCRC system with the three-category immune classification system of solid tumors, termed ‘desert’, ‘excluded’, and ‘inflamed’ phenotypes (**Chen and Mellman, 2017; Rosenberg et al., 2016**). Two pathologists evaluated the histopathologic characteristics for each subtype under the microscope. The 254 CRC samples in the TCGA-CRC dataset were categorized as these three phenotypes based on the abundance of lymphocytes and their spatial location with malignant epithelial cells (**Supplementary file 1i**). According to the three-category immune classification system, the C4 subtype was enriched with an inflamed phenotype characterized by abundant lymphocytes in direct contact with malignant cells (**Figure 2A**). The C2 subtype was mostly categorized as an excluded phenotype. The C1 and C3 subtypes were mainly classified into the desert phenotype, whereas the C3 subtype was more frequently classified as an excluded phenotype than the C1 subtype. Notably, the lymphocytes of C2 subtype were more frequently intermixed with intratumor stromal components, whereas the lymphocytes of C3 subtype were more frequently excluded from the tumor bed and intermixed with adjacent-tumor stromal components, both of which were classified as excluded phenotype according to the three-category immune classification system.

The above differences in the histopathologic characteristics among the CCCRC subtypes were based on the semi-quantitative analysis results of two pathologists, which are subjective to a certain extent. Therefore, we used hematoxylin and eosin (HE)-stained image-based deep learning to establish the



**Figure 2.** Differences in histological characteristics between comprehensive characterization of colorectal cancer (CCRC) subtypes. **(A)** Sankey plot shows overlap of CCRC subtypes with the three-category immune classification system ('desert', 'excluded', and 'inflamed' phenotypes), and their representative hematoxylin and eosin (HE)-stained whole-slide images (WSIs). C1: TCGA-AA-3955; C2: TCGA-A6-6654; C3: TCGA-CK-4948; and C4: TCGA-AD-6963. Scale bars 200  $\mu$ m. **(B)** Representative WSI (top) and the colorectal cancer (CRC)-multiclass model-inference segmentation of eight tissue types: tumor, stroma, lymphocyte, normal colon mucosa, debris, adipose, mucin, and muscle (bottom). Scale bar 500  $\mu$ m. Box plots show differences in the abundance of tumors **(C)**, lymphocyte infiltration (lym) **(D)**, and stroma **(E)** in the core tumor (CT) region. Box plots show differences in the lymphocyte infiltration to tumor content ratio **(F)** and lymphocyte infiltration to stromal content ratio **(G)** in the CT region. Box plots show differences in the abundance of lymphocytes infiltration **(H)** and stroma **(I)** in the invasive margin (IM) region. **(J)** Box plots show differences in the ratio of lymphocyte infiltration in the IM region to the CT region.

Figure 2 continued on next page



Figure 2 continued

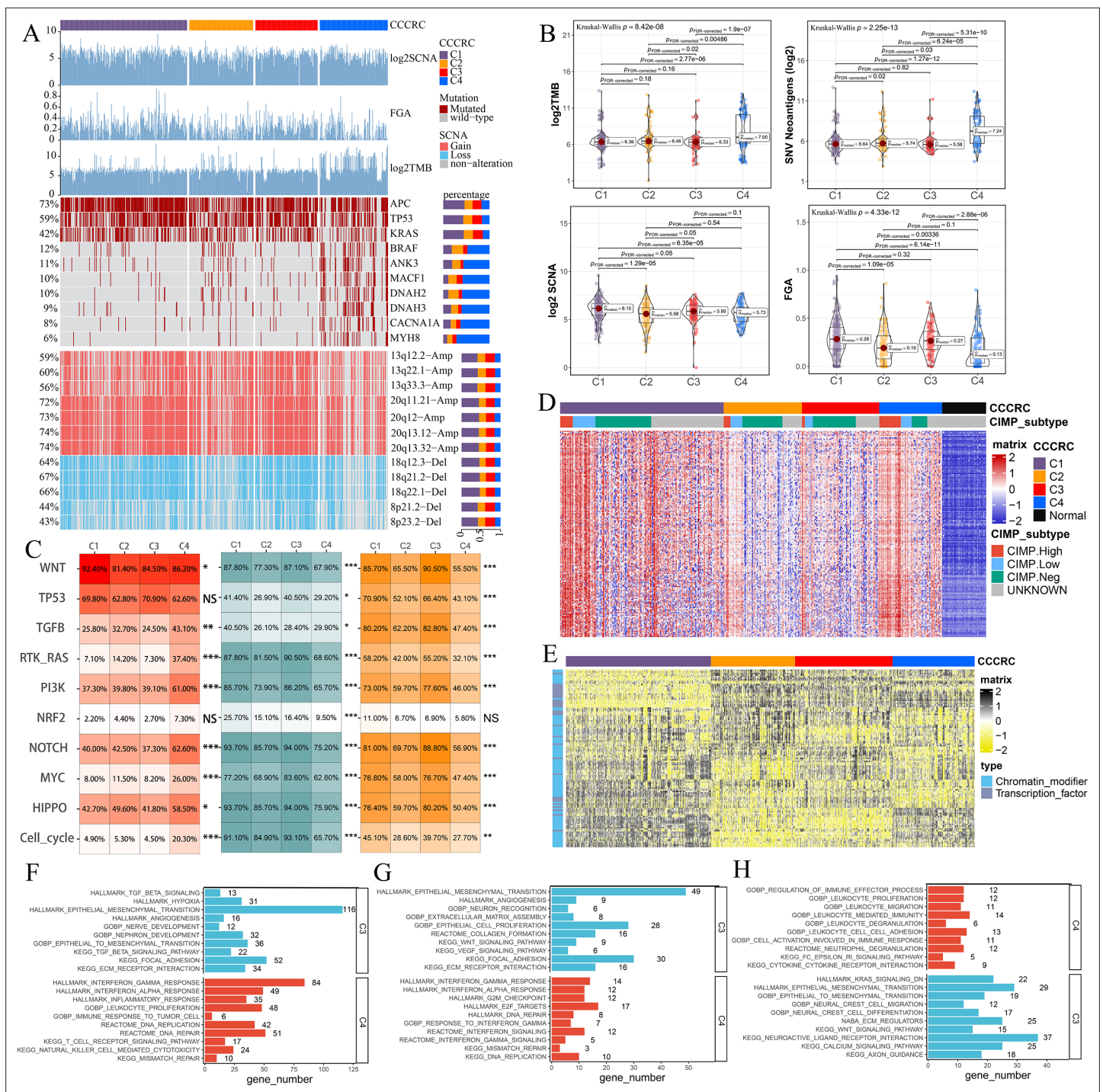
The online version of this article includes the following figure supplement(s) for figure 2:

**Figure supplement 1.** The performance of the colorectal cancer (CRC)-multiclass model: a muscle/non-muscle classifier a seven-tissue classifier.

CRC-multiclass model and to evaluate the abundance and spatial distribution of the tumor, lymphocytes, and stroma. The CRC-multiclass model consisted of a muscle/non-muscle classifier and a seven-tissue classifier that could classify eight CRC tissue types: adipose (ADI), debris (DEB), lymphocytes (LYM), mucus (MUC), smooth muscle (MUS), normal colon mucosa (NORM), cancer-associated stroma (STR), and colorectal adenocarcinoma epithelium (TUM) (Materials and methods). The muscle/non-muscle model demonstrated well performance on the internal test set of 14,681 patches, achieving an area under the curve (AUC) of approximately 0.99 and an accuracy of 0.99 (Figure 2—figure supplement 1A, B). Meanwhile, the seven-tissue classifier also performed well, with AUCs for the different tissue types above 0.99 and an accuracy of 0.95 on the internal test set of 5741 patches (Figure 2—figure supplement 1C, D). When evaluated on an external test set of 4288 patches, the muscle/non-muscle model achieved an AUC of 0.95 and an accuracy of 0.91 (Figure 2—figure supplement 1E, F). Of the 3633 patches identified as non-muscle by the muscle/non-muscle model, the seven-tissue classifier achieved AUCs ranging from 0.97 to 1 for different tissue types and an accuracy of 0.91 (Figure 2—figure supplement 1G, H). The tissue heatmap showed our model prediction results for a TCGA-CRC WSI (Figure 2B). In the core tumor (CT) region, the C1 subtype had a highly increased abundance of the tumor; the C4 subtype had increased lymphocyte infiltration and decreased stromal content; the C2 subtype had elevated lymphocyte and stromal infiltration; and the C3 subtype had the highest abundance of stroma, but less lymphocyte infiltration was detected (Figure 2C–E). We also observed that C4 subtype had the highest lymphocyte infiltration to tumor content ratio and lymphocyte infiltration to stromal content ratio, followed by C2 subtype and C3 subtype had the lowest (Figure 2F, G). In the invasive margin (IM) region, different degrees of lymphocyte infiltration and stromal components were observed for each subtype (Figure 2H,I). Importantly, the ratio of lymphocyte infiltration in the IM region of the C3 subtype to the CT region was the highest, which confirmed that the stromal components excluded lymphocytes from the CT region in the C3 subtype (Figure 2J). AI-enabled spatial analysis of WSIs confirmed the semi-quantitative results of the pathologists, with the C1 subtype belonging to the cold phenotype, C2 subtype belonging to the immunosuppressive phenotype, C3 subtype belonging to the excluded phenotype, and C4 subtype belonging to the hot phenotype. Collectively, our CCCRC system further refined the three-category immune classification system of solid tumors (Chen and Mellman, 2017; Rosenberg et al., 2016) and conformed to the four-category immune classification system, termed ‘hot’, ‘cold’, ‘immune-excluded’, and ‘immunosuppressive’ phenotypes (Galon and Bruni, 2019; Kirchhammer et al., 2022).

### Biological characterization of CCCRC subtypes

We further elucidated the differences in biological characteristics among the CCCRC subtypes using multi-omics data, including genomics, epigenetics, transcriptomics, and proteomics data. As for the genomic alterations, the C4 subtype had the highest TMB and neoantigen values and the lowest prevalence of chromosomal instability (CIN), including somatic copy number alteration (SCNA) counts and fraction of the genome altered (FGA) scores, compared with the other subtypes (Figure 3A, B). Conversely, C1 and C3 subtypes displayed the highest CIN levels, as described by SCNA counts and FGA scores, and the lowest TMB and neoantigen values (Figure 3A, B). The C2 subtype displayed median CIN levels, TMB and neoantigen values. Among the frequently mutated genes (>5%), the mutation frequencies of APC (85.8%), TP53 (64.9%), and KRAS (46.7%) were the highest in the C1 subtype compared to the other subtypes (all p-value <0.05), followed by the C3, C2, and C4 subtypes, which are closely associated with the occurrence of CRC (Figure 3A, Supplementary file 2a). The C4 subtype was significantly enriched in mutations of DNAH2 (26.0%), MYH8 (26.8%), and BRAF (26.0%) genes (all p-value <0.05), whereas the mutation frequency of C1, C2, and C3 subtypes was low. In terms of the differences in SCNA, the C1 subtype with the highest CIN level harbored significantly more amplified chromosomal regions (20q12, 20q13.12, 20q11.21, and 20q13.32) and deleted chromosomal regions (18q21.2, 18q22.1, and 18q12.3) (all p-value <0.05) (Figure 3A, B, Supplementary file 2b). The C3 subtype was significantly enriched in the amplified chromosomal regions of 13q33.3, 13q22.1, and 13q12.2 and the deleted chromosomal regions of 8p21.2 and 8p23.2 (all, p-value <0.05).



**Figure 3.** Biological characterization of comprehensive characterization of colorectal cancer (CCRC) subtypes based on multi-omics data. **(A)** Distribution of driver gene mutations and somatic copy number alterations (SCNAs) among the CCCRC subtypes in the TCGA-CRC dataset. **(B)** Box plots show differences in tumor mutation burden (TMB), neoantigens, SCNA counts, and fraction of the genome altered (FGA) scores among the four CCCRC subtypes in the TCGA-CRC dataset. The color of the box represents the different types of genomic alterations (red, mutation; blue, amplification; yellow, deletion), and the color saturation represents the frequency. The color of the p-value represents which oncogenic pathway had the highest frequency of the genomic alterations. Heatmap shows differentially methylated genes derived from each CCCRC subtype versus normal tissues **(D)** and regulon activity profiles for transcription factors and chromatin modifiers **(E)**. **(F)** Significantly enriched gene sets among genes upregulated in the C4 subtype (red bars) and the C3 subtype (blue bars). **(G)** Significantly enriched gene sets among proteins upregulated in the C4 subtype (red bars) and the C3 subtype (blue bars). **(H)** Significantly enriched gene sets of methylated genes with downregulated DNA methylation in the C4 subtype compared to the C3 subtype (red bars) or C4 subtype (blue bars)

Figure 3 continued on next page

Figure 3 continued

with upregulated DNA methylation in the C4 subtype compared to the C3 subtype (blue bars). \*p-value <0.05; \*\*p-value <0.01; \*\*\*p-value <0.001; NS, p-value >0.05.

The online version of this article includes the following figure supplement(s) for figure 3:

**Figure supplement 1.** Biological characterization of the comprehensive characterization of colorectal cancer (CCCRC) subtypes based on multi-omics data.

No SCNA was significantly enriched in C2 and C4 subtypes. The single alteration events could not adequately delineate the CCCRC subtypes, we further computed the fraction of the altered samples per oncogenic pathway in each CCCRC subtype. The C4 subtype had the highest frequency of mutations in the cell cycle, HIPPO, MYC, NOTCH, PI3K, TGFB, and RTK-RAS pathways (all p-value <0.05) (**Figure 3C, Supplementary file 2c**). Notably, the C1 subtype had the highest frequency of mutations in the WNT pathway (p-value = 0.019). The frequency of mutations in the TP53 pathway was not significantly different between CCCRC subtypes. The 10 oncogenic pathways had higher frequencies of amplification (all p-value <0.05), and 9 oncogenic pathways (except the NRF2 pathway) had higher frequencies of deletion (all p-value <0.05) in C1 and C3 subtypes compared with C2 and C4 subtypes. Although none of genomic alterations was limited to or specific to a particular subtype, the apparent enrichment of certain alteration events within the CCCRC subtypes might highlight the TME heterogeneity and the genotype-CCCRC correlations of CRC.

Subsequently, we found that the different CCCRC subtypes displayed highly diverse epigenetic, transcriptional, and proteomic profiles. As expected, the analysis of differentially methylated genes (DMGs) between CRC and normal tissues demonstrated that the C4 subtype had the most DMGs ( $n = 145$ ) compared to the C1 subtype ( $n = 109$ ), C2 subtype ( $n = 12$ ), and C3 subtype ( $n = 23$ ), and the C4 subtype exhibited extensive hypermethylation with the highest frequency of the CpG island methylator phenotype (CIMP) compared with the other subtypes (**Figure 3D**). We further analyzed the regulon activity of critical chromatin modifiers and transcription factors in CRC, which could better evaluate their combinatorial biological effects. The regulon activity of the chromatin modifiers of the C1 subtype was generally higher than that of the other subtypes (**Figure 3E**). The differences in the regulon activity of the chromatin modifiers might indicate that epigenetically driven transcriptional networks contributed to the remodeling of the TME, especially in the C1 subtype. Meanwhile, we observed that each subtype had different transcription factor activities (**Figure 3E**). As shown in **Figure 3—figure supplement 1A**, C1-specific upregulated genes (false discovery rate [FDR]<0.001, top 1000 by log<sub>2</sub>FC) were enriched for the pathways associated with tumor proliferation and metabolism. C2-specific upregulated genes were enriched for the pathways associated with immune function, stroma, and neurons. C3-specific upregulated genes were enriched for the pathways associated with stroma and neurons. Both C2- and C3-specific upregulated genes were enriched in neuron-associated pathways, suggesting that neuronal development might be involved in the formation of ECM. C4-specific upregulated genes were enriched for the pathways associated with anti-tumor immune function. The lists of differentially expressed genes (DEGs) between CCCRC subtypes were compiled in **Supplementary file 2d**. In addition, we also analyzed the correlations between all genes and the exclusion and dysfunction scores obtained from the Tumor Immune Dysfunction and Exclusion (TIDE) website, as well as the MSI expression signature, which provided a useful reference for subsequent research in the discovery of new therapeutic targets (**Supplementary file 2e**). The CCCRC-specific upregulated methylation genes (FDR <0.001, top 1000 by difference) and the CCCRC-specific upregulated proteins (p-value <0.05) were also enriched for analogous biological functional categories (**Figure 3—figure supplement 1B, C**). Gene expression differences among the CCCRC subtypes were validated in the CRC-RNAseq cohort (**Figure 3—figure supplement 1D–G**). DMGs (FDR <0.001, top 1000 by log<sub>2</sub>FC), DEGs (FDR <0.001, top 1000 by log<sub>2</sub>FC), and differentially expressed proteins (DEPs) (p-value <0.05) between each subtype were enriched for similar biological functional categories. Indeed, DEGs and DEPs upregulated in the C4 subtype compared with the C3 subtype were significantly enriched for immune-related pathways, whereas DEGs and DEPs upregulated in the C3 subtype compared with the C4 subtype were highly enriched for TGF-beta signaling, EMT, and angiogenesis (**Figure 3F, G**). Similarly, genes with decreased DNA methylation in the C3 subtype compared with the C4 subtype were enriched for EMT and ECM regulation, whereas genes with decreased DNA methylation in the C4 subtype were significantly enriched for immune-related pathways (**Figure 3H**).

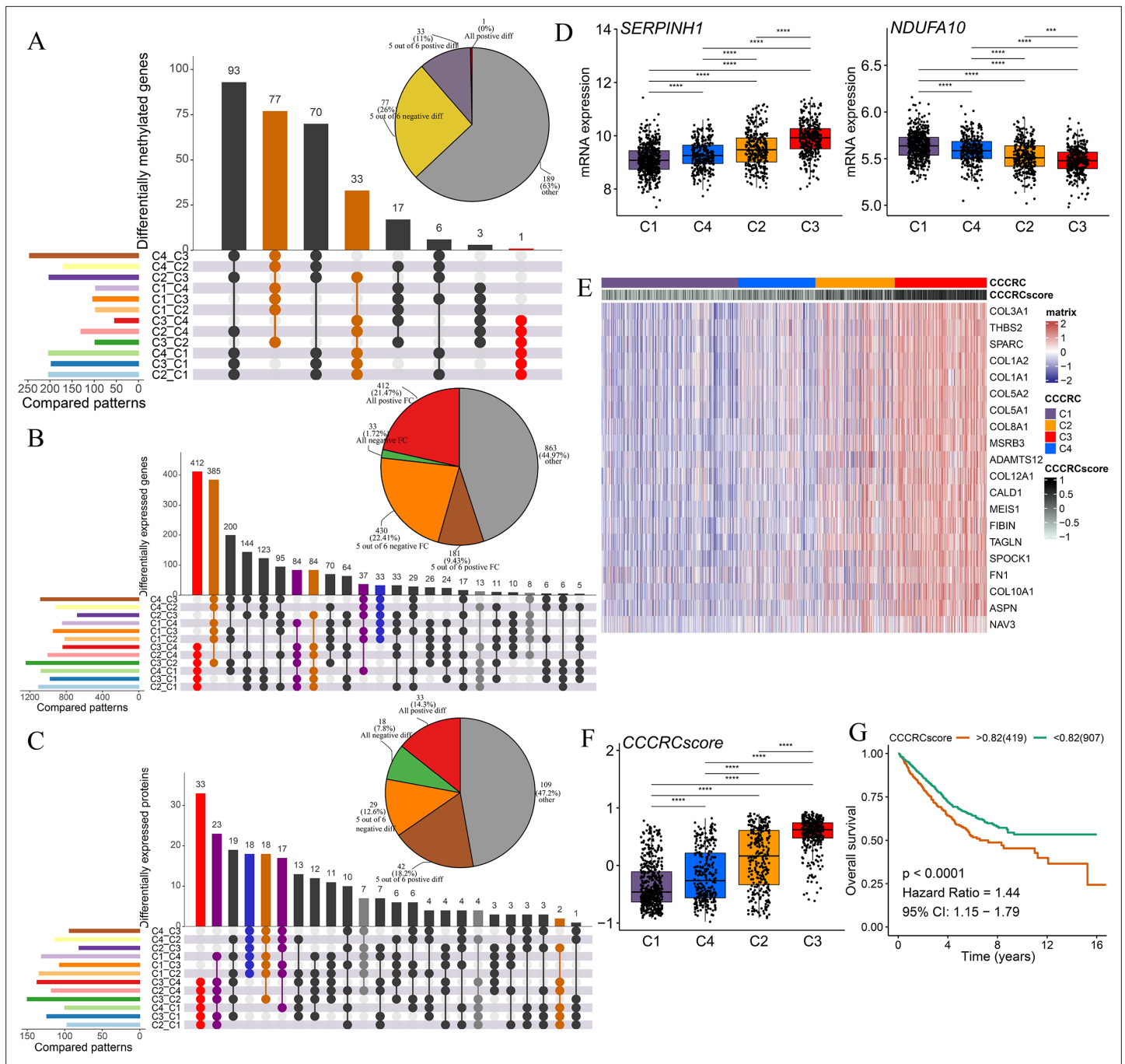
Collectively, the similar differential biological patterns of DNA methylation, gene expression, and protein levels among the CCCRC subtypes highlighted their role in influencing the TME of CRC.

## Discovery of a nongenetic tumor evolution pattern

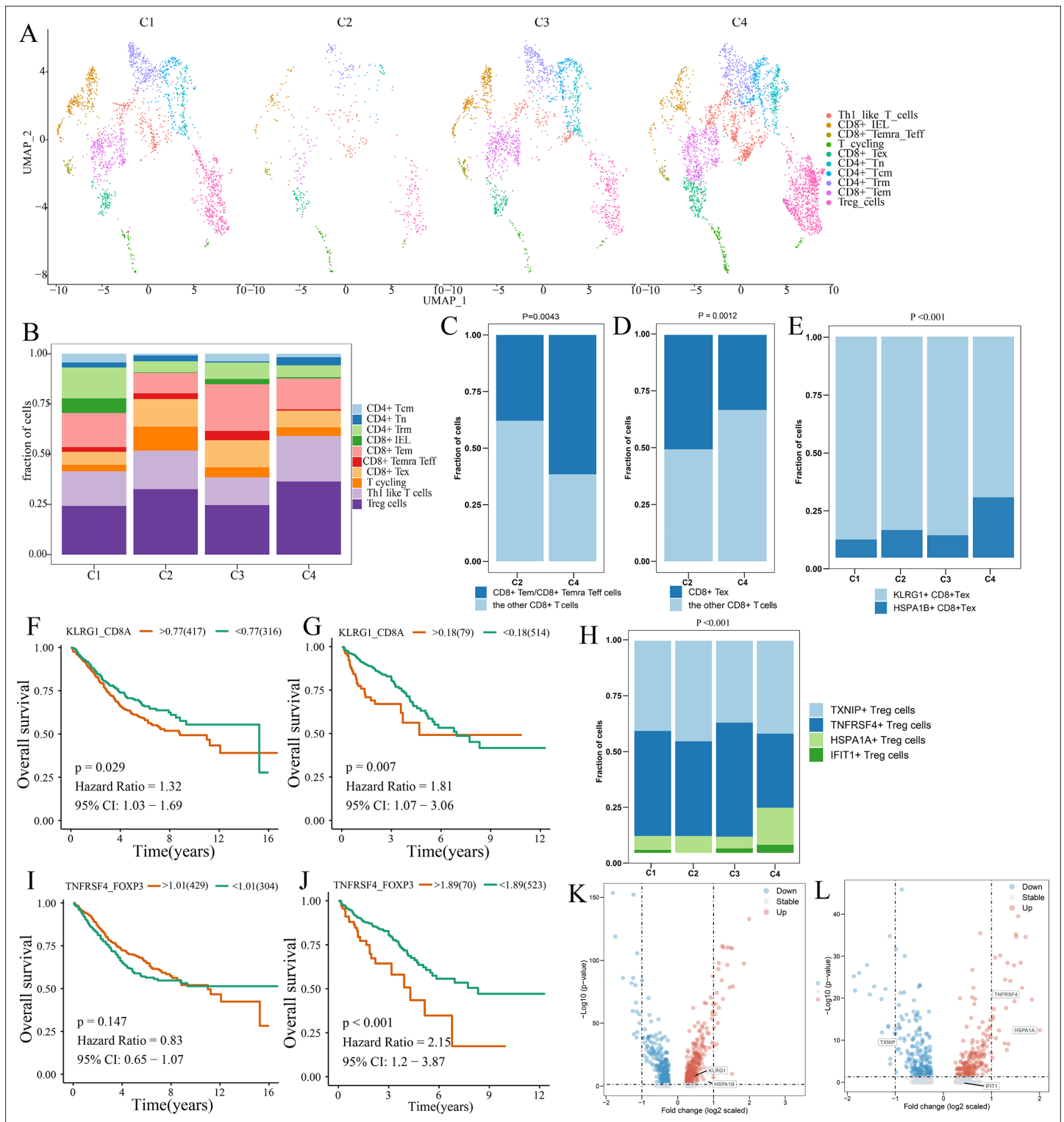
Our study sought to investigate whether there is a dominant evolutionary pattern among the four CCCRC subtypes. According to the theory of nongenetic tumor evolution, a dominant evolutionary pattern would involve gradual changes in certain molecular features across tumor subtypes, either increasing or decreasing (Tavernari *et al.*, 2021). To begin with, we obtained the common set of all DMGs (FDR <0.05), DEGs (FDR <0.05), and DEPs ( $p$ -value <0.5) between each subtype, including both upregulated and downregulated ones, by taking their intersection. Next, we separately intersected the sets of upregulated and downregulated DMGs, DEGs, and DEPs among each subtype in the common set to explore potential evolutionary patterns. From the UpSet plot (Lex *et al.*, 2014), it was evident that the evolutionary pattern from C1 to C4, followed by C2, and C3 subtypes had the same sign in difference or FC and were dominant: either all positive for increasing DNA methylation/gene expression/protein levels (red bar), or all negative for decreasing DNA methylation/gene expression/protein levels (blue bar) (Figure 4A–C). For instance, the top DEGs SERPINH1 and NDUFA10 showed a gradual increase and decrease in expression from C1, C4, C2 to C3 subtypes, respectively (Figure 4D). Meanwhile, we found the possibility of interconversion between C2 and C3 subtypes (yellow bar), between C1 and C4 subtypes (purple bar), and between C2 and C4 subtypes (gray bar) in this evolutionary pattern. Furthermore, we intersected all the positives for increasing gene expression from C1 to C4, C2, and C3 subtypes in the CRC-AFFY and CRC-RNAseq cohorts and obtained 20 CCCRC genes (Figure 4E, Supplementary file 2f), which were associated with TGF-beta signaling and infiltrating nerves. High expression of all 20 genes was significantly associated with poor DFS prognosis. To quantify the evolutionary pattern of individual CRC patients, we performed GSVA to generate CCCRC scores. To better evaluate the molecular features of the CCCRC scores, we also analyzed the correlation between the CCCRC scores and the TME-related signatures. As expected, the CCCRC scores were strongly associated with the immunosuppressive signatures, including M2 macrophages, EMT, and angiogenesis (Figure 3—figure supplement 1H). The CCCRC score was the highest in the C3 subtype than in the other subtypes (Figure 4F), and the high CCCRC score was significantly associated with shorter OS (Figure 4G). Overall, our analysis implied that the four CCCRC subtypes not only had their own unique biological characteristics, but also had a dominant evolutionary pattern driven by epigenetic, transcriptional, and proteomic reprogramming.

## Differences in T cell function between CCCRC subtypes

We obtained the GEP for 7766T cells from 12 patients with CRC, including four patients with the C1 subtype, one patient with the C2 subtype, two patients with the C3 subtype, and four patients with the C4 subtype (Supplementary file 2g). A total of five CD4<sup>+</sup> and four CD8<sup>+</sup> T cell clusters were identified in tumor and normal tissues, including CD8<sup>+</sup> intraepithelial lymphocytes (CD8<sup>+</sup> IELs), effector memory CD8<sup>+</sup> T cells (CD8<sup>+</sup> Tem), recently activated effector memory or effector CD8<sup>+</sup> T cells (CD8<sup>+</sup> Temra/Teff), exhausted CD8<sup>+</sup> T cells (CD8<sup>+</sup> Tex), central memory CD4<sup>+</sup> T cells (CD4<sup>+</sup> Tcm) and naive CD4<sup>+</sup> T cells (CD4<sup>+</sup> Tn), tissue-resident memory CD4<sup>+</sup> T (CD4<sup>+</sup> Trm) cells, TH1-like cells, Treg cells, and T cycling cells (Figure 5—figure supplement 1A, B). The markers of the T cell clusters were summarized in Supplementary file 2h. Figure 5A and B shows the distribution of the 10T cell clusters among each CCCRC subtype. The bulk RNAseq analyses demonstrated that C2 and C4 subtypes showed relative upregulation of immune components. Notably, we found that the C4 subtype was enriched in CD8<sup>+</sup> Tem and CD8<sup>+</sup> Temra/Teff cells, but lacked CD8<sup>+</sup> Tex cells compared with the C2 subtype (Figure 5C, D). Within the subset of CD8<sup>+</sup> Tex cells, we distinguished two smaller subsets according to their gene expression markers, KLRG1<sup>+</sup> CD8<sup>+</sup> Tex cells and HSPA1B<sup>+</sup> CD8<sup>+</sup> Tex cells (Figure 5—figure supplement 1C, D). KLRG1<sup>+</sup> CD8<sup>+</sup> Tex cells were more enriched in C1, C2, and C3 subtypes than the C4 subtype (Figure 5E), which resemble terminally exhausted T cells, and they were associated with non-response to ICB therapy (Luoma *et al.*, 2022). Moreover, the higher ratio of KLRG1-to-CD8A expression, the worse the OS of patients in CRC-AFFY and CRC-RNAseq cohorts (Figure 5F, G). Meanwhile, we re-clustered the Treg cells and identified four Treg cell subsets, namely, TXNIP<sup>+</sup> Treg cells, TNFRSF4<sup>+</sup> Treg cells, HSPA1A<sup>+</sup> Treg cells, and IFIT1<sup>+</sup> Treg cells (Figure 5—figure supplement 1E–H). We found that TNFRSF4<sup>+</sup> Treg cells were significantly



**Figure 4.** Discovery of a nongenetic tumor evolution pattern. The UpSet plots are used to visualize the intersection of differentially methylated genes (A), all differentially expressed genes (B), and all differentially expressed proteins (C) between each comprehensive characterization of colorectal cancer (CCCRC) subtype. As an example, ‘C4\_C3’ means that the molecular features of the C4 subtype were upregulated compared with the C3 subtype. Pie chart (top right) distributions of the sign of differences or fold changes computed for all differentially methylated genes (A), differentially expressed genes (B), and differentially expressed proteins (C). (D) mRNA expression of two top differentially expressed genes (y-axis) in the CRC-AFFY cohort stratified by the CCCRC subtypes (x-axis). (E) Heatmap shows gene expression levels of 20 CCCRC genes among the four CCCRC subtypes. (F) Box plots show differences in the CCCRC score among the four CCCRC subtypes in the CRC-AFFY cohort. (G) Kaplan–Meier method of overall survival (OS) among the four CCCRC subtypes in the CRC-AFFY and CRC-RNAseq cohorts. FC: fold change; diff: difference. \*\*\*p-value < 0.001; \*\*\*\*p-value < 0.0001.



**Figure 5.** Differences in T cell function between comprehensive characterization of colorectal cancer (CCCRC) subtypes. **(A)** UMAP shows the composition of T cells colored by cluster and divided by the CCCRC subtype in colorectal cancer (CRC) tissues. **(B)** Histogram shows the cell distribution of 10 T cell types in the different CCCRC subtypes. **(C)** Proportion of effector memory CD8+ T cells (CD8+ Tem), recently activated effector memory or effector CD8+ T cells (CD8+ Temra/Teff), and the other CD8+ T cells (shown in the histogram) in the C2 and C4 subtypes. **(D)** Proportion of exhausted CD8+ T cells (CD8+ Tex) and the other CD8+ T cells (shown in the histogram) in the C2 and C4 subtypes. **(E)** Histogram shows the cell distribution of KLRG1+ CD8+ Tex and HSPA1B+ CD8+ Tex cells in the different CCCRC subtypes. Kaplan–Meier method with log-rank test of overall survival (OS) in the CRC-AFFY cohort **(F)** and the CRC-RNAseq cohort **(G)** between low and high ratios of KLRG1-to-CD8A expression in patients. **(H)** Histogram shows

Figure 5 continued on next page

Figure 5 continued

the cell distribution of TXNIP+ Treg cells, TNFRSF4+ Treg cells, HSPA1A+ Treg cells, and IFIT1+ Treg cells in the different CCCRC subtypes. Kaplan–Meier method with log-rank test of OS in the CRC-AFFY cohort (I) and the CRC-RNAseq cohort (J) between low and high ratios of TNFRSF4-to-CD8A expression in patients. (K) Volcano plot shows differentially expressed genes between tumor (red dots) and normal CD8+ T cells (blue dots). (L) Volcano plot shows differentially expressed genes between tumor (red dots) and normal Treg cells (blue dots).

The online version of this article includes the following figure supplement(s) for figure 5:

**Figure supplement 1.** Differences in T cell function between comprehensive characterization of colorectal cancer (CCCRC) subtypes.

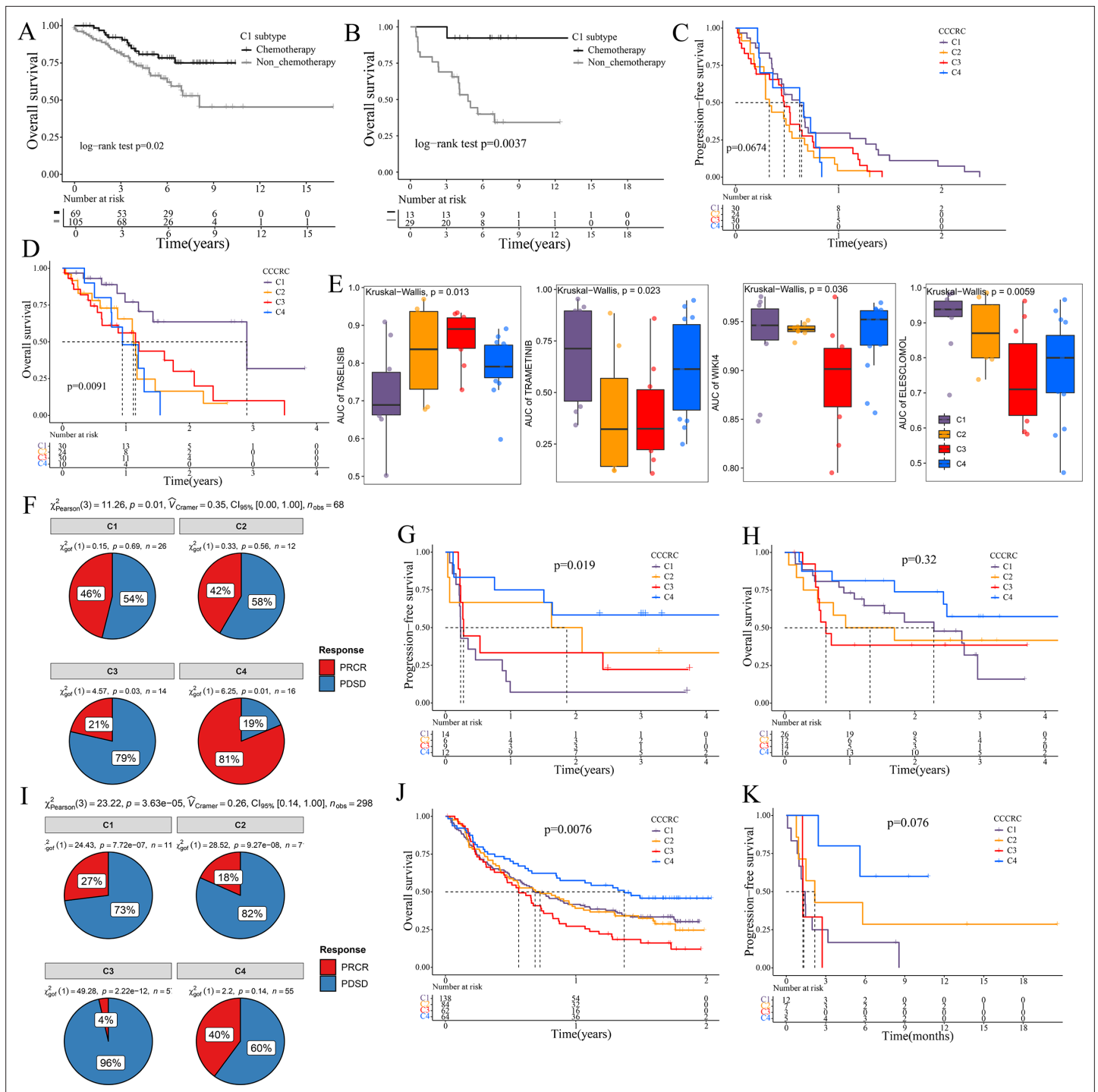
**Figure supplement 2.** Survival analyses of T cell subsets.

more enriched in C2 and C3 subtypes than the C4 subtype (**Figure 5H**), which might indicate that TNFRSF4+ Treg cells were closely related to the formation of the tumor stroma. The higher ratio of TNFRSF4-to-FOXP3 expression, the worse the OS of patients in CRC-AFFY and CRC-RNAseq cohorts (**Figure 5I, J**). Equally important, patients with a high ratio of KLRG1-to-CD8A expression or a high ratio of TNFRSF4-to-FOXP3 expression who received anti-PD1/PDL1 therapy had a shorter OS and progression-free survival (PFS) than those with a low ratio of KLRG1-to-CD8A expression or a low ratio of TNFRSF4-to-FOXP3 expression in Gide, Hugo, Jung, and IMvigor210 datasets (**Figure 5—figure supplement 2A–H**). We also found that the expression of KLRG1 and TNFRSF4 was higher in CD8+ T cells and Treg cells, respectively, in tumor tissues than in adjacent tissues (**Figure 5K, L**). Overall, we used scRNAseq data to analyze the differences in T cell function among the different CCCRC subtypes, and the C2 subtype did show more immunosuppression than the C4 subtype, which was consistent with the bulk RNAseq analyses.

### Significance of CCCRC in guiding clinical treatment of CRC

The 5-fluorouracil (5-FU)-based chemotherapy, anti-VEGF (vascular endothelial growth factor) (bevacizumab), and anti-EGFR (epidermal growth factor receptor) (cetuximab) therapies are the first-line treatment options for CRC (**Andrei et al., 2022**). We further explored whether the different CCCRC subtypes could predict therapeutic efficacy. In the CRC-AFFY cohort, 564 patients with stage II and III CRC had chemotherapy-related clinical information, including 323 who were not treated by chemotherapy and 241 who were treated by chemotherapy. Furthermore, 155 stage II and III CRC patients with or without chemotherapy in the GSE103479 dataset were also included in our study. We found that C1 patients with stage II and III CRC receiving chemotherapy had a better OS than those who did not and were more suitable for 5-FU-based chemotherapy in the CRC-AFFY cohort and the GSE103479 dataset (**Figure 6A, B**). Furthermore, 79 mCRC patients in the GSE104645 dataset were treated with chemotherapy and 83 were treated with a combination of chemotherapy and bevacizumab. The response rate (RR) after chemotherapy (including partial response and complete response) of C1 and C4 subtypes tended to be higher than that of C2 and C3 subtypes (**Figure 6—figure supplement 1A**), whereas the RR of the C2 subtype treated with a combination of chemotherapy and bevacizumab tended to be higher than that of the other subtypes (**Figure 6—figure supplement 1B, C**). In addition, the RR tended to be higher in the C2 subtype treated with 5-FU-based chemotherapy and bevacizumab than in those treated with chemotherapy alone (**Figure 6—figure supplement 1D**). The GSE104645 dataset also contained 94 mCRC patients without the RAS mutation who were treated with anti-EGFR antibody. The disease control (DC) rates (DCR) after anti-EGFR therapy (including partial response, complete response, and stable disease) were 87% for C1, 64% for C2, 68% for C3, and 80% for C4, respectively (p-value = 0.23) (**Figure 6—figure supplement 1E**). The DC rate of the C1 subtype with anti-EGFR therapy tended to be higher than that of the other subtypes (p-value = 0.07) (**Figure 6—figure supplement 1F**). Notably, PFS of the C1 subtype with anti-EGFR therapy tended to be better than that of the other subtypes (log-rank p-value = 0.067) and OS of the C1 subtype was significantly better than that of the other subtypes (log-rank p-value = 0.0091) (**Figure 6C, D**). The above results suggested that the C1 subtype may benefit from chemotherapy and anti-EGFR treatment, whereas the C2 subtype may benefit from a combination of 5-FU-based chemotherapy and bevacizumab, but there was no evidence that the C3 subtype would benefit from these treatments.

To further explore the treatment strategies of the CCCRC subtypes, we trained a pre-clinical model based on a filtered gene set comprised of 335 CCCRC subtype-specific and cancer cell-intrinsic gene



**Figure 6.** Significance of comprehensive characterization of colorectal cancer (CCCRC) in guiding the clinical treatment of colorectal cancer (CRC). Kaplan–Meier method of overall survival (OS) between stage II and III CRC C1 patients with or without chemotherapy in the CRC-AFFY cohort (A) and the GSE103479 (B) dataset. Kaplan–Meier method of progression-free survival (PFS) (C) and OS (D) among the four CCCRC subtypes in the GSE104645 dataset. (E) Box plots show the differences in the area under the receiver operator characteristics curve (AUC) of drug responses among the four CCCRC subtypes. (F) Pie chart shows the differences in the proportion of responses to immune checkpoint blockade treatment among the four CCCRC subtypes in the two independent melanoma cohorts (Gide and Hugo datasets,  $n = 68$ ) treated with anti-PD1 therapy. Kaplan–Meier method with log-rank test of PFS (G) and OS (H) among the four CCCRC subtypes in the two independent melanoma cohorts (Gide and Hugo datasets,  $n = 68$ ) treated with anti-PD1 therapy. (I) Pie chart shows the differences in the proportion of responses to immune checkpoint blockade treatment among the four CCCRC subtypes in the urothelial carcinoma cohort ( $n = 298$ ) treated with anti-PDL1 therapy. Kaplan–Meier method with log-rank test of OS and PFS

Figure 6 continued on next page



Figure 6 continued

among the four CCCRC subtypes in the urothelial carcinoma cohort ( $n = 348$ ) (J) and the lung cancer cohort ( $n = 27$ ) (K) treated with anti-PD1/PDL1 therapy. PRCR: partial response and complete response; PDS: progressive disease and stable disease.

The online version of this article includes the following figure supplement(s) for figure 6:

**Figure supplement 1.** Significance of comprehensive characterization of colorectal cancer (CCRC) in guiding clinical treatment of colorectal cancer (CRC).

**Figure supplement 2.** Spatial transcriptomics analysis of the Cytassist sample.

**Figure supplement 3.** Spatial transcriptomics analysis of the Visium sample.

**Figure supplement 4.** Establishment of the comprehensive characterization of colorectal cancer (CCRC) classifier.

markers (**Supplementary file 3a**). The pre-clinical model was publicly available at [https://github.com/XiangkunWu/pre\\_clinical\\_model](https://github.com/XiangkunWu/pre_clinical_model) (copy archived at **Wu, 2023b**). The 71 human CRC cell lines were classified into four CCCRC subtypes (**Supplementary file 3b**). As expected, the C4 subtype cell lines demonstrated the highest TMB values and MSI frequency while exhibiting the lowest FGA scores when compared to other subtypes (**Figure 6—figure supplement 1G–I**). In contrast, C1 and C3 subtype cell lines showed significantly higher FGA scores and significantly lower TMB values and MSI frequency. The C2 subtype cell lines had median FGA scores, TMB values, and MSI frequency. Our analysis revealed that the AUCs of several pathway inhibitors differed significantly among CCCRC subtypes (**Figure 6E, Figure 6—figure supplement 1J**). A lower AUC value on the dose–response curve of a drug indicates an increased sensitivity of this drug to tumor treatment. Specifically, the AUCs of taselisib and GSK2110183, the PI3K pathway inhibitor, were significantly lower in the C1 subtype. The AUCs of trametinib, a MEK 1/2 inhibitor, were significantly lower in the C2 and C3 subtypes. Using GSEA in the CRC-AFFY cohort, we also found that the KEGG MAPK signaling pathway was upregulated in C2 and C3 subtypes than the other subtypes (**Figure 6—figure supplement 1K, L**). The AUCs of WIK14, an inhibitor of the WNT pathway, elesclomol, an inducer of oxidative stress and cuproptosis, (5Z)-7-oxozeaenol, a TAK1 inhibitor, and WEHI-539, a BCL-XL inhibitor, were significantly lower in the C3 subtype. In contrast, the AUC of 5-FU was significantly higher in the C3 subtype.

ICB therapy has recently emerged as a highly promising therapeutic strategy for various malignancies, but it lacks effective markers to identify suitable patients (**Jin and Sinicrope, 2022**). We collected multiple ICB therapy-associated datasets to evaluate whether the CCCRC classification system could be used as a tool to predict ICB therapy efficacy. In the Gide and Hugo datasets treated with anti-PD1 therapy, patients were classified into the four CCCRC subtypes. As expected, the RR of anti-PD1 treatment was 81% for the C4 subtype compared to 21% for the C3 subtype (**Figure 6F**), and PFS and OS were prolonged for the C4 subtype (**Figure 6G, H**). Similar findings were observed in the IMvigor210 and Jung datasets treated with anti-PD1/PDL1. RR was significantly higher in patients with the C4 subtype (40%) compared with the other subtypes (C1 with 17%, C2 with 18%, C3 with 4%) in the IMvigor210 dataset (**Figure 6I**). The C4 subtype in the IMvigor210 and Jung datasets had the best prognosis, while patients with the C3 subtype had the worst prognosis (**Figure 6J, K**). In conclusion, we have used extensive data analysis to develop treatment protocols for each subtype, but this study is pre-clinical and still requires extensive experimentation and clinical trials for validation.

## Spatial transcriptomics analysis

We conducted a re-analysis of two CRC spatial transcriptomics (ST) data to explore the spatial distribution relationship between four CCCRC subtypes of tumor cells, T cells, and mesenchymal cells (Materials and methods). The Cytassist and Visium samples had a total of 9080 and 2660 spots, respectively. We used 'ssGSEA' method to quantify the six cell subpopulations of each spot and also visualized only the spots corresponding to the top 25% of the score ranking for each cell type (**Figure 6—figure supplement 2A, B, Figure 6—figure supplement 3A, B**). In Cytassist samples, we observed different spatial distribution patterns of the four subtypes of tumor cells (**Figure 6—figure supplement 2B**). Specifically, the C3 subtype of tumor cells was predominantly located in the tumor periphery with an enrichment of mesenchymal cells and T cells (areas selected by black dashed circles). In contrast, the C4 subtype of tumor cells was mainly present in the center of the tumor, accompanied by the presence of T cells. The C1 and C2 subtypes of tumor cells were distributed in relatively uniform areas, mainly in the tumor periphery, with fewer mesenchymal cells and T cells. However, the distribution areas of

C2 and C3 subtypes of tumor cells also partially were in overlap (the area selected by red dashed circles). The same distribution patterns can also be observed in the Visium sample (**Figure 6—figure supplement 3B**). Further analysis of the correlation between the ssGSEA scores of each cell type in the cell-type-rich regions and those of other cell types was conducted (**Figure 6—figure supplement 2D, E, Figure 6—figure supplement 3D, E**). We found that in the C3 subtype-rich region of tumor cells, the C3 subtype score of tumor cells was significantly positively correlated with the mesenchymal cell score, while in the T cell-rich region, the C3 subtype score of tumor cells was significantly negatively correlated with the T cell score. The C4 subtype score of tumor cells was significantly positively correlated with the T cell score and negatively correlated with the mesenchymal cell score in the C4 subtype-rich, T cell-rich, and mesenchymal cell-rich regions. The C1 and C2 subtype scores of tumor cells were negatively correlated with mesenchymal cell and T cell scores. Overall, these results were generally consistent with previous histopathologic analysis findings.

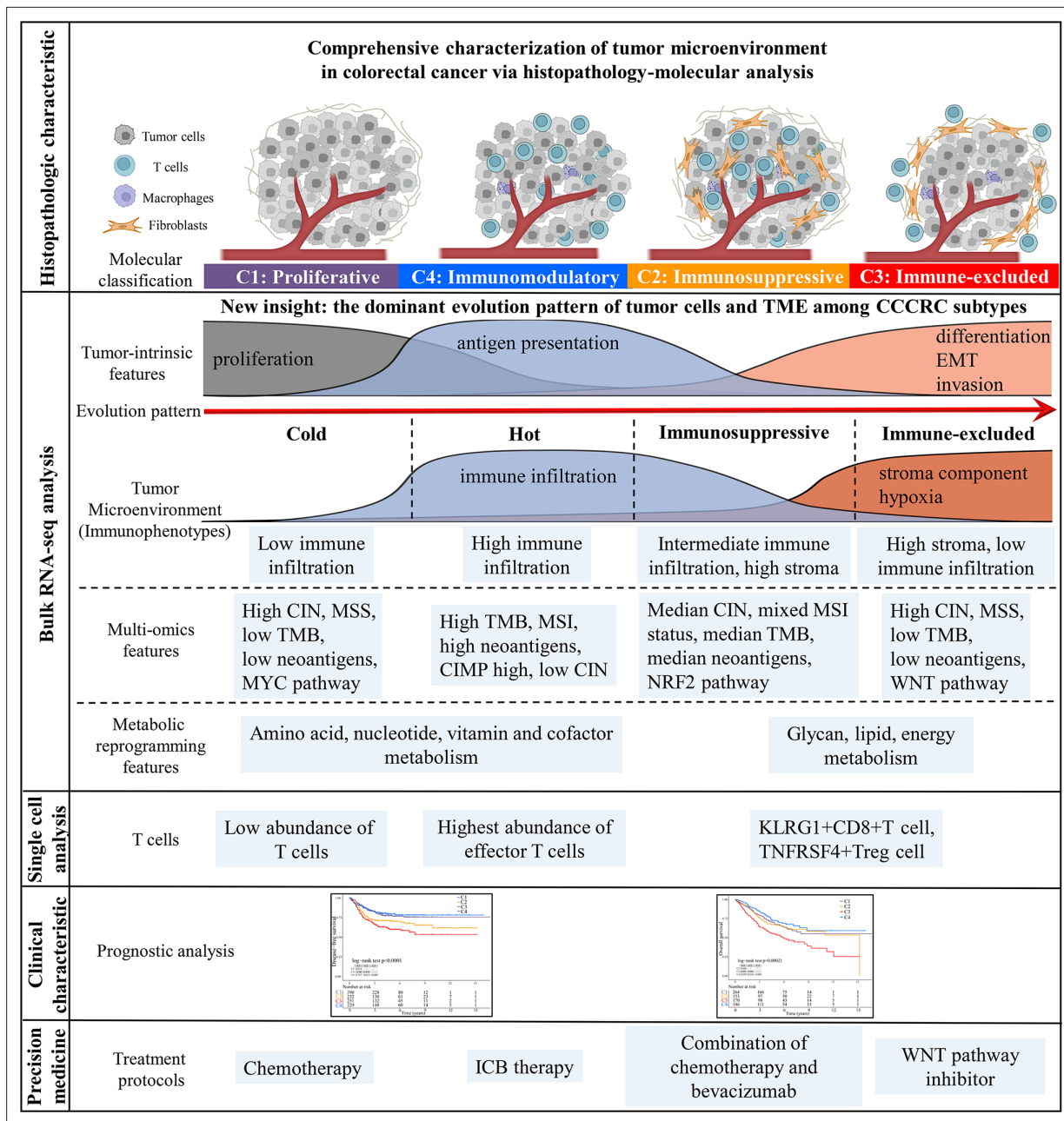
## Construction and validation of CCCRC classifier

To promote the widespread application of the CCCRC classification system, we developed a simple gene classifier to predict CCCRC subtypes. The CRC-AFFY cohort was randomly divided into the training and validation sets at a ratio of 7:3. We established the CCCRC classifier on the training set by utilizing multiple machine learning algorithms based on the GEP of 80 upregulated subtype-specific genes (Materials and methods, **Supplementary file 3c**). Upon application to the test set, GSE14333, and GSE17536 datasets, the performance of the eXtreme Gradient Boosting (xgboost) algorithm was the best with the highest accuracy values and *F1* scores compared to the random forest (RF), support vector machine (SVM), and logistic regression algorithms (**Figure 6—figure supplement 4**). Notably, the CCCRC classifier based on the xgboost algorithm displayed robust performance across gene expression platforms, Affymetrix and RNA-sequencing platforms, exhibiting a balanced accuracy of >80% for all subtypes (**Supplementary file 3d**). These findings demonstrated the stability and cross-platform applicability of our classifier. The CCCRC classifier based on the xgboost algorithm is publicly available at [https://github.com/XiangkunWu/CCRC\\_classifier](https://github.com/XiangkunWu/CCRC_classifier) (copy archived at **Wu, 2023c**) and the CCCRC subtype information of CRC patients can be obtained by directly inputting the GEP of 80 upregulated subtype-specific genes. The CCCRC classifier might facilitate the discovery of new biomarkers and the personalized treatment of clinical patients with CRC.

## Discussion

The key role of the TME in dynamically regulating tumor progression and affecting treatment outcomes has been widely recognized, and treatment strategies targeting the TME have become a promising approach for cancer therapy (**Bejarano et al., 2021; Binnewies et al., 2018; Pandey et al., 2022; Tang et al., 2021**). However, there are few comprehensive analyses that consider the tumor cells and the TME as a whole. The comprehensive dissection of the crosstalk between tumor cells and TME may reveal new tumor biology concepts and identify therapeutic targets, and ultimately achieve precise medical treatment (**Bejarano et al., 2021; Jia et al., 2022**). Thus, we collected the molecular features of the tumor cells and TME to reconstruct the whole tumor composition and performed integrated analyses to understand the TME. The four CCCRC subtypes had distinct molecular and histopathologic characteristics, therapeutic efficacy, and prognosis (**Figure 7**). We identified a nongenetic evolutionary pattern from C1, C4, C2, and C3 was associated with an evolution from a cold (C1) to a hot (C4) and eventually suppressive (C2) and excluded (C3) microenvironment (**Figure 7**).

In this study, we identified four subtypes with distinct TME features through unsupervised clustering analysis of approximately 2000 CRC patients. C1 and C4 subtypes are typical desert and inflamed tumors, respectively, while C2 and C3 subtypes were difficult to classify into one of the classical immunophenotypes of the three-category immune classification system ('desert', 'excluded', and 'inflamed' phenotypes) (**Chen and Mellman, 2017; Rosenberg et al., 2016**) based on TME features due to the unclear distribution of stromal components and lymphocytes. Our pathologists evaluated the histopathologic characteristics for each subtype under the microscope and observed that the C2 subtype was mainly categorized as an excluded phenotype and the C3 subtype was mainly classified as a desert and an excluded phenotype. However, the WSIs showed that lymphocytes in the C2 subtype were more frequently intermixed with the stroma within but not adjacent to the main



**Figure 7.** Overview of characteristics of comprehensive characterization of colorectal cancer (CCRC) subtypes. The graphical abstract summarizes histopathologic characteristics, tumor microenvironment features, multi-omics features, scRNAseq features, treatment strategies, and prognostic value for CCCRC subtypes.

tumor mass, and lymphocytes in the C3 subtype were more frequently excluded from the tumor mass but not intermixed with lymphocytes within the main tumor mass, both of which were classified as the excluded phenotype. Notably, we used AI-enabled spatial analysis of WSIs to confirm the semi-quantitative results of the pathologist, that is, the C2 subtype had increased lymphocyte and stromal infiltration in CT and IM regions and the C3 subtype had the highest abundance of stroma and less lymphocyte infiltration in the CT region, while lymphocyte infiltration and stromal components were observed in the IM region, which were more visually confirmed in the results of ST data analysis. We also found that the C2 subtype had the highest T cell dysfunction score and the C3 subtype had the highest T cell exclusion score. GSEA demonstrated that the terminally exhausted CD8+ T cell signature was upregulated in the C2 subtype compared with the C4 subtype. scRNA-seq analysis

showed that KLRG1+ CD8+ T cells were significantly more enriched in C2 and C3 subtypes than the C4 subtype. KLRG1+ CD8+ T cells were associated with nonresponse to ICB therapy, which were more terminally differentiated than KLRG1− CD8+ T cells and had lower proliferative capacity (Luoma et al., 2022). KLRG1 is a marker of terminal differentiation of CD8+ T cells (Luoma et al., 2022), and the inhibitory receptor of ILC1s (group 1 innate lymphoid cells), ILC2s, and NK cells (Chiosso et al., 2018). ILC1s in tumors express high levels of the KLRG1 gene and pro-angiogenic activity and may even promote tumor progression in TGF-beta-rich tumors (Chiosso et al., 2018). Therefore, we defined C2 and C3 subtypes as immunosuppressed and immune-excluded, respectively. Our CCCRC classification system refined the three-category immune classification system (Chen and Mellman, 2017; Rosenberg et al., 2016). Moreover, we defined for the first time the four-category immune classification system based on the integrative multi-omics analysis and histopathologic characteristics ('hot', 'immunosuppressed', 'excluded', and 'cold' phenotypes) (Galon and Bruni, 2019; Kirchner et al., 2022).

We observed an intriguing nongenetic evolution pattern among the CCCRC subtypes, wherein a shift occurred from the C1 (proliferative subtype) to C4 (immunomodulatory subtype) subtype, followed by the C2 (immunosuppressed subtype) subtype and finally, the C3 (immune-excluded subtype) subtype. We proposed that in this evolutionary pattern, immune infiltration gradually increased with the augmentation of genomic alterations and tumor immunogenicity, while the stromal components also increased in a gradual manner. These components are crucial in the progression of tumor, as they contribute to the exhaustion and eventual exclusion of lymphocytes from the tumor bed (Calon et al., 2015; Galon and Bruni, 2019; Kirchner et al., 2022). Meanwhile, our research revealed that both C2 and C3 subtypes exhibited a high level of tumor stroma, while C1 and C4 subtypes were characterized by active DNA damage and repair and high tumor proliferation. Additionally, C2 and C4 subtypes had an abundance of immune components. This was consistent with our finding that there may be interconversion between the C1 and C4 subtypes, between the C4 and C2 subtypes, and between the C2 and C3 subtypes in this evolutionary pattern. The interconversion between C2 and C4 subtypes in this evolutionary pattern was the rarest situation, indicating that once the tumor enters the C2 subtype, it was difficult to reverse and will progress to the C3 subtype. Not incidentally, this evolutionary pattern was consistent with the findings of Tavernari et al., 2021, who demonstrated that the progression from lepidic to solid histology of lung adenocarcinoma was associated with a transition from a cold (lepidic) to a hot (papillary and especially acinar), and eventually suppressive and excluded (solid) TME. Using spatially resolved molecular profiles, they uncovered the coexistence of cold, inflamed, and excluded regions within individual tumors. Similarly, our ST analysis revealed the simultaneous presence of multiple CCCRC subtypes in the same sample. What is more, we identified a gene list that promoted this evolutionary pattern and proposed CCCRC score based on the gene list to quantify the evolutionary pattern of individual CRC patients, which served as an independent prognostic predictor. However, it is important to note that further experimental evidence is required to verify this bold speculation of the evolutionary pattern, and a large collective effort is needed to arrive at a consensus.

The CCCRC subtypes were significantly correlated with the CMS subtypes (Guinney et al., 2015) and clinicopathological characteristics. The CMS classification system integrated six independent classification systems utilizing a network-based approach (Guinney et al., 2015), which is considered as the most robust classification system that is used to predict prognoses and to guide ICB therapy, chemotherapy, and anti-EGFR therapy as well as to screen new potential targeted drugs (Eide et al., 2017; Kawazoe et al., 2020; Linnekamp et al., 2018; Sveen et al., 2018; Thota et al., 2021; Zhan et al., 2021). In our study, the transcriptomic data used to identify CCCRC subtypes were essentially the same as the cohort used to develop the CMS subtypes (Guinney et al., 2015). The C1, C2/C3, and C4 subtypes exhibited partial overlap with the CMS2, CMS4, and CMS1 subtypes, respectively. Notably, we observed a higher prevalence of right-sided lesions in the C4 subtype, consistent with the CMS1 subtype, and a higher prevalence of left-sided lesions in the C1/C3 subtype, consistent with the CMS4 subtype. Previous studies indicated that left-sided CRCs had greater genetic instability and right-sided CRCs were associated with MSI and CIMP-high features (Bufill, 1990; Lee et al., 2017). We observed that patients with the CMS1 subtype, characterized by high immune infiltration and activation, did not have the best prognosis compared with the other CMS subtypes, while patients with the CMS2 subtype, characterized by low immune infiltration, had the best prognosis (Becht et al., 2016a;

**Guinney et al., 2015**). This might be due to the fact that CMS1 subtype had a higher accumulation of fibroblasts compared to CMS2 subtype (**Guinney et al., 2015**). Moreover, our study revealed that, in comparison with the C4 subtype, the CMS1 subtype had fewer anti-tumor immune components, more stroma components, and other immunosuppressive components, resulting in a poorer prognosis. The prognosis of patients with the C2 subtype was also better than that of patients with the CMS4 and C3 subtypes, and the C2 subtype within the CMS4 subtype also had a better prognosis than the C3 subtype within the CMS4 subtype, which might be related to the presence of more immune infiltration in the C2 subtype than in the C3 and CMS4 subtypes. Our CCCRC classification system further subdivided CMS4 into two subtypes, the C2 and C3 subtypes, each with distinct molecular and prognostic characteristics. Meanwhile, the C4 subtype with MSI exhibited higher immune infiltration and less stromal component compared with the CMS4 subtype with MSI. Overall, our CCCRC classification system can be combined with the CMS classification system and other molecular subtypes to enhance the understanding of the association between TME heterogeneity and different clinical phenotypes.

The CCCRC classification system hold promise for discovering novel biomarkers and therapeutic targets in the TME based on one or more immune evasion mechanisms, which can be further validated by combining TIDE score (**Jiang et al., 2018**), CMS classification system (**Guinney et al., 2015**) and forthcoming molecular classification systems. The C1, C2, and C3 subtypes corresponded to immune evasion mechanisms of low immunogenicity, T cell dysfunction, and T cell exclusion, respectively. In contrast, the C4 subtype exhibited the highest degree of infiltration by anti-tumor immune components and the most favorable prognosis, potentially indicating immunological control of the tumor, known as immune equilibrium. For instance, KDM1A and RAD21 are specific genes for C1 subtype, and it has been reported that interfering with them could activate interferon signaling and induce tumor intrinsic immunogenicity (**Deng et al., 2022; Nguyen et al., 2022; Supplementary file 2d**). Similarly, SPP1, a feature gene for the C3 subtype, has been reported to be positively correlated with colon cancer liver metastasis in SPP1+ macrophages and highly expressed in CMS4 (mesenchymal subtype) (**Lee et al., 2020; Liu et al., 2022**). SPP1+ macrophages interacted with CAFs to induce extracellular matrix remodeling and the formation of an immune-excluded phenotype, thereby limiting immune infiltration into the tumor core (**Liu et al., 2023**). SERPINB9 is significantly upregulated in the C2 subtype, and previous research has demonstrated its pivotal function in enhancing tumor cell survival and promoting the existence of immune-suppressive cells within the TME (**Jiang et al., 2020**). In this study, the expression levels of SPP1 and SERPINB9 were found to be significantly positively correlated with the exclusion and dysfunction scores (TIDE score), respectively (**Supplementary file 2e**). Meanwhile, we identified a large number of mutant genes significantly enriched in the C4 subtype, which mutated to cause substantial immune infiltration and could be candidate genes for mRNA vaccine development. The RNA-mediated immunotherapy regulating the TME is known as the next era of cancer treatment (**Pandey et al., 2022**). Additionally, the CCCRC classification system has the potential to offer a deeper understanding of the underlying biology of CRC by integrating forthcoming molecular classification systems. First, it is necessary to identify the commonalities and differences between different classification systems and how they can complement each other. This can be achieved by conducting comparative studies and analyzing the overlap between different classification systems. Second, it is essential to leverage the strengths of each classification system to identify new biomarkers or therapeutic targets. For example, the CCCRC classification can provide valuable insights into the underlying immune evasion mechanisms of CRC. Third, it is important to combine the discovered potential therapeutic targets with different classification systems for mutual verification. Overall, extensive research has demonstrated a significant correlation between the CCCRC subtype-specific genes and immune evasion mechanisms, underscoring their potential as promising therapeutic targets for developing precision medicine approaches against cancer within the TME. However, realizing the full potential of these classification systems will require a collaborative effort among researchers and clinicians to ensure their effective integration and utilization. Through integration, we can leverage the strengths of different classification systems to identify new biomarkers and therapeutic targets that can ultimately improve patient outcomes.

The CCCRC classification system may contribute to the design of future clinical trials. In our study, we found that the C1 subtype may be suitable for chemotherapy and cetuximab treatment, which is consistent with the predicted results of CMS2 in the CALGB/SWOG 80405I clinical trial (**Lenz et al., 2019**). Additionally, we observed that the C2 subtype could potentially benefit from a combination

of chemotherapy and bevacizumab treatment. By analyzing the molecular characteristics of the C4 subtype and validating it in multiple cohorts receiving ICB therapy, we proposed that ICB therapy may represent a promising therapeutic strategy for the treatment of the C4 subtype. However, our predicted results showed that the C3 subtype was resistant to chemotherapy, cetuximab, bevacizumab, and ICB therapy. Therefore, we used cell line drug sensitivity assay data to explore further treatment options for the CCCRC subtypes. We observed that the C1 subtype exhibited higher PI3K pathway activity compared to the other subtypes. As expected, our drug sensitivity analyses suggested that the C1 subtype may benefit from PI3K inhibitors. Notably, the potential pro-inflammatory effect of PI3K-mTOR pathway inhibitors might offer advantages for cancer immunotherapy (*Fruman et al., 2017*), indicating that the C1 subtype might be suitable for PI3K inhibitors combined with ICB treatment. Similarly, we also found upregulation of the KEGG MAPK signaling pathway in the C2 and C3 subtypes. Drug sensitivity analyses suggested that the C2 and C3 subtypes may benefit from treatment with trametinib, an MEK 1/2 inhibitor. *Borriello et al., 2017* demonstrated that activation of CAFs in primary human neuroblastoma depends on co-activation of JAK2/STAT3 and MEK/ERK1/2 signaling in tumor cells, and trametinib can enhance the response of tumor cells to etoposide. Moreover, *Datta et al., 2022* showed that combined inhibition of MEK and STAT3 can alleviate stromal inflammation, reduce the abundance of iCAF, myCAF, and apCAF, and remodel the TME to overcome immune therapy resistance in pancreatic ductal adenocarcinoma. We also found that the C3 subtype showed increased sensitivity to the WNT pathway inhibitor WIKI4. The C3 subtypes had an upregulation of the WNT signaling pathway, and activation of the WNT signaling pathway can cause T cells in highly immunogenic tumors to be excluded from the tumor core (*Takeuchi et al., 2021*). Overall, our classification system can assist in identifying populations suitable for appropriate treatments in a clinical setting and may promote precision medicine, but this research is pre-clinical and requires further substantial basic experiments and clinical trials to validate.

The CCCRC classification system also can facilitate the understanding of the differences in the results of the CALGB/SWOG 80405 and FIRE-3 clinical trials (*Heinemann et al., 2014; Lenz et al., 2019*). Both trials evaluated the combination of cetuximab or bevacizumab with a different chemotherapy backbone: in the CALGB/SWOG 80405 trial, 75% of patients received oxaliplatin, while in the FIRE-3 trial, all patients received irinotecan. *Aderka et al., 2019* discussed the reasons for the differences in the results of these two clinical trials, which may be related to differences in the chemotherapy backbone used and TME heterogeneity. Based on our examination of the results summarized in **Figure 4** of the work by *Aderka et al., 2019*, we found that differences in the treatment outcomes of the CMS1 and CMS4 subtypes were the crucial factor behind the divergent results observed in the two clinical trials. The CMS1 and CMS4 subtypes had a microenvironment rich in CAFs. Our CCCRC classification results also showed that CMS1, in addition to mainly consisting of the C4 subtype, also contained a considerable number of the C2 subtype, while the CMS4 subtype mainly consisted of the C2 and C3 subtypes. Furthermore, our study results indicated that the C2 subtype was suitable for chemotherapy in combination with bevacizumab, possibly because the combination can inhibit the CAFs and abnormal blood vessel formation in the TME, thus alleviating the immune suppression of the immune cells. However, the C3 subtype was not suitable for chemotherapy in combination with bevacizumab because it only accumulated CAFs and abnormal blood vessel formation but lacked T cell infiltration. Therefore, we boldly speculated that the CMS1 and CMS4 subtypes in the CALGB/SWOG 80405 clinical trial may contain more C2 subtypes than those in the FIRE-3 clinical trial, leading to the CMS1 and CMS4 subtypes in the CALGB/SWOG 80405 clinical trial being more suitable for chemotherapy in combination with bevacizumab than cetuximab compared to the FIRE-3 clinical trial. Overall, the integration of CCCRC and CMS classification systems can provide valuable insights for understanding the divergent outcomes of the two clinical trials.

Our study has several limitations that should be considered when interpreting the results. First, the present study is a pre-clinical research and further validation of our findings requires additional basic experiments and prospective clinical trials. Second, it should be noted that due to the difficulty in evaluating the impact of removing individual TME-related signatures on clustering results, we performed cluster sensitivity analysis by removing the entire TME category. Third, given the relatively small sample size of scRNAseq and ST data used in this study, larger datasets are needed for analysis to better understand the complexity of the TME.

To conclude, our study proposed the CCCRC classification system and performed integrated data analysis to clearly characterize the molecular features and histopathologic characteristics of each CCCRC subtype, develop the corresponding personalized treatments for patients with the different CCCRC subtypes, and construct the simple gene classifier to facilitate clinical application. We believe that our study will serve as a research paradigm for dissecting the TME and for transitioning from molecular classification to clinical translation, thereby accelerating the understanding of the TME in CRC and contributing to the development of therapeutic targets against TME.

## Materials and methods

### Acquisition and processing of GEP for the investigation of CCCRC

A total of 2196 samples were obtained from ten publicly available datasets (**Supplementary file 1c**). The eight publicly available raw microarray datasets sequenced by the Affymetrix gene chip were downloaded from the Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo/>) and renormalized by the robust multi-array average (RMA) method, including GSE13067, GSE13294, GSE14333, GSE17536, GSE33113, GSE37892, GSE38832, and GSE39582. Samples that overlapped in GSE14333 and GSE17536 datasets were excluded from the GSE14333 dataset. The two RNA sequencing datasets (TCGA and CPTAC datasets) were obtained from the TCGA data portal (March 2022) (<https://portal.gdc.cancer.gov/>), and the count data were normalized by the 'voom' method. Ensembl IDs were annotated into gene symbols using GENCODE (v36). If the gene symbol had multiple probes or duplicates, the median value was calculated as its relative GEP. Before merging the microarray or RNAseq datasets into the CRC-AFFY or CRC-RNAseq cohort, the batch effects were examined using PCA and corrected using the 'Combat' function. The selection criteria of these patients included: (1) CRC primary tissue samples; (2) coming from the same sequencing platform; (3) surgically resected specimens. The exclusion criteria included: (1) CRC metastatic tissue; (2) puncture tissues; (3) having received radiation therapy or chemotherapy prior to surgery. Detailed information on the sample size and the corresponding clinicopathological data of the CRC-AFFY and CRC-RNAseq cohorts are summarized in **Supplementary file 1c** and **Supplementary file 1f**. The CPTAC dataset was excluded from the survival analysis due to a median follow-up time of less than 3 years. Additionally, samples with a follow-up time of zero were also excluded from the follow-up analysis.

### Calculation of the TME-related signature scores

After reviewing previously published studies, the Molecular Signatures Database (MSigDB; <http://www.gsea-msigdb.org/gsea/msigdb/index.jsp>), and the Reactome pathway portal (<https://reactome.org/PathwayBrowser/>), we identified relevant biomarker genes for tumor, immune, stromal, and metabolic reprogramming signatures. The 4525 genes from each of the 61 TME-related signatures are listed in **Supplementary file 1a**, as well as the source of each signature. GSVA with default parameters using the R package 'GSVA' was performed to calculate the signature score of each TME-related signature for each sample of each cohort separately based on the relative GEP (**Hänzelmann et al., 2013**).

### Normal tissue versus tumor tissue analysis

To assess the distribution of normal and tumor samples in the GSE39582 ( $n = 19$  normal) and TCGA ( $n = 41$  normal) datasets, the GEP of each dataset were re-normalized, including the normal samples (consistent with the description of data normalization above). PCOA based on 'Euclidean' distance was used to analyze the distribution between normal and CRC samples (**Dixon, 2003**). PERMANOVA test (**Zhu et al., 2021**) was used to evaluate whether the difference in 'Euclidean' distances between the normal and CRC samples was statistically significant (obtained using the R package 'vegan'; **Desgarenes et al., 2014**).

### Comprehensive characterization of CRC

The 'ConsensusClusterPlus' function of the R package 'ConsensusClusterPlus' (**Monti et al., 2003**) was applied to identify the optimal number of CCCRC based on the TME-related signatures in the CRC-AFFY cohort (partitioning around medoids (pam) clustering; 'Pearson' distance; 1000 iterations; from 2 to 6 clusters). The stability of the clusters was evaluated using the consensus matrices heatmap,

the empirical CDF plot, and the delta area plot. The purpose of the consensus matrices heatmap was to find the 'cleanest' cluster partition, with dark blue color block indicating that samples were often clustered together with a high consensus and white color block indicating that samples were not often clustered together with a low consensus. CDF plot displayed the consensus distribution for each  $k$  and the delta area plot represented the change in the area under the CDF curves, and both were used to determine the maximum consensus distribution. To verify the repeatability and robustness of the CCCRC subtypes in the CRC-RNaseq cohort, we used the 'pamr.predict' function of the R package 'pamr' (Tibshirani *et al.*, 2002) to extract centroids of each subtype and establish a PAMR classifier in the CRC-AFFY cohort. A threshold with minimum 10-fold cross-validation error was selected to identify the TME-related signatures that exhibit at least one non-zero difference between each subtype (seed = 11). PAM is a statistical technique to identify subsets of features that best characterize each class using nearest shrunken centroids (Tibshirani *et al.*, 2002). The technique is general and can be used in many other classification problems. The TME-related signature scores were normalized by the Z-scores before performing 'pamr.predict' analysis. PCOA based on 'Euclidean' distance was used to analyze the distribution of the CCCRC subtypes.

### Estimation of the TME cell abundance with other methods

The cell abundance of each sample was estimated based on the GEP using the MCP-counter algorithm (Becht *et al.*, 2016b) and the CIBERSORT (Newman *et al.*, 2015) algorithm, both of which have been validated using the GEP of the corresponding cell populations and the degree of cellular infiltration estimated by immunohistochemistry. The MCP-counter algorithm estimated the cell abundance of nine immune and stromal cell populations. The CIBERSORT algorithm, which applies the LM22 matrix, estimated the cell fraction of 22 immune cell populations. The ESTIMATE algorithm with default parameters was utilized to estimate the degree of infiltration of the total immune cells and stromal cells in the TME of each sample, as well as the tumor purity (Yoshihara *et al.*, 2013).

### Calculation of the other biological pathway enrichment scores

Human metabolism-related pathways were obtained from the KEGG database (<https://www.genome.jp/kegg/>). The 1660 genes assigned to 86 human metabolism-related pathways are listed in **Supplementary file 3e**. Ten oncogenic signatures containing 331 genes and the terminally exhausted T cell signature were retrieved from a previously published study (Beltra *et al.*, 2020; Sanchez-Vega *et al.*, 2018; **Supplementary file 3e**). GSEA was performed to calculate the enrichment score of each signature for each sample of each cohort separately based on the relative GEP. To identify the potential differences in the biological functions of genes among CCCRC subtypes, GSEA was performed based on the gene signatures using the R package 'clusterprofiler' (Yu *et al.*, 2012).

### Histopathologic examination of the TCGA-CRC samples

A total of 616 TCGA CRC diagnostic HE-stained WSIs were downloaded from the TCGA data portal (March 2022) (<https://portal.gdc.cancer.gov/>), and the WSIs were examined blindly by two experienced pathologists. A total of 254 WSIs were included after removing the WSIs with poor quality and without views of the IM (**Supplementary file 1i**). According to the semi-quantitative pathological assessment of lymphocytes and their spatial location with malignant epithelial cells, the pathologist classified CRC into three immunophenotypes: 'desert', 'excluded', and 'inflamed', as previously described (Chen and Mellman, 2017; Rosenberg *et al.*, 2016). The inflamed phenotype was characterized by abundant lymphocytes in direct contact with malignant cells, the excluded phenotype was characterized by lymphocytes merely present in the stroma within or adjacent to the main tumor mass, and the desert phenotype was characterized by the lack of lymphocytes and stroma.

### Artificial intelligence-enabled spatial analysis of CRC-WSIs

We performed artificial intelligence (AI)-enabled spatial analysis of WSIs and developed a CRC-multiclass model to identify eight tissue types: tumor, stroma, lymphocyte, normal colon mucosa, debris, adipose, mucin, and muscle, and quantified the abundances of the tumor, stroma, and lymphocytes in the CT region and the IM region, respectively. The CRC-multiclass model consisted of two sequential parts: a muscle/non-muscle classifier that could distinguish each muscle (MUS) patch in HE-stained WSIs, and a seven-tissue classifier that could classify seven tissue types: adipose



(ADI), debris (DEB), lymphocytes (LYM), mucus (MUC), normal colon mucosa (NORM), cancer-associated stroma (STR), and colorectal adenocarcinoma epithelium (TUM). To develop the CRC-multiclass model, we used three dataset including NCT-CRC-HE-100k dataset, NCT-CRC-HE-7k dataset (<https://zenodo.org/record/1214456#.YyRJGWB6RmM>), and TCGA-CRC dataset. At first, pathologists conducted quality control on the NCT-CRC-HE-100K dataset. Tissues with inaccurate labeling were removed from the dataset. The stromal tissue dataset was reduced from 10,446 patches to 9638 patches, and the lymphocyte tissue dataset was reduced from 11,555 patches to 10,453 patches. After up-sampling stroma tissue patch, 4000 additional patches were added. Total of 90,686 patches were selected. The dataset contains patches that are 224 pixels by 224 pixels in size and was generated from 85 patients. Then, 9 WSIs from 9 patients in the TCGA-CRC dataset were annotated with tissue label, resulting in a total of 4288 patches. The preprocessing step involved identifying the outline of the entire tissue and then segmenting it. Segmentation was performed by non-overlapping 224 × 224 patch extraction based on the tissue outline, and a patch was considered valid if 75% of its area fell within the annotation boundaries. We did not do any modification on NCT-CRC-7k dataset, which includes 7180 patches from 50 patients with size of 224 pixels by 224 pixels. To develop the muscle/non-muscle model, The NCT-CRC-HE-100k and NCT-CRC-HE-7k datasets were merged and subsequently partitioned into an internal training set of 68,506 patches, an internal validation set of 14,679 patches, and an internal test set of 14,681 patches. This model predicts whether a tissue patch is a non-muscle patch. If the probability of a sample belonging to the non-muscle class is greater than 0.99, it is classified as non-muscle and then would be fed to our seven-tissue classifier. The seven-tissue classifier was trained with 54,597 patches and validated with 23,400 patches on the CRC-HE-100k dataset that underwent quality control and removal of muscle tissue, and the muscle-free CRC-HE-7k dataset of 5741 patches was used as the internal test set. The 4288 patches of TCGA-CRC dataset are the external testing set for evaluating externally for these two models.

The WSI tissue type prediction pipeline was as follows. First, the background was removed by the preprocessing steps. The preprocessing step for WSI tissue segmentation involved converting the color space from RGB to HSV, applying a median filter to the saturation channel to remove noise, and finally using a thresholding approach to segment the WSI tissue from background. Second, the tissues regions of WSIs were segmented into non-overlapping image patches at a resolution of 0.5 μm/pixel (×20 magnification). It is worth noting that if the WSI consisted of ×40 magnification, it was down-sampled to ×20 magnification. Next, the image patches were fed into the CRC-multiclass model. If an image patch was determined to be non-muscle by the muscle/non-muscle classifier, it was fed into the seven-tissue classifier to predict its tissue class. Both the muscle/non-muscle classifier and the seven-tissue classifier used the same model architecture, which involved utilizing a pre-trained ResNet50 model for transfer learning. The last layer of the feature extraction layer was connected to two fully connected layers, with ReLU as the activation function and a dropout of 0.4 applied to the first fully connected layer. The final activation function used was softmax. During training, the optimizer used was stochastic gradient descent, with a learning rate of 0.001 and momentum of 0.9. The muscle/non-muscle classifier was trained for 20 epochs, while the seven-tissue classifier was trained for 10 epochs. During this experiment, we tested three model architectures, including ResNet50, vgg16, and Inception V3 for the CRC-tissue classifier. According to the accuracy of seven tissues (tumor, stroma, lymphocytes, normal colon mucosa, debris, adipose, and mucin) in the NCT-CRC-HE-7k dataset, the performance of ResNet50 was the best, which was the reason we selected ResNet50 as the basic model architecture.

After recognizing the CRC tissue types by our deep learning model automatically, we quantified the abundances of the tumor, lymphocytes, and stroma in the CT region and the IM region. The quantification pipeline consisted of three steps. First, we used the open-source software QuPath-0.3.2 (<https://qupath.github.io/>; *Bankhead et al., 2017*) to delineate the CT and IM region. The IM region was defined as 500 mm outside the CT region (*Kather et al., 2018*). The CT and IM regions were manually annotated by two experienced pathologists to reduce bias. Next, we utilized our CRC-multiclass model to predict the tissue type of patches within the CT and IM regions after segmenting images within the annotated CT and IM contours into 224-pixel by 224-pixel patches. This enabled us to determine the number of lymphocyte (LYM), stroma (STR), and tumor (TUM) patches within these regions. Then, the abundances of the tumor, lymphocytes, and stroma in different regions of each WSI

were quantified with an area ratio of their area to region area. A total of 254 TCGA-CRC WSIs were quantified with this quantification pipeline.

### Acquisition of signatures associated with the ICB therapy response

The TIDE score was calculated using GEP, and it was used to evaluate the degree of T cell dysfunction and T cell exclusion (*Jiang et al., 2018*). The higher the score, the later the dysfunction stage of T cells or the higher the degree of T cell exclusion. The gene expression average of all samples in each cohort was used as the normalized control and the normalized GEP was uploaded to the TIDE website (<http://tide.dfci.harvard.edu/>).

### Acquisition and processing of CRC multi-omics data

Masked somatic mutation data ( $n = 571$  samples), masked copy number segment data ( $n = 609$  samples), and DNA methylation beta values (Illumina human methylation 450) (45 normal samples and 390 tumor samples) were download from the TCGA data portal (March 2022) (<https://portal.gdc.cancer.gov/>). The liquid chromatography–tandem mass spectrometry (LC–MS/MS)-based proteomic data for the TCGA CRC samples ( $n = 88$  samples) were obtained from a previously published study (*Zhang et al., 2014*). The R package ‘maftools 2.6.05’ with default parameters was used to analyze the somatic mutation data. Synonymous mutations were regarded as wild-type, and genes with mutation rates <5% were excluded. Nonsynonymous mutations were used to calculate tumor mutation burden (TMB). SCNAs defined by the GISTIC2.0 module on the GenePattern website (<https://www.genepattern.org/>), including arm-level gain (1), and high amplification (2), diploid/normal (0), arm-level deletion (–1), and deep deletion (–2). The CINmetrics algorithm was used to calculate CIN signature, including SCNAs count and FGA, which was proposed by Vishalozza et al. (<https://rdrr.io/github/lasseignelab/CINmetrics/>; *Oza et al., 2023*) based on previously published studies (*Baumbusch et al., 2013*; *Chin et al., 2007*; *Davison et al., 2014*). If somatic mutation events or SCNAs occurred in one or more genes in the oncogenic pathway, the tumor sample was considered altered in a given pathway. The MSI status was obtained from the CMS website (<https://www.synapse.org/#!Synapse:syn2623706>). Tumor neoantigen signature and the ITH data in the CRC-RNAseq cohort were obtained from a previously published study (*Thorsson et al., 2018*). The prevalence of somatic mutation events or SCNAs was compared among CCCRC cases using Fisher’s exact test or chi-square test. For the DNA methylation data, probes located in promoter CpG islands were extracted, including TSS200, 1stExon, TSS1500, and 5’UTR. The probes detected on X and Y chromosomes or any probe with NA value were removed. For genes with multiple probes mapped to the promoter, the median beta value was calculated as the degree of gene methylation. The beta-value difference was defined as the difference between the mean beta value of each CCCRC sample and normal samples, and Wilcoxon rank-sum test was used to test whether the difference was statistically significant. DMGs between normal and CRC samples were defined as  $|\text{mean beta value}| < 0.2$  in normal samples and  $|\text{mean beta value}| > 0.5$  in CRC samples with FDR <0.05.

We performed differential analysis between each of the four subtypes and the remaining subtypes using Wilcoxon rank-sum test to identify DMGs with FDR <0.001 and DEPs with p-value <0.05. The difference of DMGs and DEPs was defined as the difference between each of the four subtypes and the remaining subtypes. The ‘limma’ package was used to identify DEGs with FDR <0.001 between each of the four subtypes and the remaining subtypes. To identify subtype-specific DMGs, DEPs, and DEGs in one of the subtypes, we excluded those that were found to be differentially expressed in comparisons between one of the other subtypes and the remaining subtypes. Additionally, we performed differential analysis of methylation gene, gene expression, and protein levels between each subtype based on the above method. Enrichment analysis was performed by the R package ‘clusterProfiler’ (*Yu et al., 2012*). p-values were adjusted for multiple comparisons by the Benjamini–Hochberg-corrected FDR.

### Regulon analysis

The R package ‘RTN’ was used to reconstruct the transcriptional regulatory networks of regulons (*Robertson et al., 2017*), including 31 transcription factors and 82 chromatin remodeling genes, that were associated with CRC (*Vymetalkova et al., 2019*; *Xu et al., 2021*; *Supplementary file 3f*). Mutual information and Spearman’s correlation analysis were utilized to infer the possible associations

between a regulator and all possible targets from the GEP, and the permutation algorithm was used to eliminate associations with an FDR  $>1 \times 10^{-5}$ . Unstable associations were removed by bootstrap analysis ( $n = 1000$ ), and the weakest association in triangles consisting of two regulators and common targets were eliminated by the data processing inequality algorithm. Two-tailed gene set enrichment analysis was used to calculate the regulon activity score for each sample.

### Publicly available CRC classification systems

To classify CRC samples into different CRC subtypes according to the previously published gene classifier, gene lists for the four classifiers were extracted from relevant publications and summarized (**Supplementary file 3g**). These classifiers included the Budinska classification system (**Budinska et al., 2013**), the De Sousa classification system (**De Sousa E Melo et al., 2013**), the Roepman classification system (**Roepman et al., 2014**), and the Sadanandam classification system (**Sadanandam et al., 2013**). The nearest template prediction (NTP) algorithm was employed to classify the samples and to generate an FDR to assess the classification robustness. For NTP implementation, we screened genes that were specifically and positively associated with one subtype according to the screening strategies of a previously published study (**Medico et al., 2015**).

### Bulk RNAseq and scRNAseq data processing of the GSE108989 dataset

A total of 12 CRC samples with bulk RNAseq and scRNAseq data were obtained from the **GSE108989** dataset (**Zhang et al., 2018; Supplementary file 2g**). To identify the CCCRC subtypes, bulk RNAseq with transcripts per million (TPM) was further log<sub>2</sub>-transformed, and GSVA was performed to calculate the signature score of each TME-related signature in each sample based on the GEP. The PAMR classifier was used to classify CRC samples into four CCCRC subtypes based on the TME-related signatures (seed = 11, threshold = 0.566). For scRNAseq data processing, the raw GEP were normalized and selected according to the following criteria: cells with  $>200$  genes and  $<7000$  genes and  $<20\%$  of mitochondrial gene expression in UMI counts, which was determined using the R package 'Seurat'. Counts of the filtered matrix for each gene were normalized to the total library size with the Seurat 'NormalizeData' function. The 'FindVariableGenes' function was used to identify 2000 hypervariable genes for unsupervised clustering. Next, each integrated feature was centered to a mean of zero and scaled by the standard deviation with the Seurat 'ScaleData' function. The 'RunPCA' function was used for PCA. We identified diverse T cell clusters using the 'FindClusters' function, and set the resolution parameter to 0.5. Each cell cluster was compared to the other clusters by the 'FindAllMarkers' function to identify DEGs (only pos: TRUE, min.Pct: 0.25, logFc.threshold: 0.25). Cell annotation was carried out by consulting the latest cell marker databases, such as CellMarker (<https://www.biolegend.com/en-us/cell-markers>) and PanglaoDB (<https://ngdc.cncb.ac.cn/databasecommons/database/id/6917>), combined with a previously published study (**Zhang et al., 2018**). To define the feature genes for each CCCRC subtype, differential expression analysis between CCCRC subtypes was performed using the 'FindMarkers' function. FDR  $<0.05$  were considered statistically significant.

### Collection and processing of therapy-associated datasets

Therapy-associated datasets were used to explore the treatment strategies for each CCCRC subtype. GEP of GSE103479 and GSE104645 datasets were downloaded from the GEO database. If the gene symbol was annotated with multiple probes, the median value was used as the expression of the gene. The clinical data of the GSE104645 dataset were obtained from the supplementary file of a study by **Okita et al., 2018**. The GSE103479 dataset contained 156 stage II and III CRC patients with or without 5-FU-based chemotherapy. The GSE104645 dataset contained 193 mCRC patients treated with chemotherapy, a combination of chemotherapy and bevacizumab, or anti-EGFR therapies. The available RNAseq expression dataset of patients treated with anti-PD-1 therapy was also downloaded. The Gide (PRJEB23709) dataset was downloaded, and the raw fastq files were re-analyzed. The RNA reads were aligned using STAR v2.5.3 and quantified as TPM using RSEM v1.3.0 and log<sub>2</sub>-transformed. Ensembl IDs were annotated into gene symbols using GENCODE v36. The GEP of Hugo (GSE78220) and Jung (GSE135222) datasets and the corresponding clinical data were downloaded from the GEO database, and the FPKM values were converted to log<sub>2</sub>-transformed TPM values. We obtained the GEP ( $n = 348$ ) of urothelial carcinoma patients in the IMvigor210 dataset treated with

anti-PD-L1 therapy and the corresponding clinical data using the R package 'IMvigor210CoreBiologies 1.0.0' (IMvigor210 dataset), and the count values were converted to log<sub>2</sub>-transformed TPM values. To reduce batch effects and tissue type-specific effects, we first performed GSVA analysis of the TME-related signatures in each dataset, and the signature scores were normalized by Z-scores method. Next, we used the PAMR classifier to classify the samples into the four CCCRC subtypes based on the TME-related signatures in each dataset (seed = 11, threshold = 0.566). Detailed information on the sample size and the corresponding treatment data of the therapy-associated datasets are summarized in **Supplementary file 3h**.

## Construction of a pre-clinical model

To explore the treatment for each CCCRC subtype using cancer cell line drug sensitivity experiments, we developed a pre-clinical model based on subtype-specific, cancer cell-intrinsic gene markers according to a previously published study (Eide *et al.*, 2017). Firstly, the 'limma' package was used to identify DEGs with FDR <0.05 between each of the four subtypes and the remaining subtypes in the CRC-AFFY cohort. To identify subtype-specific genes in one of the subtypes, we excluded those that were found to be differentially expressed in comparisons between one of the other subtypes and the remaining subtypes. The upregulated subtype-specific genes (log<sub>2</sub>FC >0 and FDR <0.05) were ranked based on their log<sub>2</sub>FC and selected the top 500 genes for further gene screening. Secondly, the GEP of human CRC tissues versus patient-derived xenografts (PDX) in the GSE35144 dataset by the R package 'limma' was used to remove those genes associated with stromal and immune components. DEGs with FDR >0.5 and log<sub>2</sub>FC <1 between human CRC tissues versus PDX were considered as cancer cell-intrinsic genes. Thirdly, we also utilized human CRC cell lines to obtain cancer cell-intrinsic genes. A total of 71 human CRC cell lines with RNAseq data (log<sub>2</sub>TPM) was obtained from the Genomics of Drug Sensitivity in Cancer (GDSC) database (<https://depmap.org/portal/download/all/>), 43 of which had dose–response curve (AUC) values. The MSI status, FGA and TMB information of CRC cell lines was obtained from cbiportal website ([https://www.cbiportal.org/study/summary?id=ccl\\_broad\\_2019](https://www.cbiportal.org/study/summary?id=ccl_broad_2019)). RNAseq data for 71 human CRC cell lines were used to further determine the cancer cell-intrinsic genes and genes among the top 25% within (1) the 10–90% percentile range of the largest expression values and (2) the highest expression in at least three samples. The subtype-specific genes and cancer cell-intrinsic genes were intersected to generate the gene list for developing the pre-clinical model. The pre-clinical model was developed using the NTP function of R package 'CMScaller', which can be applied to cross-tissues and cross-platform predictions (Hoshida, 2010). The GEP (log<sub>2</sub>TPM) of 71 human CRC cell lines normalized by the Z-score were input into the pre-clinical model, and the cell lines were divided into four CCCRC subtypes. We then compared the drug responses among the CCCRC subtypes by analyzing the differences in the AUC values.

## Spatial transcriptomics data processing

To investigate the spatial distribution relationship between four CCCRC subtypes of tumor cells, T cells, and stromal cells, we conducted a re-analysis of publicly available CRC ST data obtained from the 10X website (<https://www.10xgenomics.com/resources/datasets>). The Space Ranger output files were then processed with Seurat (V4.1.1) (Hao *et al.*, 2021) using SCTransform for normalization (Hafemeister and Satija, 2019). RunPCA were used to dimension reduction and RunUMAP to visualize the data. We used 'ssGSEA' method implemented in the R package 'GSVA' to score the six cell types (C1–C4 subtype cancer cells, mesenchymal cells, and T cells) (Hänzelmann *et al.*, 2013). The 'ssGSEA' method has been previously demonstrated to be highly reliable and suitable for ST data analysis (Wu *et al.*, 2022). The cell-type-rich region was defined as the ssGSEA score of each cell type from one spot larger than the 75% quantile of this cell type. The markers for the six cell types are listed in the **Supplementary file 1a** and **Supplementary file 3a**.

## Discovery and validation of the CCCRC classifier

In order to facilitate the widespread application of CCCRC classification system, we established a simple gene classifier to predict CCCRC subtypes. Firstly, we filtered genes based on their mean expression and variance in the CRC-AFFY cohort, and genes with expression and variance below the bottom 25% were removed. Then, we applied the RF algorithm in the R package 'caret' to perform feature selection on the CCCRC subtype-specific genes of the CRC-AFFY cohort. The top

20 most informative features for each subtype were ranked and selected based on the impurity measure generated by the algorithm. This allowed us to identify critical genes that are strongly associated with each CCCRC subtype and develop the CCCRC classifier. Next, we randomly divided the CRC-AFFY cohort into training and test sets at a ratio of 7:3 using 'createDataPartition' function provided in the R package 'caret' (seed = 123). The GEP was normalized with Z-scores prior to model training and validation. The CCCRC classifiers were trained with the top 80 upregulated subtype-specific genes using the RF, SVM, xgboost, and Logistic Regression algorithms implemented in the R package 'caret'. Finally, we validated the CCCRC classifier on the GSE14333 and GSE17536 datasets, as well as the CRC-AFFY cohort. The consensus clustering result of the CRC-AFFY cohort and the PAMR classifier result of the CRC-RNAseq cohort were used as 'gold-standard' ( $n = 2196$  samples). We evaluated the predictive performance of the CCCRC classifier by evaluating measures such as accuracy value and F1 score, which were generated using the 'confusionMatrix' function provided in the R package 'caret'.

## Statistical analyses

All statistical analyses were conducted by R 4.0.2 software. Statistical significance of the comparisons for continuous variables and categorical variables was assessed by the Wilcoxon rank-sum test or Kruskal–Wallis test and Fisher's exact test or chi-square test, respectively. Correlations between variables were estimated by Spearman's correlation analysis or Pearson's correlation analysis. Patients were divided into either high or low gene expression groups by the best cutoff calculated by the R package 'survminer'. The Kaplan–Meier method with log-rank test was utilized to generate the survival curves. Univariate and multivariate Cox proportional hazard regression analyses were performed to generate 95% confidence intervals and hazard ratios. Two-sided  $p$ -values  $<0.05$  were considered statistically significant.

## Acknowledgements

We thank TCGA and GEO databases and the cBioportal website for free use. We thank Nanjing Simcere Medical Laboratory Science Co, Ltd and Jiangsu Simcere Diagnostics Co, Ltd, and all the members of its AI and bioinformatics team, for generously sharing their experience and codes.

---

## Additional information

### Competing interests

Xiaoping Qu, Yiran Zheng, Minghui Ge, Linlin Yan: are affiliated with Nanjing Simcere Medical Laboratory Science Co., Ltd. and Jiangsu Simcere Diagnostics Co., Ltd. These authors have no financial interests to declare. The other authors declare that no competing interests exist.

### Funding

Funder	Grant reference number	Author
National Key R&D Program of China	No. 2021YFF1201004	Li Liang

The funders had no role in study design, data collection, and interpretation, or the decision to submit the work for publication.

### Author contributions

Xiangkun Wu, Conceptualization, Data curation, Validation, Investigation, Methodology, Writing – original draft; Hong Yan, Conceptualization, Validation, Investigation, Methodology; Mingxing Qiu, Conceptualization, Data curation, Validation, Methodology; Xiaoping Qu, Resources, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft; Jing Wang, Conceptualization, Data curation, Methodology; Shaowan Xu, Yiran Zheng, Minghui Ge, Linlin Yan, Conceptualization, Investigation, Methodology; Li Liang, Conceptualization, Supervision, Funding acquisition, Investigation, Writing – review and editing

**Author ORCIDs**Li Liang  <http://orcid.org/0000-0003-2424-4926>**Decision letter and Author response**Decision letter <https://doi.org/10.7554/eLife.86032.sa1>Author response <https://doi.org/10.7554/eLife.86032.sa2>

---

## Additional files

**Supplementary files**

- Supplementary file 1. Identification and validation of comprehensive characterization of colorectal cancer (CCCRC) classification. (a) The 61 tumor microenvironment (TME)-related signatures used to comprehensively characterize CRC. (b) Overlapping genes between 61 TME-related signatures. (c) The datasets used to identify and validate of CCCRC classification. (d) Univariate cox analysis and Kaplan–Meier method showing the ability of the 61 TME-related signatures to predict disease-free survival and overall survival in the CRC-AFFY cohort. (e) Difference analysis of metabolism signature scores between the four CCCRC subtypes in the CRC-AFFY cohort. (f) The information of clinicopathological characteristics and CCCRC subtypes in the CRC-AFFY and CRC-RNAseq cohorts. (g) Associations between clinicopathological characteristics and CCCRC subtypes. (h) Multivariate Cox proportional hazard regression analysis for disease-free survival and overall survival adjusting for the CCCRC subtypes and clinicopathological characteristics in the combined cohort (the CRC-AFFY and CRC-RNAseq cohorts). (1) A total of 254 hematoxylin and eosin (HE)-stained whole-slide images used to evaluate histological characteristics of CCCRC subtypes.
- Supplementary file 2. The multi-omics analysis of the comprehensive characterization of colorectal cancer (CCCRC) subtypes. (a) The different frequency of somatic mutations between the four CCCRC subtypes in TCGA-CRC dataset. The different frequency of somatic copy number alterations between the four CCCRC subtypes in TCGA-CRC dataset. (b) The fraction of altered samples per oncogenic pathway in each CCCRC subtype. (c) Lists of differentially expressed genes between each of the four subtypes and the remaining subtypes and between each subtype. (d) The correlation between all genes and the exclusion and dysfunction scores obtained from the Tumor Immune Dysfunction and Exclusion (TIDE) website. (e) Univariate cox analysis and Kaplan–Meier method showing the ability of 20 CCCRC genes to predict disease-free survival in the CRC-AFFY cohort. (f) A total of 7766T cells from 12 CRC patients in GSE108989 dataset. (g) The marker genes of the T cell clusters.
- Supplementary file 3. Clinical application of comprehensive characterization of colorectal cancer (CCCRC) classification system. (a) The 335 CCCRC subtype-specific and cancer cell-intrinsic gene markers used to construct the pre-clinical model. (b) The 71 human colorectal cancer cell lines grouped according to CCCRC classification system. (c) A total of 80 upregulated subtype-specific genes used to construct the CCCRC classifier. (d) The performance of the CCCRC classifier in the CRC-AFFY and the CRC-RNAseq cohort. Lists of 86 human metabolism-related signatures, 10 oncogenic signatures and terminally exhausted T cell signature. (e) Transcription factors and chromatin modifiers in colorectal cancer. (f) Publicly available CRC molecular classification systems. (g) The datasets used to reveal the therapy strategies for four CCCRC subtypes.
- MDAR checklist

**Data availability**

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article. This paper does not report original codes. The CCCRC classifier is publicly available at [https://github.com/XiangkunWu/CCCRC\\_classifier](https://github.com/XiangkunWu/CCCRC_classifier) (copy archived at [Wu, 2023c](#)). All software is freely or commercially available and is listed in the Methods. Any additional information required to reanalyze the data reported in this work paper is available from the lead contact upon request.

The following previously published datasets were used:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Jorissen RN	2009	Expression data from primary colorectal cancers	<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE13067">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE13067</a>	NCBI Gene Expression Omnibus, GSE13067
Jorissen RN	2009	Expression data from primary colorectal cancers	<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE13294">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE13294</a>	NCBI Gene Expression Omnibus, GSE13294
Jorissen RN	2010	Expression data from 290 primary colorectal cancers	<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE14333">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE14333</a>	NCBI Gene Expression Omnibus, GSE14333
Smith JJ	2009	Metastasis Gene Expression Profile Predicts Recurrence and Death in Colon Cancer Patients (Moffitt Samples)	<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE17536">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE17536</a>	NCBI Gene Expression Omnibus, GSE17536
Medema JP	2011	AMC colon cancer AJCCII	<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE33113">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE33113</a>	NCBI Gene Expression Omnibus, GSE33113
Laibe S	2012	A seven-gene signature aggregates a subgroup of stage II colon cancers with stage III	<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE37892">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE37892</a>	NCBI Gene Expression Omnibus, GSE37892
Tripathi MK	2014	NFAT transcriptional activity is associated with metastatic capacity in colon cancer	<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE38832">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE38832</a>	NCBI Gene Expression Omnibus, GSE38832
Marisa L	2013	Gene expression Classification of Colon Cancer defines six molecular subtypes with distinct clinical, molecular and survival characteristics [Expression]	<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE39582">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE39582</a>	NCBI Gene Expression Omnibus, GSE39582
Allen WL	2017	Gene expression data from stage II and III treated and untreated colorectal cancer patients	<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE103479">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE103479</a>	NCBI Gene Expression Omnibus, GSE103479
Okita A	2018	Consensus molecular subtypes classification of colorectal cancer as a predictive factor for chemotherapeutic efficacy against metastatic colorectal cancer	<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE104645">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE104645</a>	NCBI Gene Expression Omnibus, GSE104645
Gide TN et al	2019	Biomarkers of response and resistance to checkpoint blockade immunotherapy in metastatic melanoma	<a href="https://www.ncbi.nlm.nih.gov/bioproject/PRJEB23709">https://www.ncbi.nlm.nih.gov/bioproject/PRJEB23709</a>	NCBI BioProject, PRJEB23709

*Continued on next page*

Continued

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Hugo W, Zaretsky JM, Sun L, Song C, Homet-Moreno B, Hu-Lieskovan S, Berent-Maoz B, Pang J, Chmielowski B, Cherry G, Seja E, Lomeli S, Kong X, Kelley MC, Sosman JA, Johnson DB, Riba A	2016	mRNA expressions in pre-treatment melanomas undergoing anti-PD-1 checkpoint inhibition therapy	<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE78220">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE78220</a>	NCBI Gene Expression Omnibus, GSE78220
Jung H, Kim J	2019	DNA methylation loss coupled with mitotic cell division promotes immune evasion of tumours with high mutation load [RNA-seq]	<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE135222">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE135222</a>	NCBI Gene Expression Omnibus, GSE135222
Nickles D, Senbabaoglu Y, Sheinson D	2018	TGF- $\beta$ attenuates tumor response to PD-L1 blockade by contributing to exclusion of T cells	<a href="http://research-pub.gene.com/IMvigor210CoreBiologies/#transcriptome-wide-gene-expression-data">http://research-pub.gene.com/IMvigor210CoreBiologies/#transcriptome-wide-gene-expression-data</a>	IMvigor210, 10.1038/nature25501
10x GENOMICS	2023	Human Colorectal Cancer, 11 mm Capture Area (FFPE)	<a href="https://www.10xgenomics.com/resources/datasets/human-colorectal-cancer-11-mm-capture-area-ffpe-2-standard">https://www.10xgenomics.com/resources/datasets/human-colorectal-cancer-11-mm-capture-area-ffpe-2-standard</a>	10x GENOMICS, /human-colorectal-cancer-11-mm-capture-area-ffpe-2-standard
10x GENOMICS	2020	Human Intestine Cancer (FPPE)	<a href="https://www.10xgenomics.com/resources/datasets/human-intestine-cancer-1-standard">https://www.10xgenomics.com/resources/datasets/human-intestine-cancer-1-standard</a>	10x GENOMICS, /human-intestine-cancer-1-standard
Zhang L	2018	Lineage tracking reveals dynamic relationships of T cells in colorectal cancer	<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE108989">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE108989</a>	NCBI Gene Expression Omnibus, GSE108989

## References

- Aderka D**, Stintzing S, Heinemann V. 2019. Explaining the unexplainable: discrepancies in results from the CALGB/SWOG 80405 and FIRE-3 studies. *The Lancet. Oncology* **20**:e274–e283. DOI: [https://doi.org/10.1016/S1470-2045\(19\)30172-X](https://doi.org/10.1016/S1470-2045(19)30172-X), PMID: 31044725
- Anderson NM**, Simon MC. 2020. The tumor Microenvironment. *Current Biology* **30**:R921–R925. DOI: <https://doi.org/10.1016/j.cub.2020.06.081>, PMID: 32810447
- Andrei P**, Battuello P, Grasso G, Rovera E, Tesio N, Bardelli A. 2022. Integrated approaches for precision oncology in colorectal cancer: the more you know, the better. *Seminars in Cancer Biology* **84**:199–213. DOI: <https://doi.org/10.1016/j.semcan.2021.04.007>, PMID: 33848627
- Archambault AN**, Lin Y, Jeon J, Harrison TA, Bishop DT, Brenner H, Casey G, Chan AT, Chang-Claude J, Figueiredo JC, Gallinger S, Gruber SB, Gunter MJ, Hoffmeister M, Jenkins MA, Keku TO, Marchand LL, Li L, Moreno V, Newcomb PA, et al. 2021. Nongenetic determinants of risk for early-onset colorectal cancer. *JNCI Cancer Spectrum* **5**:pkab029. DOI: <https://doi.org/10.1093/jncics/pkab029>, PMID: 34041438
- Bagaev A**, Kotlov N, Nomie K, Svekolkina V, Gafurov A, Isaeva O, Osokin N, Kozlov I, Frenkel F, Gancharova O, Almog N, Tsiper M, Ataulkhanov R, Fowler N. 2021. Conserved Pan-cancer Microenvironment subtypes predict response to Immunotherapy. *Cancer Cell* **39**:845–865. DOI: <https://doi.org/10.1016/j.ccell.2021.04.014>, PMID: 34019806
- Bankhead P**, Loughrey MB, Fernández JA, Dombrowski Y, McArt DG, Dunne PD, McQuaid S, Gray RT, Murray LJ, Coleman HG, James JA, Salto-Tellez M, Hamilton PW. 2017. Qupath: open source software for Digital pathology image analysis. *Scientific Reports* **7**:16878. DOI: <https://doi.org/10.1038/s41598-017-17204-5>, PMID: 29203879



- Baumbusch LO**, Helland Å, Wang Y, Liestøl K, Schaner ME, Holm R, Etemadmoghadam D, Alsop K, Brown P, Mitchell G, Fereday S, DeFazio A, Bowtell DDL, Kristensen GB, Lingjærde OC, Børresen-Dale A-L, Australian Ovarian Cancer Study Group. 2013. High levels of Genomic aberrations in Serous ovarian cancers are associated with better survival. *PLOS ONE* **8**:e54356. DOI: <https://doi.org/10.1371/journal.pone.0054356>, PMID: 23372714
- Becht E**, de Reyniès A, Giraldo NA, Pilati C, Buttard B, Lacroix L, Selves J, Sautès-Fridman C, Laurent-Puig P, Fridman WH. 2016a. Immune and Stromal classification of colorectal cancer is associated with molecular subtypes and relevant for precision Immunotherapy. *Clinical Cancer Research* **22**:4057–4066. DOI: <https://doi.org/10.1158/1078-0432.CCR-15-2879>, PMID: 26994146
- Becht E**, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, Selves J, Laurent-Puig P, Sautès-Fridman C, Fridman WH, de Reyniès A. 2016b. Estimating the population abundance of tissue-infiltrating immune and Stromal cell populations using gene expression. *Genome Biology* **17**:249. DOI: <https://doi.org/10.1186/s13059-016-1113-y>, PMID: 27908289
- Bejarano L**, Jordão MJC, Joyce JA. 2021. Therapeutic targeting of the tumor Microenvironment. *Cancer Discovery* **11**:933–959. DOI: <https://doi.org/10.1158/2159-8290.CD-20-1808>, PMID: 33811125
- Beltra J-C**, Manne S, Abdel-Hakeem MS, Kurachi M, Giles JR, Chen Z, Casella V, Ngiew SF, Khan O, Huang YJ, Yan P, Nzingha K, Xu W, Amaravadi RK, Xu X, Karakousis GC, Mitchell TC, Schuchter LM, Huang AC, Wherry EJ. 2020. Developmental relationships of four exhausted Cd8(+) T cell Subsets reveals underlying transcriptional and epigenetic landscape control mechanisms. *Immunity* **52**:825–841. DOI: <https://doi.org/10.1016/j.immuni.2020.04.014>, PMID: 32396847
- Binnewies M**, Roberts EW, Kersten K, Chan V, Fearon DF, Merad M, Coussens LM, Gabrilovich DI, Ostrand-Rosenberg S, Hedrick CC, Vonderheide RH, Pittet MJ, Jain RK, Zou W, Howcroft TK, Woodhouse EC, Weinberg RA, Krummel MF. 2018. Understanding the tumor immune Microenvironment (TIME) for effective therapy. *Nature Medicine* **24**:541–550. DOI: <https://doi.org/10.1038/s41591-018-0014-x>, PMID: 29686425
- Borriello L**, Nakata R, Sheard MA, Fernandez GE, Spoto R, Malvar J, Blavier L, Shimada H, Asgharzadeh S, Seeger RC, DeClerck YA. 2017. Cancer-associated fibroblasts share characteristics and Protumorigenic activity with Mesenchymal Stromal cells. *Cancer Research* **77**:5142–5157. DOI: <https://doi.org/10.1158/0008-5472.CAN-16-2586>, PMID: 28687621
- Budinska E**, Popovici V, Tejpar S, D'Ario G, Lapique N, Sikora KO, Di Narzo AF, Yan P, Hodgson JG, Weinrich S, Bosman F, Roth A, Delorenzi M. 2013. Gene expression patterns unveil a new level of molecular heterogeneity in colorectal cancer. *The Journal of Pathology* **231**:63–76. DOI: <https://doi.org/10.1002/path.4212>, PMID: 23836465
- Bufill JA**. 1990. Colorectal cancer: evidence for distinct genetic categories based on proximal or distal tumor location. *Annals of Internal Medicine* **113**:779–788. DOI: <https://doi.org/10.7326/0003-4819-113-10-779>, PMID: 2240880
- Calon A**, Lonardo E, Berenguer-Llargo A, Espinet E, Hernando-Momblona X, Iglesias M, Sevillano M, Palomo-Ponce S, Tauriello DVF, Byrom D, Cortina C, Morral C, Barceló C, Tosi S, Riera A, Attolini CS-O, Rossell D, Sancho E, Batlle E. 2015. Stromal gene expression defines poor-prognosis subtypes in colorectal cancer. *Nature Genetics* **47**:320–329. DOI: <https://doi.org/10.1038/ng.3225>, PMID: 25706628
- Cancer Genome Atlas Network**. 2012. Comprehensive molecular characterization of human colon and Rectal cancer. *Nature* **487**:330–337. DOI: <https://doi.org/10.1038/nature11252>, PMID: 22810696
- Caswell DR**, Swanton C. 2017. The role of tumour heterogeneity and Clonal Cooperativity in metastasis, immune evasion and clinical outcome. *BMC Medicine* **15**:133. DOI: <https://doi.org/10.1186/s12916-017-0900-y>, PMID: 28716075
- Chen DS**, Mellman I. 2017. Elements of cancer immunity and the cancer-immune set point. *Nature* **541**:321–330. DOI: <https://doi.org/10.1038/nature21349>, PMID: 28102259
- Chen X**, Song E. 2022. The theory of tumor Ecosystem. *Cancer Communications* **42**:587–608. DOI: <https://doi.org/10.1002/cac2.12316>, PMID: 35642770
- Chin SF**, Teschendorff AE, Marioni JC, Wang Y, Barbosa-Morais NL, Thorne NP, Costa JL, Pinder SE, van de Wiel MA, Green AR, Ellis IO, Porter PL, Tavaré S, Brenton JD, Ylstra B, Caldas C. 2007. High-resolution aCGH and expression profiling identifies a novel Genomic subtype of ER negative breast cancer. *Genome Biology* **8**:R215. DOI: <https://doi.org/10.1186/gb-2007-8-10-r215>, PMID: 17925008
- Chiossone L**, Dumas PY, Vienne M, Vivier E. 2018. Natural killer cells and other innate Lymphoid cells in cancer. *Nature Reviews. Immunology* **18**:671–688. DOI: <https://doi.org/10.1038/s41577-018-0077-4>, PMID: 30315207
- Ciardiello F**, Ciardiello D, Martini G, Napolitano S, Tabernero J, Cervantes A. 2022. Clinical management of metastatic colorectal cancer in the era of precision medicine. *CA* **72**:372–401. DOI: <https://doi.org/10.3322/caac.21728>, PMID: 35472088
- Datta J**, Dai X, Bianchi A, De Castro Silva I, Mehra S, Garrido VT, Lamichhane P, Singh SP, Zhou Z, Dosch AR, Messaggio F, Ban Y, Umland O, Hosein PJ, Nagathihalli NS, Merchant NB. 2022. Combined MEK and Stat3 inhibition Uncovers Stromal plasticity by enriching for cancer-associated fibroblasts with Mesenchymal stem cell-like features to overcome Immunotherapy resistance in Pancreatic cancer. *Gastroenterology* **163**:1593–1612. DOI: <https://doi.org/10.1053/j.gastro.2022.07.076>, PMID: 35948109
- Davison JM**, Yee M, Krill-Burger JM, Lyons-Weiler MA, Kelly LA, Sciulli CM, Nason KS, Luketich JD, Michalopoulos GK, LaFramboise WA. 2014. The degree of segmental Aneuploidy measured by total copy number abnormalities predicts survival and recurrence in superficial gastroesophageal adenocarcinoma. *PLOS ONE* **9**:e79079. DOI: <https://doi.org/10.1371/journal.pone.0079079>, PMID: 24454681

- Deng P**, Wang Z, Chen J, Liu S, Yao X, Liu S, Liu L, Yu Z, Huang Y, Xiong Z, Xiao R, Gao J, Liang W, Chen J, Liu H, Hong JH, Chan JY, Guan P, Chen J, Wang Y, et al. 2022. Rad21 amplification Epigenetically suppresses interferon signaling to promote immune evasion in ovarian cancer. *The Journal of Clinical Investigation* **132**:e159628. DOI: <https://doi.org/10.1172/JCI159628>, PMID: 36201246
- De Sousa E Melo F**, Wang X, Jansen M, Fessler E, Trinh A, de Rooij LPMH, de Jong JH, de Boer OJ, van Leersum R, Bijlsma MF, Rodermond H, van der Heijden M, van Noesel CJM, Tuijnman JB, Dekker E, Markowitz F, Medema JP, Vermeulen L. 2013. Poor-prognosis colon cancer is defined by a Molecularly distinct subtype and develops from Serrated precursor lesions. *Nature Medicine* **19**:614–618. DOI: <https://doi.org/10.1038/nm.3174>, PMID: 23584090
- Desgarenes D**, Garrido E, Torres-Gomez MJ, Peña-Cabrales JJ, Partida-Martinez LP. 2014. Diazotrophic potential among bacterial communities associated with wild and cultivated Agave species. *FEMS Microbiology Ecology* **90**:844–857. DOI: <https://doi.org/10.1111/1574-6941.12438>, PMID: 25314594
- Dixon P**. 2003. VEGAN, a package of R functions for community Ecology. *Journal of Vegetation Science* **14**:927–930. DOI: <https://doi.org/10.1111/j.1654-1103.2003.tb02228.x>
- Eide PW**, Bruun J, Lothe RA, Sveen A. 2017. Cmscaller: an R package for consensus molecular Subtyping of colorectal cancer pre-clinical models. *Scientific Reports* **7**:16618. DOI: <https://doi.org/10.1038/s41598-017-16747-x>, PMID: 29192179
- Elia I**, Haigis MC. 2021. Metabolites and the tumour Microenvironment: from cellular mechanisms to systemic metabolism. *Nature Metabolism* **3**:21–32. DOI: <https://doi.org/10.1038/s42255-020-00317-z>, PMID: 33398194
- Fruman DA**, Chiu H, Hopkins BD, Bagrodia S, Cantley LC, Abraham RT. 2017. The Pi3K pathway in human disease. *Cell* **170**:605–635. DOI: <https://doi.org/10.1016/j.cell.2017.07.029>, PMID: 28802037
- Galon J**, Bruni D. 2019. Approaches to treat immune hot, altered and cold tumours with combination Immunotherapies. *Nature Reviews. Drug Discovery* **18**:197–218. DOI: <https://doi.org/10.1038/s41573-018-0007-y>, PMID: 30610226
- Guinney J**, Dienstmann R, Wang X, de Reyniès A, Schlicker A, Song S, Marisa L, Roepman P, Nyamundanda G, Angelino P, Bot BM, Morris JS, Simon IM, Gerster S, Fessler E, De Sousa E Melo F, Missiaglia E, Ramay H, Barras D, Homiczko K, et al. 2015. The consensus molecular subtypes of colorectal cancer. *Nature Medicine* **21**:1350–1356. DOI: <https://doi.org/10.1038/nm.3967>, PMID: 26457759
- Hafemeister C**, Satija R. 2019. Normalization and variance stabilization of single-cell RNA-Seq data using Regularized negative binomial regression. *Genome Biology* **20**:296. DOI: <https://doi.org/10.1186/s13059-019-1874-1>, PMID: 31870423
- Hänzelmann S**, Castelo R, Guinney J. 2013. GSEA: Gene set variation analysis for Microarray and RNA-Seq data. *BMC Bioinformatics* **14**:7. DOI: <https://doi.org/10.1186/1471-2105-14-7>, PMID: 23323831
- Hao Y**, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, Lee MJ, Wilk AJ, Darby C, Zager M, Hoffman P, Stoeckius M, Papalexi E, Mimitou EP, Jain J, Srivastava A, Stuart T, Fleming LM, Yeung B, Rogers AJ, et al. 2021. Integrated analysis of Multimodal single-cell data. *Cell* **184**:3573–3587. DOI: <https://doi.org/10.1016/j.cell.2021.04.048>, PMID: 34062119
- He Y**, Jiang Z, Chen C, Wang X. 2018. Classification of triple-negative breast cancers based on Immunogenomic profiling. *Journal of Experimental & Clinical Cancer Research* **37**:327. DOI: <https://doi.org/10.1186/s13046-018-1002-1>, PMID: 30594216
- Heinemann V**, von Weikersthal LF, Decker T, Kiani A, Vehling-Kaiser U, Al-Batran S-E, Heintges T, Lerchenmüller C, Kahl C, Seipelt G, Kullmann F, Stauch M, Scheithauer W, Hielscher J, Scholz M, Müller S, Link H, Niederle N, Rost A, Höffkes H-G, et al. 2014. FOLFIRI plus Cetuximab versus FOLFIRI plus Bevacizumab as first-line treatment for patients with metastatic colorectal cancer (FIRE-3): a randomised, open-label, phase 3 trial. *The Lancet. Oncology* **15**:1065–1075. DOI: [https://doi.org/10.1016/S1470-2045\(14\)70330-4](https://doi.org/10.1016/S1470-2045(14)70330-4), PMID: 25088940
- Hirata E**, Sahai E. 2017. Tumor Microenvironment and differential responses to therapy. *Cold Spring Harbor Perspectives in Medicine* **7**:a026781. DOI: <https://doi.org/10.1101/cshperspect.a026781>, PMID: 28213438
- Hoshida Y**. 2010. Nearest template prediction: a single-sample-based flexible class prediction with confidence assessment. *PLOS ONE* **5**:e15543. DOI: <https://doi.org/10.1371/journal.pone.0015543>, PMID: 21124904
- Jia Q**, Wang A, Yuan Y, Zhu B, Long H. 2022. Heterogeneity of the tumor immune Microenvironment and its clinical relevance. *Experimental Hematology & Oncology* **11**:24. DOI: <https://doi.org/10.1186/s40164-022-00277-y>, PMID: 35461288
- Jiang P**, Gu S, Pan D, Fu J, Sahu A, Hu X, Li Z, Traugh N, Bu X, Li B, Liu J, Freeman GJ, Brown MA, Wucherpennig KW, Liu XS. 2018. Signatures of T cell dysfunction and exclusion predict cancer Immunotherapy response. *Nature Medicine* **24**:1550–1558. DOI: <https://doi.org/10.1038/s41591-018-0136-1>, PMID: 30127393
- Jiang L**, Wang Y-J, Zhao J, Uehara M, Hou Q, Kasinath V, Ichimura T, Banouni N, Dai L, Li X, Greiner DL, Shultz LD, Zhang X, Sun Z-YJ, Curtin I, Vangos NE, Yeoh ZC, Geffken EA, Seo H-S, Liu Z-X, et al. 2020. Direct tumor killing and Immunotherapy through anti-Serpinb9 therapy. *Cell* **183**:1219–1233. DOI: <https://doi.org/10.1016/j.cell.2020.10.045>, PMID: 33242418
- Jin Z**, Sinicrope FA. 2022. Mismatch repair-deficient colorectal cancer: building on Checkpoint blockade. *Journal of Clinical Oncology* **40**:2735–2750. DOI: <https://doi.org/10.1200/JCO.21.02691>, PMID: 35649217
- Kather JN**, Suarez-Carmona M, Charoentong P, Weis C-A, Hirsch D, Bankhead P, Horning M, Ferber D, Kel I, Herpel E, Schott S, Zörnig I, Utikal J, Marx A, Gaiser T, Brenner H, Chang-Claude J, Hoffmeister M, Jäger D, Halama N. 2018. Topography of cancer-associated immune cells in human solid tumors. *eLife* **7**:e36967. DOI: <https://doi.org/10.7554/eLife.36967>, PMID: 30179157

- Kawazoe A**, Kuboki Y, Shinozaki E, Hara H, Nishina T, Komatsu Y, Yuki S, Wakabayashi M, Nomura S, Sato A, Kuwata T, Kawazu M, Mano H, Togashi Y, Nishikawa H, Yoshino T. 2020. Multicenter phase I/II trial of Napabucasin and Pembrolizumab in patients with metastatic colorectal cancer (Epic1503/SCOOP trial). *Clinical Cancer Research* **26**:5887–5894. DOI: <https://doi.org/10.1158/1078-0432.CCR-20-1803>, PMID: 32694160
- Kaymak I**, Williams KS, Cantor JR, Jones RG. 2021. Immunometabolic interplay in the tumor Microenvironment. *Cancer Cell* **39**:28–37. DOI: <https://doi.org/10.1016/j.ccell.2020.09.004>, PMID: 33125860
- Kirchhammer N**, Trefny MP, Auf der Maur P, Läubli H, Zippelius A. 2022. Combination cancer Immunotherapies: emerging treatment strategies adapted to the tumor Microenvironment. *Science Translational Medicine* **14**:eabo3605. DOI: <https://doi.org/10.1126/scitranslmed.abo3605>, PMID: 36350989
- Lee MS**, Menter DG, Kopetz S. 2017. Right versus left colon cancer biology: integrating the consensus molecular subtypes. *Journal of the National Comprehensive Cancer Network* **15**:411–419. DOI: <https://doi.org/10.6004/jnccn.2017.0038>, PMID: 28275039
- Lee H-O**, Hong Y, Etliglu HE, Cho YB, Pomella V, Van den Bosch B, Vanhecke J, Verbandt S, Hong H, Min J-W, Kim N, Eum HH, Qian J, Boeckx B, Lambrechts D, Tsantoulis P, De Hertogh G, Chung W, Lee T, An M, et al. 2020. Lineage-dependent gene expression programs influence the immune landscape of colorectal cancer. *Nature Genetics* **52**:594–603. DOI: <https://doi.org/10.1038/s41588-020-0636-z>, PMID: 32451460
- Leng D**, Zheng L, Wen Y, Zhang Y, Wu L, Wang J, Wang M, Zhang Z, He S, Bo X. 2022. A benchmark study of deep learning-based multi-Omics data fusion methods for cancer. *Genome Biology* **23**:171. DOI: <https://doi.org/10.1186/s13059-022-02739-2>, PMID: 35945544
- Lenz H-J**, Ou F-S, Venook AP, Hochster HS, Niedzwiecki D, Goldberg RM, Mayer RJ, Bertagnolli MM, Blanke CD, Zemla T, Qu X, Wirapati P, Tejpar S, Innocenti F, Kabbarah O. 2019. Impact of consensus molecular subtype on survival in patients with metastatic colorectal cancer: results from CALGB/SWOG 80405 (alliance). *Journal of Clinical Oncology* **37**:1876–1885. DOI: <https://doi.org/10.1200/JCO.18.02258>, PMID: 31042420
- Lex A**, Gehlenborg N, Strobel H, Vuillemot R, Pfister H. 2014. Upset: visualization of intersecting SETS. *IEEE Transactions on Visualization and Computer Graphics* **20**:1983–1992. DOI: <https://doi.org/10.1109/TVCG.2014.2346248>, PMID: 26356912
- Li C**, Sun Y-D, Yu G-Y, Cui J-R, Lou Z, Zhang H, Huang Y, Bai C-G, Deng L-L, Liu P, Zheng K, Wang Y-H, Wang Q-Q, Li Q-R, Wu Q-Q, Liu Q, Shyr Y, Li Y-X, Chen L-N, Wu J-R, et al. 2020. Integrated Omics of metastatic colorectal cancer. *Cancer Cell* **38**:734–747. DOI: <https://doi.org/10.1016/j.ccell.2020.08.002>, PMID: 32888432
- Linnekamp JF**, Hooff S, Prasetyanti PR, Kandimalla R, Buikhuisen JY, Fessler E, Ramesh P, Lee K, Bochove GGW, de Jong JH, Cameron K, Leersum R, Rodermond HM, Franitza M, Nürnberg P, Mangiapane LR, Wang X, Clevers H, Vermeulen L, Stassi G, et al. 2018. Consensus molecular subtypes of colorectal cancer are Recapitulated in in vitro and in vivo models. *Cell Death and Differentiation* **25**:616–633. DOI: <https://doi.org/10.1038/s41418-017-0011-5>, PMID: 29305587
- Liu T**, Han C, Wang S, Fang P, Ma Z, Xu L, Yin R. 2019. Cancer-associated fibroblasts: an emerging target of anti-cancer Immunotherapy. *Journal of Hematology & Oncology* **12**:86. DOI: <https://doi.org/10.1186/s13045-019-0770-1>, PMID: 31462327
- Liu Y**, Zhang Q, Xing B, Luo N, Gao R, Yu K, Hu X, Bu Z, Peng J, Ren X, Zhang Z. 2022. Immune Phenotypic linkage between colorectal cancer and liver metastasis. *Cancer Cell* **40**:424–437. DOI: <https://doi.org/10.1016/j.ccell.2022.02.013>, PMID: 35303421
- Liu Y**, Xun Z, Ma K, Liang S, Li X, Zhou S, Sun L, Liu Y, Du Y, Guo X, Cui T, Zhou H, Wang J, Yin D, Song R, Zhang S, Cai W, Meng F, Guo H, Zhang B, et al. 2023. Identification of a tumour immune barrier in the HCC Microenvironment that determines the efficacy of Immunotherapy. *Journal of Hepatology* **78**:770–782. DOI: <https://doi.org/10.1016/j.jhep.2023.01.011>, PMID: 36708811
- Luoma AM**, Suo S, Wang Y, Gunasti L, Porter CBM, Nabils N, Tadros J, Ferretti AP, Liao S, Gurer C, Chen Y-H, Criscitiello S, Ricker CA, Dionne D, Rozenblatt-Rosen O, Uppaluri R, Haddad RI, Ashenberg O, Regev A, Van Allen EM, et al. 2022. Tissue-resident memory and circulating T cells are early responders to pre-surgical cancer Immunotherapy. *Cell* **185**:2918–2935. DOI: <https://doi.org/10.1016/j.cell.2022.06.018>, PMID: 35803260
- Marisa L**, de Reyniès A, Duval A, Selves J, Gaub MP, Vescovo L, Etienne-Grimaldi M-C, Schiappa R, Guenot D, Ayadi M, Kirzin S, Chazal M, Fléjou J-F, Benchimol D, Berger A, Lagarde A, Pencreach E, Piard F, Elias D, Parc Y, et al. 2013. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and Prognostic value. *PLOS Medicine* **10**:e1001453. DOI: <https://doi.org/10.1371/journal.pmed.1001453>, PMID: 23700391
- Medico E**, Russo M, Picco G, Cancelliere C, Valtorta E, Corti G, Buscarino M, Isella C, Lamba S, Martinoglio B, Veronese S, Siena S, Sartore-Bianchi A, Beccuti M, Mottolese M, Linnebacher M, Cordero F, Di Nicolantonio F, Bardelli A. 2015. The molecular landscape of colorectal cancer cell lines UNVEILS clinically actionable kinase targets. *Nature Communications* **6**:7002. DOI: <https://doi.org/10.1038/ncomms8002>, PMID: 25926053
- Monti S**, Tamayo P, Mesirov J, Golub T. 2003. Consensus clustering: A Resampling-based method for class discovery and visualization of gene expression Microarray data. *Machine Learning* **52**:91–118. DOI: <https://doi.org/10.1023/A:1023949509487>
- Newman AM**, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA. 2015. Robust enumeration of cell Subsets from tissue expression profiles. *Nature Methods* **12**:453–457. DOI: <https://doi.org/10.1038/nmeth.3337>, PMID: 25822800
- Nguyen EM**, Taniguchi H, Chan JM, Zhan YA, Chen X, Qiu J, de Stanchina E, Allaj V, Shah NS, Uddin F, Manoj P, Liu M, Cai SF, Levine R, Quintanal-Villalonga Á, Sen T, Chow A, Rudin CM. 2022. Targeting Lysine-specific

- demethylase 1 Rescues major Histocompatibility complex class I antigen presentation and overcomes programmed death-ligand 1 blockade resistance in SCLC. *Journal of Thoracic Oncology* **17**:1014–1031. DOI: <https://doi.org/10.1016/j.jtho.2022.05.014>, PMID: 35691495
- Okita A**, Takahashi S, Ouchi K, Inoue M, Watanabe M, Endo M, Honda H, Yamada Y, Ishioka C. 2018. Consensus molecular subtypes classification of colorectal cancer as a predictive factor for chemotherapeutic efficacy against metastatic colorectal cancer. *Oncotarget* **9**:18698–18711. DOI: <https://doi.org/10.18632/oncotarget.24617>, PMID: 29721154
- Oza VH**, Fisher JL, Darji R, Lasseigne BN. 2023. Cinmetrics: an R package for analyzing copy number aberrations as a measure of Chromosomal instability. *PeerJ* **11**:e15244. DOI: <https://doi.org/10.7717/peerj.15244>, PMID: 37123011
- Pandey PR**, Young KH, Kumar D, Jain N. 2022. RNA-mediated Immunotherapy regulating tumor immune Microenvironment: next wave of cancer Therapeutics. *Molecular Cancer* **21**:58. DOI: <https://doi.org/10.1186/s12943-022-01528-6>, PMID: 35189921
- Robertson AG**, Kim J, Al-Ahmadie H, Bellmunt J, Guo G, Cherniack AD, Hinoue T, Laird PW, Hoadley KA, Akbani R, Castro MAA, Gibb EA, Kanchi RS, Gordenin DA, Shukla SA, Sanchez-Vega F, Hansel DE, Czerniak BA, Reuter VE, Su X, et al. 2017. Comprehensive molecular characterization of muscle-invasive bladder cancer. *Cell* **171**:540–556. DOI: <https://doi.org/10.1016/j.cell.2017.09.007>, PMID: 28988769
- Roepman P**, Schlicker A, Tabernero J, Majewski I, Tian S, Moreno V, Snel MH, Chresta CM, Rosenberg R, Nitsche U, Macarulla T, Capella G, Salazar R, Orphanides G, Wessels LFA, Bernards R, Simon IM. 2014. Colorectal cancer intrinsic subtypes predict chemotherapy benefit, deficient mismatch repair and epithelial-to-Mesenchymal transition. *International Journal of Cancer* **134**:552–562. DOI: <https://doi.org/10.1002/ijc.28387>, PMID: 23852808
- Rosenberg JE**, Hoffman-Censits J, Powles T, van der Heijden MS, Balar AV, Necchi A, Dawson N, O'Donnell PH, Balmanoukian A, Loriot Y, Srinivas S, Retz MM, Grivas P, Joseph RW, Galsky MD, Fleming MT, Petrylak DP, Perez-Gracia JL, Burris HA, Castellano D, et al. 2016. Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: a single-arm. *Lancet* **387**:1909–1920. DOI: [https://doi.org/10.1016/S0140-6736\(16\)00561-4](https://doi.org/10.1016/S0140-6736(16)00561-4), PMID: 26952546
- Sadanandam A**, Lyssiotis CA, Homicsko K, Collisson EA, Gibb WJ, Wullschlegel S, Ostos LCG, Lannon WA, Grotzinger C, Del Rio M, Lhermitte B, Olshen AB, Wiedenmann B, Cantley LC, Gray JW, Hanahan D. 2013. A colorectal cancer classification system that Associates cellular phenotype and responses to therapy. *Nature Medicine* **19**:619–625. DOI: <https://doi.org/10.1038/nm.3175>, PMID: 23584089
- Sanchez-Vega F**, Mina M, Armenia J, Chatila WK, Luna A, La KC, Dimitriadou S, Liu DL, Kantheti HS, Saghafinia S, Chakravarty D, Daian F, Gao Q, Bailey MH, Liang W-W, Foltz SM, Shmulevich I, Ding L, Heins Z, Ochoa A, et al. 2018. Oncogenic signaling pathways in the cancer genome Atlas. *Cell* **173**:321–337. DOI: <https://doi.org/10.1016/j.cell.2018.03.035>, PMID: 29625050
- Saxena M**, van der Burg SH, Melief CJM, Bhardwaj N. 2021. Therapeutic cancer vaccines. *Nature Reviews. Cancer* **21**:360–378. DOI: <https://doi.org/10.1038/s41568-021-00346-0>, PMID: 33907315
- Schlicker A**, Beran G, Chresta CM, McWalter G, Pritchard A, Weston S, Runswick S, Davenport S, Heathcote K, Castro DA, Orphanides G, French T, Wessels LFA. 2012. Subtypes of primary colorectal tumors correlate with response to targeted treatment in colorectal cell lines. *BMC Medical Genomics* **5**:66. DOI: <https://doi.org/10.1186/1755-8794-5-66>, PMID: 23272949
- Siegel RL**, Miller KD, Wagle NS, Jemal A. 2023. Cancer statistics, 2023. *CA* **73**:17–48. DOI: <https://doi.org/10.3322/caac.21763>, PMID: 36633525
- Sveen A**, Bruun J, Eide PW, Eilertsen IA, Ramirez L, Murumägi A, Arjama M, Danielsen SA, Kryeziu K, Elez E, Tabernero J, Guinney J, Palmer HG, Nesbakken A, Kallioniemi O, Dienstmann R, Lothe RA. 2018. Colorectal cancer consensus molecular subtypes translated to Preclinical models uncover potentially Targetable cancer cell Dependencies. *Clinical Cancer Research* **24**:794–806. DOI: <https://doi.org/10.1158/1078-0432.CCR-17-1234>, PMID: 29242316
- Takeuchi Y**, Tanegashima T, Sato E, Irie T, Sai A, Itahashi K, Kumagai S, Tada Y, Togashi Y, Koyama S, Akbay EA, Karasaki T, Kataoka K, Funaki S, Shintani Y, Nagatomo I, Kida H, Ishii G, Miyoshi T, Aokage K, et al. 2021. Highly Immunogenic cancer cells require activation of the WNT pathway for immunological escape. *Science Immunology* **6**:eabc6424. DOI: <https://doi.org/10.1126/sciimmunol.abc6424>, PMID: 34767457
- Tang T**, Huang X, Zhang G, Hong Z, Bai X, Liang T. 2021. Advantages of targeting the tumor immune Microenvironment over blocking immune Checkpoint in cancer Immunotherapy. *Signal Transduction and Targeted Therapy* **6**:72. DOI: <https://doi.org/10.1038/s41392-020-00449-4>, PMID: 33608497
- Tavernari D**, Battistello E, Dheilly E, Petruzzella AS, Mina M, Sordet-Dessimoz J, Peters S, Krueger T, Gfeller D, Riggi N, Oricchio E, Letovanec I, Ciriello G. 2021. Nongenetic evolution drives lung adenocarcinoma spatial heterogeneity and progression. *Cancer Discovery* **11**:1490–1507. DOI: <https://doi.org/10.1158/2159-8290.CD-20-1274>, PMID: 33563664
- Thorsson V**, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang T-H, Porta-Pardo E, Gao GF, Plaisier CL, Eddy JA, Ziv E, Culhane AC, Paull EO, Sivakumar IKA, Gentles AJ, Malhotra R, Farshidfar F, Colaprico A, Parker JS, Mose LE, et al. 2018. The immune landscape of cancer. *Immunity* **48**:812–830. DOI: <https://doi.org/10.1016/j.immuni.2018.03.023>, PMID: 29628290
- Thota R**, Yang M, Pflieger L, Schell MJ, Rajan M, Davis TB, Wang H, Presson A, Pledger WJ, Yeatman TJ. 2021. APC and Tp53 mutations predict Cetuximab sensitivity across consensus molecular subtypes. *Cancers* **13**:5394. DOI: <https://doi.org/10.3390/cancers13215394>

- Tibshirani R**, Hastie T, Narasimhan B, Chu G. 2002. Diagnosis of multiple cancer types by shrunken Centroids of gene expression. *PNAS* **99**:6567–6572. DOI: <https://doi.org/10.1073/pnas.082099299>, PMID: 12011421
- Vasaikar S**, Huang C, Wang X, Petyuk VA, Savage SR, Wen B, Dou Y, Zhang Y, Shi Z, Arshad OA, Gritsenko MA, Zimmerman LJ, McDermott JE, Clauss TR, Moore RJ, Zhao R, Monroe ME, Wang Y-T, Chambers MC, Slebos RJC, et al. 2019. Proteogenomic analysis of human colon cancer reveals new therapeutic opportunities. *Cell* **177**:1035–1049. DOI: <https://doi.org/10.1016/j.cell.2019.03.030>, PMID: 31031003
- Vymetalkova V**, Vodicka P, Vodenkova S, Alonso S, Schneider-Stock R. 2019. DNA methylation and Chromatin modifiers in colorectal cancer. *Molecular Aspects of Medicine* **69**:73–92. DOI: <https://doi.org/10.1016/j.mam.2019.04.002>, PMID: 31028771
- Wilkerson MD**, Hayes DN. 2010. Consensusclusterplus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* **26**:1572–1573. DOI: <https://doi.org/10.1093/bioinformatics/btq170>, PMID: 20427518
- Wu Y**, Yang S, Ma J, Chen Z, Song G, Rao D, Cheng Y, Huang S, Liu Y, Jiang S, Liu J, Huang X, Wang X, Qiu S, Xu J, Xi R, Bai F, Zhou J, Fan J, Zhang X, et al. 2022. Spatiotemporal immune landscape of colorectal cancer liver metastasis at single-cell level. *Cancer Discovery* **12**:134–153. DOI: <https://doi.org/10.1158/2159-8290.CD-21-0316>, PMID: 34417225
- Wu X**. 2023a. Pamr\_Classifier. swh:1:rev:c8b87c46eab6c4fc2a528c136ce62e51b9ad71af. Software Heritage. [https://archive.softwareheritage.org/swh:1:dir:95a4d6ea7545906ca3d539264028bc27992704b5;origin=https://github.com/XiangkunWu/PAMR\\_classifier;visit=swh:1:snp:9958e99a51b91d6f8082d985955cdf0592832593;anchor=swh:1:rev:c8b87c46eab6c4fc2a528c136ce62e51b9ad71af](https://archive.softwareheritage.org/swh:1:dir:95a4d6ea7545906ca3d539264028bc27992704b5;origin=https://github.com/XiangkunWu/PAMR_classifier;visit=swh:1:snp:9958e99a51b91d6f8082d985955cdf0592832593;anchor=swh:1:rev:c8b87c46eab6c4fc2a528c136ce62e51b9ad71af)
- Wu X**. 2023b. Pre\_Clinical\_Model. swh:1:rev:5adf0c170407f30df33e15faf19c1091af9f752f. Software Heritage. [https://archive.softwareheritage.org/swh:1:dir:3bb6de7704e96a75adc2191c195c215b0559c11c;origin=https://github.com/XiangkunWu/pre\\_clinical\\_model;visit=swh:1:snp:7c6cee1d3bc38b71641b5ac3e0acaaf7a96ed7;anchor=swh:1:rev:5adf0c170407f30df33e15faf19c1091af9f752f](https://archive.softwareheritage.org/swh:1:dir:3bb6de7704e96a75adc2191c195c215b0559c11c;origin=https://github.com/XiangkunWu/pre_clinical_model;visit=swh:1:snp:7c6cee1d3bc38b71641b5ac3e0acaaf7a96ed7;anchor=swh:1:rev:5adf0c170407f30df33e15faf19c1091af9f752f)
- Wu X**. 2023c. Cccrc\_Classifier. swh:1:rev:8de317a40f03e7272fd6031e12663595d595fa5f. Software Heritage. [https://archive.softwareheritage.org/swh:1:dir:ab578816e920631223538b5fbbf5fab46608d10d;origin=https://github.com/XiangkunWu/CCRC\\_classifier;visit=swh:1:snp:62f1de5d883ba11c4abb1b33df4c406601efd177;anchor=swh:1:rev:8de317a40f03e7272fd6031e12663595d595fa5f](https://archive.softwareheritage.org/swh:1:dir:ab578816e920631223538b5fbbf5fab46608d10d;origin=https://github.com/XiangkunWu/CCRC_classifier;visit=swh:1:snp:62f1de5d883ba11c4abb1b33df4c406601efd177;anchor=swh:1:rev:8de317a40f03e7272fd6031e12663595d595fa5f)
- Xu H**, Liu L, Li W, Zou D, Yu J, Wang L, Wong CC. 2021. Transcription factors in colorectal cancer: molecular mechanism and therapeutic implications. *Oncogene* **40**:1555–1569. DOI: <https://doi.org/10.1038/s41388-020-01587-3>, PMID: 33323976
- Yoshihara K**, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, Treviño V, Shen H, Laird PW, Levine DA, Carter SL, Getz G, Stenke-Hale K, Mills GB, Verhaak RGW. 2013. Inferring tumour purity and Stromal and immune cell Admixture from expression data. *Nature Communications* **4**:2612. DOI: <https://doi.org/10.1038/ncomms3612>, PMID: 24113773
- Yu G**, Wang LG, Han Y, He QY. 2012. clusterProfiler: an R package for comparing biological themes among Gene clusters. *Omics* **16**:284–287. DOI: <https://doi.org/10.1089/omi.2011.0118>, PMID: 22455463
- Zhan T**, Faehling V, Rauscher B, Betge J, Ebert MP, Boutros M. 2021. Multi-Omics integration identifies a selective vulnerability of colorectal cancer subtypes to Ym155. *International Journal of Cancer* **148**:1948–1963. DOI: <https://doi.org/10.1002/ijc.33393>, PMID: 33186476
- Zhang B**, Wang J, Wang X, Zhu J, Liu Q, Shi Z, Chambers MC, Zimmerman LJ, Shaddox KF, Kim S, Davies SR, Wang S, Wang P, Kinsinger CR, Rivers RC, Rodriguez H, Townsend RR, Ellis MJC, Carr SA, Tabb DL, et al. 2014. Proteogenomic characterization of human colon and Rectal cancer. *Nature* **513**:382–387. DOI: <https://doi.org/10.1038/nature13438>, PMID: 25043054
- Zhang L**, Yu X, Zheng L, Zhang Y, Li Y, Fang Q, Gao R, Kang B, Zhang Q, Huang JY, Konno H, Guo X, Ye Y, Gao S, Wang S, Hu X, Ren X, Shen Z, Ouyang W, Zhang Z. 2018. Lineage tracking reveals dynamic relationships of T cells in colorectal cancer. *Nature* **564**:268–272. DOI: <https://doi.org/10.1038/s41586-018-0694-x>, PMID: 30479382
- Zhang J**, Huang D, Saw PE, Song E. 2022. Turning cold tumors hot: from molecular mechanisms to clinical applications. *Trends in Immunology* **43**:523–545. DOI: <https://doi.org/10.1016/j.it.2022.04.010>, PMID: 35624021
- Zhao Q**, Wang F, Chen Y-X, Chen S, Yao Y-C, Zeng Z-L, Jiang T-J, Wang Y-N, Wu C-Y, Jing Y, Huang Y-S, Zhang J, Wang Z-X, He M-M, Pu H-Y, Mai Z-J, Wu Q-N, Long R, Zhang X, Huang T, et al. 2022. Comprehensive profiling of 1015 patients' Exomes reveals Genomic-clinical associations in colorectal cancer. *Nature Communications* **13**:2342. DOI: <https://doi.org/10.1038/s41467-022-30062-8>, PMID: 35487942
- Zhu Z**, Satten GA, Mitchell C, Hu YJ. 2021. Constraining PERMANOVA and LDM to within-set comparisons by projection improves the efficiency of analyses of matched sets of Microbiome data. *Microbiome* **9**:133. DOI: <https://doi.org/10.1186/s40168-021-01034-9>, PMID: 34108046