Immune depletion of the methylated phenotype of colon cancer is closely related to resistance to immune checkpoint inhibitors

Chengqian Zhong[1†], Tingjiang Xie[2†], Long Chen[2†], Xuejing Zhong[3], Xinjing Li[4], Xiumei Cai[1], Kaihong Chen[5*] and Shiqian Lan[1*]

[1]Department of Digestive Endoscopy center, Longyan First Affiliated Hospital of Fujian Medical University, Longyan, China, [2]Department of Gastrointestinal Surgery, Longyan First Affiliated Hospital of Fujian Medical University, Longyan, China, [3]Department of Science and Education, Longyan First Affiliated Hospital of Fujian Medical University, Longyan, China, [4]Department of Pathology, Longyan First Affiliated Hospital of Fujian Medical University, Longyan, China, [5]Department of Cardiology, Longyan First Affiliated Hospital of Fujian Medical University, Longyan, China

Background: Molecular typing based on single omics data has its limitations and requires effective integration of multiple omics data for tumor typing of colorectal cancer (CRC).

Methods: Transcriptome expression, DNA methylation, somatic mutation, clinicopathological information, and copy number variation were retrieved from TCGA, UCSC Xena, cBioPortal, FireBrowse, or GEO. After pre-processing and calculating the clustering prediction index (CPI) with gap statistics, integrative clustering analysis was conducted *via* MOVICS. The tumor microenvironment (TME) was deconvolved using several algorithms such as GSVA, MCPcounter, ESTIMATE, and PCA. The metabolism-relevant pathways were extracted through ssGSEA. Differential analysis was based on limma and enrichment analysis was carried out by Enrichr. DNA methylation and transcriptome expression were integrated *via* ELMER. Finally, nearest template or hemotherapeutic sensitivity prediction was conducted using NTP or pRRophetic.

Results: Three molecular subtypes (CS1, CS2, and CS3) were recognized by integrating transcriptome, DNA methylation, and driver mutations. CRC patients in CS3 had the most favorable prognosis. A total of 90 differentially mutated genes among the three CSs were obtained, and CS3 displayed the highest tumor mutation burden (TMB), while significant instability across the entire chromosome was observed in the CS2 group. A total of 30 upregulated mRNAs served as classifiers were identified and the similar diversity in clinical outcomes of CS3 was validated in four external datasets. The heterogeneity in the TME and metabolism-related pathways were also observed in the three

CSs. Furthermore, we found CS2 tended to loss methylations while CS3 tended to gain methylations. Univariate and multivariate Cox regression revealed that the subtypes were independent prognostic factors. For the drug sensitivity analysis, we found patients in CS2 were more sensitive to ABT.263, NSC.87877, BIRB.0796, and PAC.1. By Integrating with the DNA mutation and RNA expression in CS3, we identified that SOX9, a specific marker of CS3, was higher in the tumor than tumor adjacent by IHC in the in-house cohort and public cohort.

Conclusion: The molecular subtypes based on integrated multi-omics uncovered new insights into the prognosis, mechanisms, and clinical therapeutic targets for CRC.

Introduction

Colorectal cancer (CRC) is the third most common malignant tumor in the world and the fourth major cause of cancer death (1). The diagnosis of CRC is often in the middle and late stages with poor prognosis, and distant metastasis is the main cause of death in colorectal cancer patients. With the continuous improvement in medical level, comprehensive treatment measures such as surgery, radiotherapy, and chemotherapy, targeted therapy, and immunotherapy have improved the overall survival (OS) of patients with CRC, but their overall efficacy is still poor, and the 5-year survival rate of patients with metastatic CRC is only about 14% (1). Therefore, how to effectively evaluate the prognosis of different CRC patients is an urgent problem to be solved.

At present, the most widely used prognostic staging system for CRC is the TNM (Tumor, Node, and Metastasis) staging system, which is easy to observe from clinical information and is the benchmark for the establishment of clinical treatment plans

---

Abbreviations: CRC, Colorectal Cancer; TME, Tumor Microenvironment; TMB, Tumor Mutation Burden; OS, Overall Survival; MSI, Microsatellite Instability; MMR, Mismatch Repair; SNV, Single Nucleotide Variation; TCGA, The Cancer Genome Atlas; COAD, Colon Adenocarcinoma; READ, Rectum Adenocarcinoma; FPKM, Fragments Per Kilobase Million; CNV, Copy Number Variation; RMA, Robust Multichip Average; TPM, Transcripts Per Kilobase Million; PCA, Principal Component Analysis; CPI, Clustering Prediction Index; GSVA, Gene Set Variation Analysis; MeTIL, Methylation of Tumour-Infiltrating Lymphocyte; GSEA, Gene Set Enrichment Analysis; ssGSEA, single-sample GSEA; MSigDB, Molecular Signatures Database; DMP, Differentially Methylated Probe; DPI, Data Processing Inequality; CS, Cancer Subtype; PFS, Progression-Free Survival; CIRC, Co-ordinate Immune Response Cluster.

for patients. However, the TNM staging system mainly relies on expert opinions, and the features used are relatively single. The abnormal phenomenon of the TNM staging system in CRC (the prognosis of patients at stage IIB/C is significantly worse than that of patients at stage III A) results in its limited ability of personalized and accurate clinical decision (2). At the same time, as a population-based system, the TNM staging system has been questioned about its application to individual patients (3). The latest eighth edition of the TNM staging system included biomarkers as new prognostic factors in some cancer staging (3). Therefore, it is necessary to introduce new prognostic factors to the existing TNM staging system in order to more accurately assess the prognosis of patients and formulate treatment plans.

Cancer is a complex disease with high heterogeneity, even patients with the same histopathological classification will have different gene mutations (4). Hence, personalized prevention, diagnosis, and treatment should be done according to the clinical and omics characteristics of different patients (5). For CRC, microsatellite instability (MSI), DNA mismatch repair (MMR), and the results of molecular tests such as RAS mutation and BRAF VE6000 are used to determine the prognosis (3, 6). It is possible to combine clinical and omics information for more personalized prognostic analysis of cancer, but it is difficult for a single omics data to fully account for all factors in a complex disease such as cancer, making it difficult for researchers to derive data from millions of single-nucleotide variations (SNV) to find the key gene that actually causes the disease (7). In recent years, more and more researchers have carried out integrated analysis of various omics data and obtained certain results (8, 9). However, most prognostic studies of CRC are limited to one set of omics, such as gene expression (10) or DNA methylation (11), and few studies that consider multiple omics data have failed to effectively combine multiple omics data with clinical data (12).

Therefore, how to integrate clinical data and omics data and apply them to the prognosis of CRC is of great significance.

The Cancer Genome Atlas (TCGA) is a platform that integrates clinical data, survival information, and multiple omics data for 33 cancers. Through the integration and analysis of multiple omics, cancer subtype classification, biomarker discovery, and survival prognosis analysis can be carried out (13–15). Herein, using data from TCGA and other public databases, we developed a classifier based on multi-omics integration for the prognosis prediction of CRC for the first time. We evaluated the differences in genomic heterogeneity, transcriptome biomarkers, TME landscape, metabolism-related pathways, epigenetic regulation, and chemotherapeutic drug sensitivity among the molecular subtypes of CRC. Multivariate Cox regression analysis confirmed the independent prognostic value of our subtype system. In summary, the molecular subtypes based on integrated multi-omics uncovered new insights into the prognosis, mechanisms, and clinical therapeutic targets for patients with CRC.

Materials and methods

Study population

Molecular data of patients diagnosed with CRC were retrieved from TCGA (13). Transcriptome expression profiles of the TCGA-COAD (colon adenocarcinoma) and TCGA-READ (rectum adenocarcinoma) projects quantified by the number of fragments per kilobase million (FPKM) were downloaded from the UCSC Xena (https://xenabrowser.net/), including 616 fresh-frozen samples with primary malignancy and 51 adjacent normal samples. The DNA methylation profile quantified by Illumina HumanMethylation 450K-array platform was downloaded from the UCSC Xena (https://xenabrowser.net/) under the projects of TCGA-COAD and TCGA-READ, respectively, including a total of 387 primary colorectal tumour samples and 45 adjacent normal samples. Somatic mutation data, patients' clinicopathological information, and survival data were retrieved from cBioPortal (http://www.cbioportal.org/datasets) (16). Copy number variation (CNV) data was collected from FireBrowse (http://firebrowse.org/) (17). For the purpose of multi-omics integrative clustering, 306 primary colorectal tumour samples with available transcriptome expression, DNA methylation, and somatic mutation profiles were identified for this study. Another four independent cohorts downloaded from GEO, including GSE14333 (18), GSE17538 (19), GSE38832 (20), and GSE39582 (21), comprised of a total of 1,159 CRCs with gene expression matrix and corresponding clinicopathological information. Of these external validation cohorts, gene expression matrices were profiled by Affymetrix Human Genome U133 Plus 2.0 Array. The Robust Multichip Average (RMA) algorithm was used for background correction and normalization (22).

Data pre-processing for gene expression and DNA methylation profiles

For the FPKM data of high-throughput sequencing from TCGA, Ensembl IDs for mRNAs were transformed to gene symbols by GENCODE 22. The FPKM values were transferred into transcripts per kilobase million (TPM) values, which showed more similarity to those derived from microarray and more comparable between samples (23). For microarray data retrieved from GEO database, we performed RMA normalization and processing using default settings for background correction and normalization by R package affy (24). Affymetrix probe ID was annotated with gene symbols according to the GPL570 platform. For multiple probes that mapped to one gene, mean value of expression was considered. The potential cross-dataset batch effect was removed under an empirical Bayes framework, namely, ComBat, by the R package sva (25), and the batch effect was further investigated using principal component analysis (PCA) for transcriptome profiles. For DNA methylation, we performed logit transforms β-values before ComBat adjustment and then computed the reverse logit transformation following the ComBat adjustment (26). Subsequently, we used R package ChAMP to comprehensively filter the methylation matrix. To be specific, probes with detection P value > 0.01, probes with <3 beads in at least 5% of samples per probe, all non-CpG probes, all SNP-related probes, all multi-hit probes, and probes located on sex chromosomes were removed in the first place (26, 27).

Integrative clustering based on multi-omics profiles

To perform integrative clustering analysis, we processed the TCGA multi-omics data sets to form three data matrices with columns corresponding to the common samples (n = 306) and rows corresponding to the omics features. The transcriptome expression profile was first $\log_2$ transformed. For the methylation data, we extracted probes located in promoter CpG islands, and for genes having more than one probe mapping to its promoter, the median β value was considered to identify 10,263 methylated genes. For the mutation matrix, a gene was considered mutated (entry of 1) if it contained at least one type of the following nonsynonymous variations: frameshift deletion/insertion, in-frame deletion/insertion, missense/nonsense/nonstop mutation, splice site or translation start site mutation; otherwise, 0 was used to designate wild-type status. To better fit the model and accelerate the clustering efficiency, features with flat values were removed. Specifically, we used the top 1,500 most variable mRNAs, and methylation genes according to the median absolute deviation. Additionally, 20 genes that were previously identified as driver mutations for

colorectal carcinoma were selected for cancer subtyping (28). To find an optimal clustering number, we calculated the clustering prediction index (CPI) and gap statistics using R package MOVICS (29). Consequently, integrative clustering of the TCGA cohort was conducted by R package MOVICS using a Bayesian latent variable model (29, 30).

Deconvolution of tumour microenvironment

To estimate the cell abundance of TME, we retrieved from the previous study a compendium of microenvironment genes related to specific microenvironment cell subsets, which consisted of 364 genes representing 24 microenvironment cell types (31, 32). We then used gene set variation analysis (GSVA) on these gene sets to generate enrichment scores for each cell using the R package GSVA (33). Additionally, quantification of the absolute abundance of eight immune and two stromal cell populations in heterogeneous tissues from transcriptomic data was conducted by the R package MCPcounter (34). The presence of infiltrating immune/stromal cells in the tumour tissue was estimated by the R package ESTIMATE (35). Additionally, the individual DNA methylation of tumour-infiltrating lymphocyte (MeTIL) score in the TCGA cohort was calculated using PCA according to the protocols described in the literature (36).

Single sample enrichment for metabolism-relevant pathways

The 115 metabolism-relevant gene signatures were achieved from previously published study (37), and were quantified by using single-sample GSEA (ssGSEA) approach through R package GSVA (38). Specifically, we extracted three main categories of these metabolism-relevant pathways, including carbohydrate metabolism, amino acid metabolism, and lipid metabolism.

Differential analysis and functional enrichment

Differential expression analyses were conducted using the R package "limma" (39). Gene set enrichment analysis (GSEA) was performed based on pre-ranked gene list according to the descending ordered $\log_2$FoldChange value derived from differential expression analysis; we then leveraged R package clusterProfiler to determine functional enrichment based on Hallmark gene set background that was retrieved from Molecular Signatures Database (MSigDB) (40, 41). The

differentially methylated probes (DMPs) were obtained by R package ChAMP (26). Specifically, we considered probe to have significantly gained methylation if its corresponding mean β-value was greater than 0.3 in the specific subtype but less than 0.2 in the reference subtype with $P<0.05$ and FDR<0.05; vice versa for probes that significantly lost methylation. Gene-list based enrichment analysis was conducted by an integrative and collaborative website tool (Enrichr; https://maayanlab.cloud/Enrichr/) (42).

Cancer subtype characterization and visualization

As previously developed R package MOVICS provides powerful functions to comprehensively characterize cancer subtypes and create feature rich customizable visualizations with minimal effort, we therefore characterized the identified colorectal subtypes from multiple aspects, including survival rate, mutational frequency, fraction of copy number-altered genome (FGA), and clinical characteristics. All parameters were set to default values (29).

Integrative analysis of DNA methylation and transcriptome expression

We used R package ELMER to investigate the crosstalk between DNA methylation and transcriptome expression under an integrative analytic pipeline (43). For probes that are located in promoters, we identified putative genes that were significantly downregulated due to the hypermethylation of promoter probes. Next, the closest 20 upstream and downstream genes were collected for each probe, and for each candidate probe-gene pair, the Mann-Whitney U test was harnessed to test the null hypothesis that overall gene expression in the specific group was less than or equal to that in the reference group. For probes that are located in enhancers (distal probes that are at least 2Kb away from transcription start site on human chromosomes), hypomethylated enhancer mode with overexpressed gene expression pattern was investigated accordingly.

Nearest template prediction

Gene-expression signature-based classification was conducted using NTP algorithm, which provided a convenient model-free approach to make category prediction at single-sample level using only a list of signature genes and a test dataset, which was flexible and beneficial in external cohort application (44, 45).

Analysis of regulons

Transcriptional regulatory networks (regulons) were constructed for 71 candidate regulators associated with cancerous chromatin remodelling (46). As described in the previous study (31), potential associations between a regulator and all possible target genes were revealed by mutual information and Spearman's correlation, and associations were dropped *via* permutation analysis if the corresponding FDR was greater than 0.00001. Unstable associations were also eliminated through bootstrapping (1,000 re-samplings, consensus bootstrap>95%), and the weakest associations were removed by data processing inequality (DPI) filtering embedded in the R package RTN (47). Regulon activity scores for all samples were calculated by two-tailed GSEA.

Therapeutic response analysis

We employed R package pRRophetic to predict the chemotherapeutic sensitivity for each colorectal sample using the parameters by default (48, 49). For immunotherapy, we retrieved a published data set consisting of 47 patients with melanoma who responded to anti-CTLA4 or anti-PD1 blockades (50), and then harnessed subclass mapping to predict the clinical response to immune checkpoint blockade (51).

Immunohistochemical staining

The 50 pairs of CRC tumor and adjacent normal tissue Microarray (D216Re01) were purchased from Xi'an bioaitech Co., Ltd (Xi'an, China). Immunohistochemical staining was performed on normal and the paired tumor tissue slides. The slides were incubated with rabbit polyclonalanti-SOX9 (EPR14335, 1:2000); antibodies at 4℃ overnight. SOX9 expression was evaluated by using a system considering the staining intensity (0 means negative 1 means weak; 2 means moderate; and 3 means strong) and the percentage of positively stained cells (<5%=05% to <25%=1, 25% to 50%=2, >50 to <75%=3, >75%=4). The final score was calculated by multiplying the extent score by the intensity score.

Statistical analyses

All statistical analyses were performed by R (Version 4.0.2). We used Fisher's exact test for categorical data, Kruskal–Wallis one-way analysis of variance for continuous data, a log-rank test for Kaplan-Meier curve, and Cox regression for hazard ratio. For all comparisons, a two-sided $P < 0.05$ was considered statistically significant.

Results

Multi-omics integrative molecular subtype of colorectal cancer
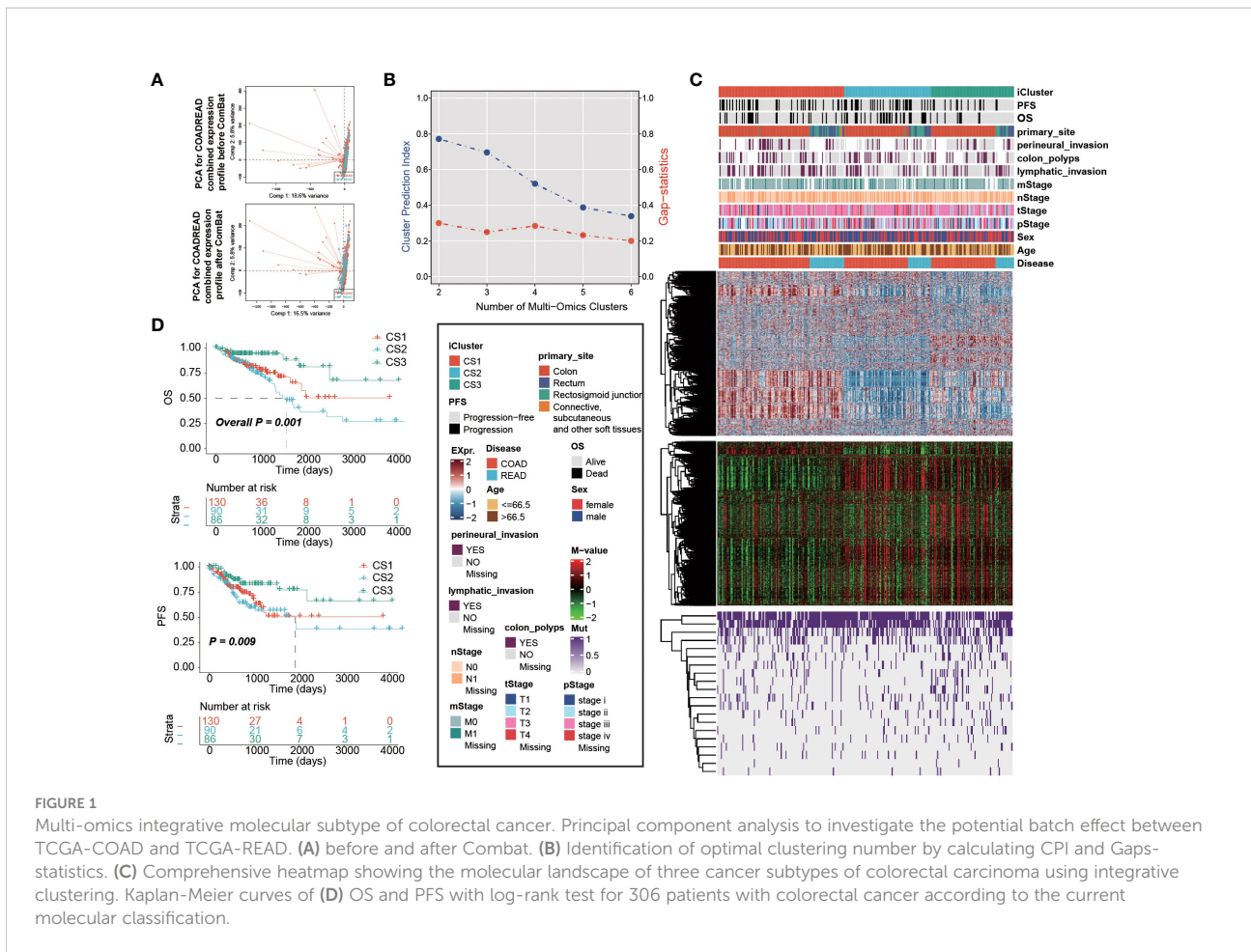
We combined expression profiles of TCGA-COAD and TCGA-READ, and further removed the potential batch effect (Figure 1A). We determined the optimal cluster number of three taking into account two clustering statistics and previous molecular classifications (Figure 1B). Subsequently, integrative clustering identified three robust cancer subtypes (CSs), which were characterized by distinct molecular patterns across transcriptome mRNA expression, DNA methylation and colorectal cancerous driver mutations (Figure 1C). Of note, these classifications were not associated with major clinical features (all $P > 0.05$; Supplementary Table S1); our classification system was tightly associated with overall survival rate (OS; $P = 0.001$; Figure 1D) and progression-free survival rate (PFS; $P = 0.009$);. Generally, CS3 showed the most favourable prognosis among three clusters.

Genomic heterogeneity of colorectal cancer subtype

To investigate the genomic heterogeneity of these molecular subtypes further, we investigate the differentially mutated genes among our classifications, leading to a total of 90 genes (FDR < 0.05 and mutational frequency > 10%; Figure 2A). Among these 90 genes, 11 genes were previously identified as driver mutations in colorectal cancer, including *PIK3CA*, *APC*, *BRAF*, *KRAS*, *TP53*, *FBXW7*, *AMER1*, *TCF7L2*, *SOX9*, *ARID1A*, and *SMAD4* (Supplementary Table S2). Additionally, we found that CS3 showed a significantly higher tumour mutation burden (TMB, $P = 0.002$; Figure 2B) than the other two subtypes. We then investigated chromosomal instability by calculating the FGA scores and found that CS2 had significant instability across the entire chromosome as compared to the other two subtypes with significantly higher copy number loss or gain ($P < 0.001$; Figure 2C). We showed three types distinguishing composite copy number profiles: gistic score (Figure 2D), and percentage/frequency.

Identification of transcriptome biomarkers for colorectal cancer subtype

Given that transcriptome-level data were the most commonly used molecular profiles in cancer research, we identified 30 mRNAs with uniquely and significantly upregulated expression as classifiers for each subtype in the TCGA cohort, and a 90-gene signature was generated (Figure 3F; Supplementary Table S3). To

FIGURE 1

Multi-omics integrative molecular subtype of colorectal cancer. Principal component analysis to investigate the potential batch effect between TCGA-COAD and TCGA-READ. **(A)** before and after Combat. **(B)** Identification of optimal clustering number by calculating CPI and Gaps-statistics. **(C)** Comprehensive heatmap showing the molecular landscape of three cancer subtypes of colorectal carcinoma using integrative clustering. Kaplan-Meier curves of **(D)** OS and PFS with log-rank test for 306 patients with colorectal cancer according to the current molecular classification.

test the reproducibility of our identified colorectal molecular subtypes, we combined four external datasets as GEO cohort of which expression profiles were measured by microarray platform; batch effect across different datasets were removed (Figures 3A, B). We then predict the identified molecular subtypes in the GEO cohort (n = 1,159) using NTP algorithm, which classified each sample in the GEO cohort as one of the identified CS (Figure 3C). Of note, a total of 961 cases of GEO cohort were predicted with confidence (FDR < 0.05) and those cases were used for the downstream analyses. Likewise, CS3 presented with the most favourable clinical outcome out of the three subtypes (*P* = 0.008; Figures 3D, E).

Delineation of metabolism-related pathways in colorectal cancer subtype

Oncogenic heatmap with cancer associated mutations in tcga coadread (Figure 4A). Boxplot for oncogenetic pathways in iclusters of tcga coadread(Figure 4B). Similarly, GSEA is run for each subtype based on its corresponding DEA result to identify subtype-specific functional pathways (Figures 4C, D). Since

Metabolic pathways regulate colorectal cancer initiation and progression, we further explored whether distinct subtypes had different metabolic characteristics in both TCGA and GEO cohort (Figure 4E). Of note, we found global dysfunction of metabolism-related pathways among three molecular subtypes, and generally CS3 showed relatively higher enrichment level of carbohydrate, amino acid, and lipid metabolism-relevant pathways, which may suggest that these colorectal cancers preserved the default metabolic program of normal colon and rectum, leading to a generally good clinical outcome.

Tumour microenvironment landscape of colorectal cancer subtype

Since cancer immunity plays a critical role in tumour progression, we suspected that the tumour microenvironment may vary a lot among these molecular subtypes. Since cancer immunity plays a critical role in tumour progression, we suspected that the tumour microenvironment may vary a lot among these molecular subtypes. Therefore, we investigated the specific immune cell infiltration status of samples in the TCGA
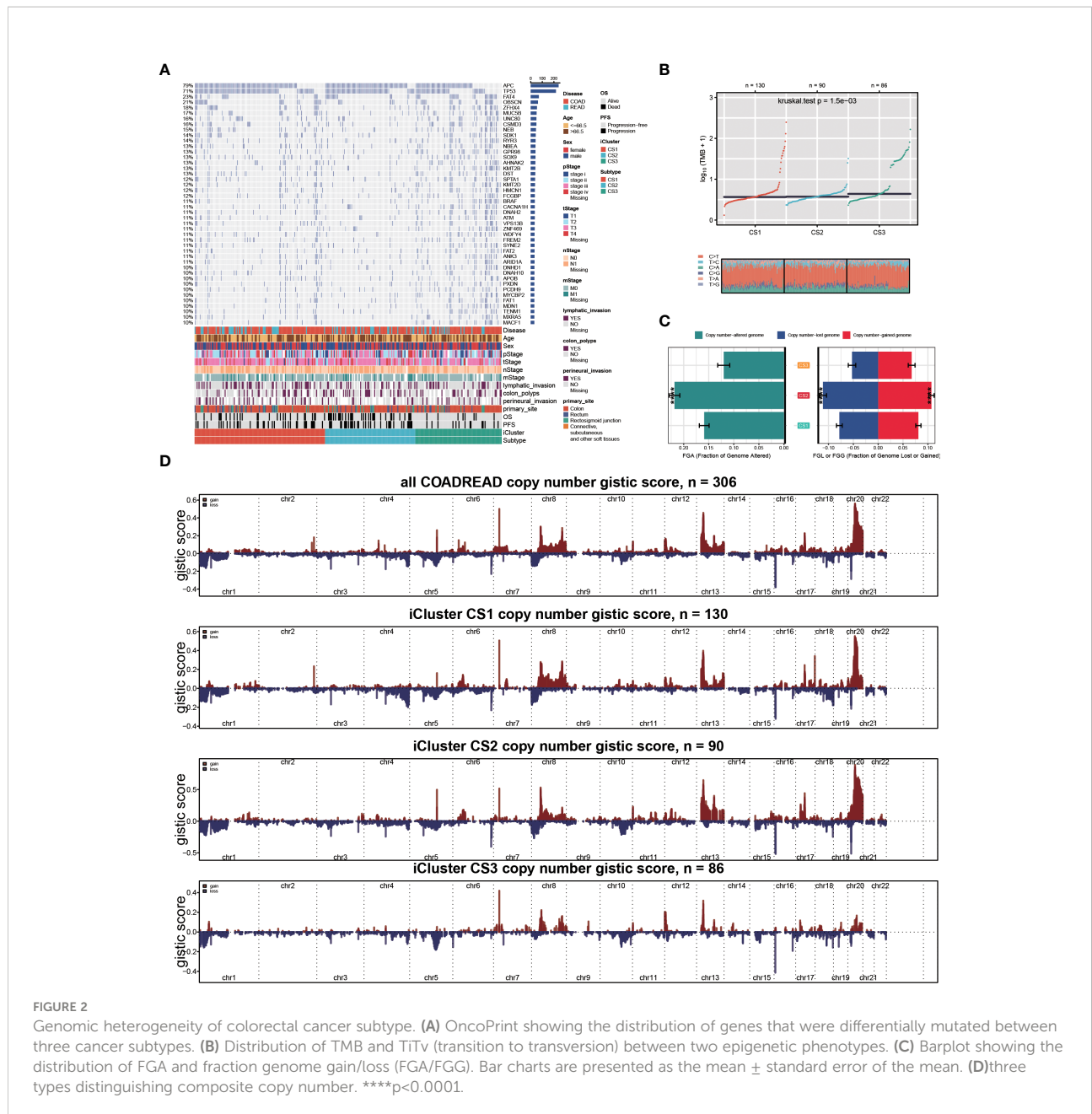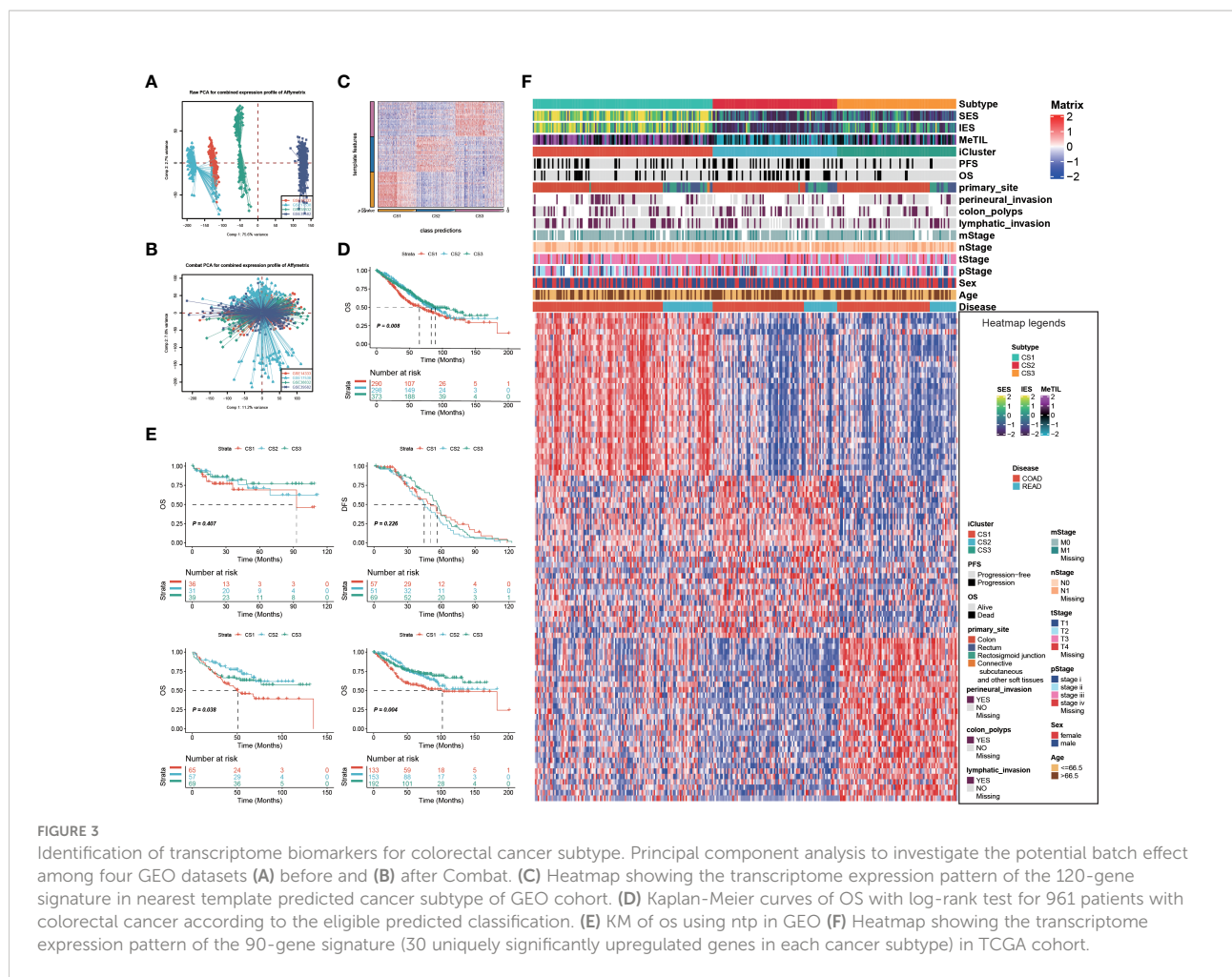
FIGURE 2
Genomic heterogeneity of colorectal cancer subtype. **(A)** OncoPrint showing the distribution of genes that were differentially mutated between three cancer subtypes. **(B)** Distribution of TMB and TiTv (transition to transversion) between two epigenetic phenotypes. **(C)** Barplot showing the distribution of FGA and fraction genome gain/loss (FGA/FGG). Bar charts are presented as the mean $\pm$ standard error of the mean. **(D)**three types distinguishing composite copy number. ****p<0.0001.

cohort. To be specific, we quantified the infiltration levels of several microenvironment cell types using different approach, and surveyed the colorectal samples for the expression of genes representing immune checkpoint targets. The analysis of gene expression signatures suggested that CS1 was highly immune-infiltrated, CS3 showed relatively higher immunocyte infiltration, while CS2 was generally immune-depleted (Figure 5A). This finding may converge to the poor overall survival of CS2 versus other molecular subtypes. Compared to the other subtypes, CS1 had relatively higher expression of several genes that represent potential targets for immunotherapy, including CD274 (PDL1), PDCD1 (PD1),

CD247 (CD3), PDCD1LG2 (PDL2), CTLA4 (CD152), TNFRSF9 (CD137), TNFRSF4 (CD134) and TLR9 (Sup_S2). Interestingly, CS1 enriched for B cell, CD8 T cells but may lack CD4 memory activated cells (Sup_S2); previous study showed the ratio of CD4/CD8 may play prognostic role in several cancer subtypes (52, 53). Additionally, we found that interferon-γ pathway was significantly activated in CS1 (FDR < 0.001; Figure 5B), which made us hypothesized that CS1 may be beneficial from immune checkpoint inhibitors. In this manner, we performed subclass mapping of TCGA cohort and revealed that only the CS1 showed high transcriptome-level similarity to a group of patients with melanoma who responded to anti-

FIGURE 3

Identification of transcriptome biomarkers for colorectal cancer subtype. Principal component analysis to investigate the potential batch effect among four GEO datasets **(A)** before and **(B)** after Combat. **(C)** Heatmap showing the transcriptome expression pattern of the 120-gene signature in nearest template predicted cancer subtype of GEO cohort. **(D)** Kaplan-Meier curves of OS with log-rank test for 961 patients with colorectal cancer according to the eligible predicted classification. **(E)** KM of os using ntp in GEO **(F)** Heatmap showing the transcriptome expression pattern of the 90-gene signature (30 uniquely significantly upregulated genes in each cancer subtype) in TCGA cohort.

CTLA4 or anti-PD1 blockades ($P < 0.05$, adjusted $P \le 0.25$; Figure 5C), which indicated that the current classification may be useful to identify ideal candidates of patients with colorectal cancer for immunotherapy. The tumour microenvironment landscape was generally validated in GEO cohort. Consistently, CS1 in GEO cohort significantly activated interferon-γ pathway, and showed higher likelihood of responding to immune checkpoint inhibitors.

Epigenetic regulation in colorectal cancer subtype

Given the different transcription profiles among the three CRC subtypes, we then asked if this could mirror the epigenetic aspect. To this end, we identified differentially methylated probes for each subtype, and we found CS2 tended to loss methylations (n = 240) as compared to other subtypes (Supplementary Table S4). Notably, these probes losing DNA methylation were significantly enriched in enhancers compared to the 450K array background ($P<0.001$; Figure 6A). As to CS3,

we found this subtype tended to gain methylations (n=249) compared to other subtypes (Supplementary Table S5), and those probes gaining methylation significantly enriched in promoter CpG islands ($P<0.001$; Figure 6B). To further investigate the crosstalk between epigenetic DNA methylation and transcriptome expression, we performed integrative analysis combining both gene expression and DNA methylation profiles using ELMER pipeline. First, for CS2, we identified distal probes that are 2Kb away from the transcription start site of the human chromosome, and performed differential methylation analysis at probe level to identify probes with difference of β-value greater than 0.1 (FDR<0.05) in CS2 compared to other subtypes, ending up with a total of 3,683 distal probes/enhancers (Supplementary Table S6). Next, ELMER searched for the nearby 20 genes corresponding to these probes, and further predicted enhancer-gene linkages using associations between DNA methylation at enhancers and expression of 20 nearby genes of the CpG sites; such analysis identified a total of 2,533 pairs corresponding to 1,003 genes (Figure 6C; Supplementary Table S7). To understand the biologic relevance of these genes that were epigenetically activated, we harnessed Enrichr and found
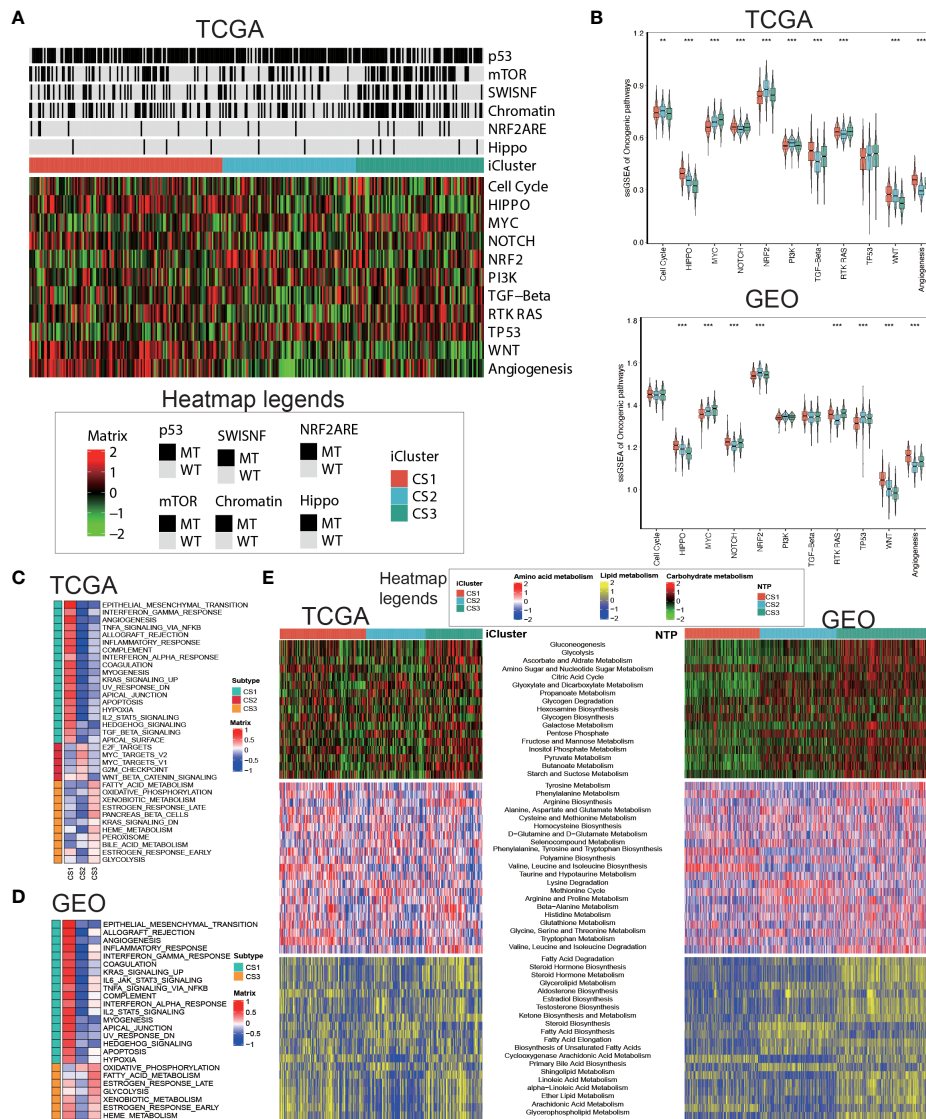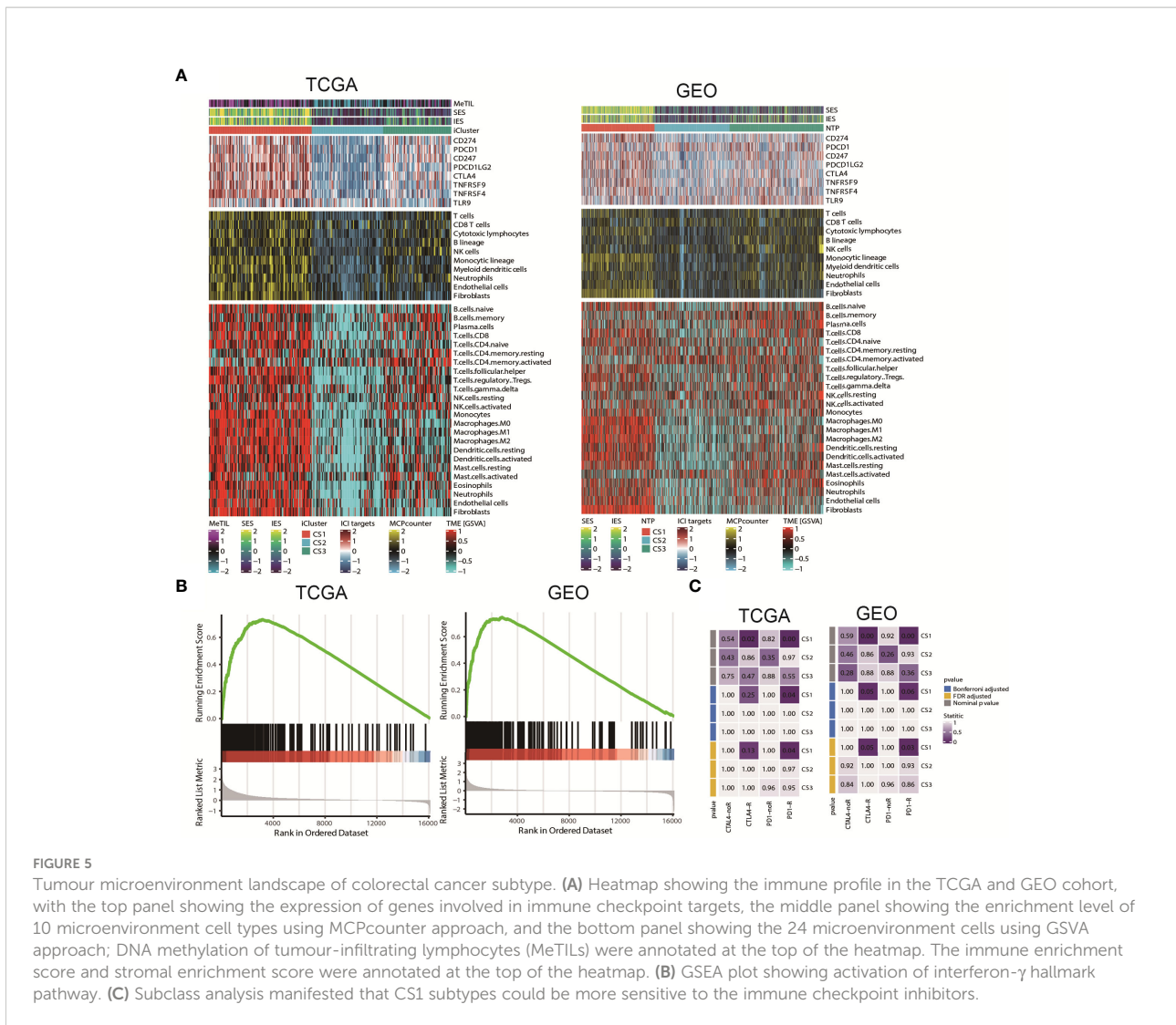
FIGURE 4

Delineation of metabolism-related pathways in colorectal cancer subtype. **(A)** Oncogenic heatmap with cancer associated mutations in tcga coadread. **(B)** Boxplot for oncogenetic pathways in iclusters of tcga coadread. **(C)** Upregulated hallmark pathway heatmap in tcga_using_upregulated_pathways. **(D)** Upregulated hallmark pathway heatmap in geo_using_upregulated_pathways. **(E)** Heatmap showing transcriptome enrichment score of three metabolic categories in TCGA and GEO cohorts. **p < 0.01; ***p < 0.001.

that these genes were significantly enriched in MYC Hallmark pathways (P =0.006, FDR=0.24; Supplementary Table S8). Previous study demonstrated that MYC oncogene was associated the suppression in tumour immunity (54), which suggest that the downregulation of MYC oncogenic pathway may contribute shaping the immune-depleted tumour microenvironment of CS2. Using the similar strategy, we investigated CS3, but we searched for promoter-gene pairs that showed epigenetically silencing mode. In this manner, ELMER identified a total of 1,063 promoters that gained methylation in CS3 versus other subtypes (Supplementary Table S9), and a total

of 3,212 promoter-gene pairs were identified to be epigenetically silenced in CS3 (Figure 6D; Supplementary Table S10). Enrichr revealed that these genes are significantly enriched in epithelia-mesenchymal transition (EMT) hallmark pathway ($P<0.001$, FDR<0.001; Supplementary Table S11). Down-regulation of EMT may decrease tumour-initiating and metastatic potential of cancer cells (55), which lead to good prognosis of CS3. In addition, activity profiles of regulons associated with chromatin remodelling highlighted additional potential regulatory differences among three colorectal cancer subtypes, indicating that epigenetically driven transcriptional networks might be

FIGURE 5
Tumour microenvironment landscape of colorectal cancer subtype. **(A)** Heatmap showing the immune profile in the TCGA and GEO cohort, with the top panel showing the expression of genes involved in immune checkpoint targets, the middle panel showing the enrichment level of 10 microenvironment cell types using MCPcounter approach, and the bottom panel showing the 24 microenvironment cells using GSVA approach; DNA methylation of tumour-infiltrating lymphocytes (MeTILs) were annotated at the top of the heatmap. The immune enrichment score and stromal enrichment score were annotated at the top of the heatmap. **(B)** GSEA plot showing activation of interferon-γ hallmark pathway. **(C)** Subclass analysis manifested that CS1 subtypes could be more sensitive to the immune checkpoint inhibitors.

important differentiators of these molecular subtypes (Figures 6E, F).

Independent prognostic value of colorectal cancer subtype

We then surveyed that whether the current classification was an independent prognostic factor in colorectal cancers from TCGA cohort. As the generally favourable prognosis of CS3, we therefore considered the CS3 as the non-aggressive subtype while patients belonged to CS1 or CS2 were aggressive in clinical setting. In this manner, univariate Cox regression model was first conducted to filter out prognostic clinical characterizations concerning OS and PFS; multivariate Cox regression was subsequently performed based on those prognosis-relevant features. Using such strategy, we found that the current classification remained the independent prognostic

factor after adjusting clinical prognostic features with respect to OS ($P = 0.026$) and PFS ($P = 0.032$) (Figures 7A, B).

Potential therapeutic strategy for colorectal cancer subtype

Considering the significantly poor clinical outcome of CS2 in colorectal cancer, we decided to infer potential anticancer agents that may show clinical efficiency for patients belonging to CS2 through an *in-sillico* drug screening approach. To this end, we constructed ridge regression model between cell lines and corresponding drug sensitivity and applied the predictive model to each of the colorectal cases in both TCGA and GEO cohorts (Supplementary Tables 12, 13). A total of four drugs were discovered to be potentially effective in treating patients with CS2 phenotype as compared to other cases, including ABT.263, NSC.87877, BIRB.0796, and PAC.1 (all, $P < 0.01$; Figures 7C, D).
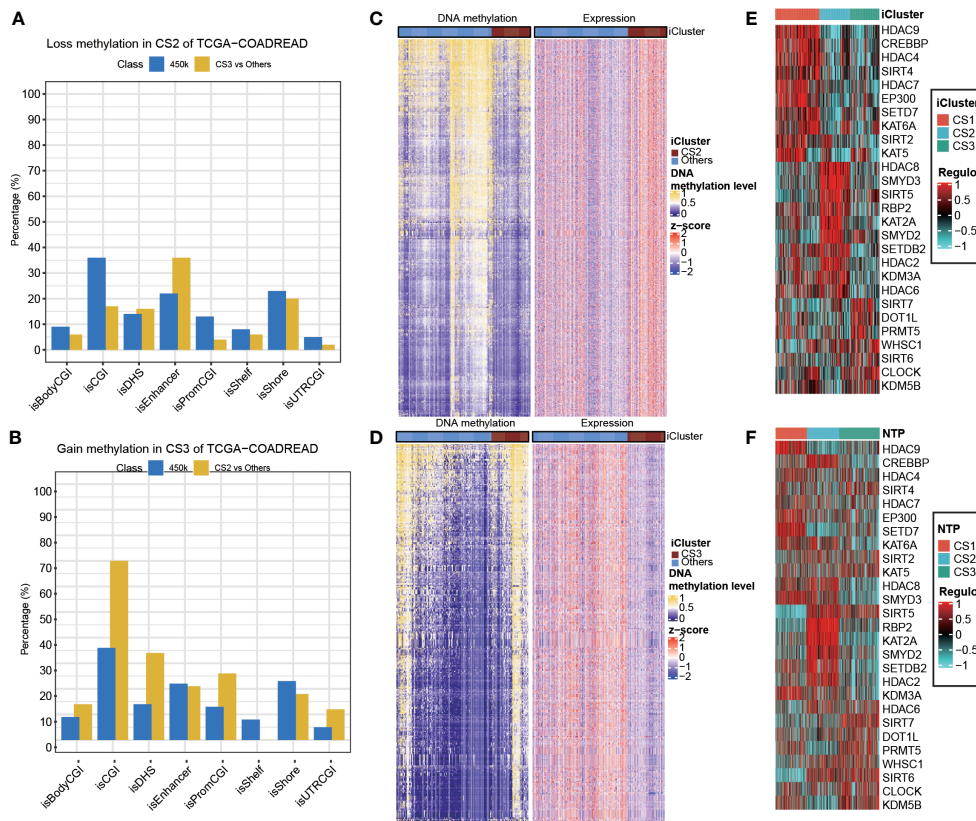
FIGURE 6

Epigenetic regulation in colorectal cancer subtype. Heatmap showing activity of regulon relevant to potential regulators associated with chromatin remodelling in both **(A)** TCGA and **(B)** GEO cohorts. Heatmap showing the association between DNA methylation and gene expression, presenting with **(C)** an epigenetic activation pattern in CS2 and **(D)** an epigenetically silencing pattern in CS3 of TCGA cohort. Barplots showing the region-specific distribution of DMPs comparing to the Illumina 450karray background for the **(E)** CS2 and **(F)** CS3 molecular classification in TCGA cohort.

Identified a biomarker for multi-omics molecular subtype

To apply the molecular subtype better in the clinic, we identified a biomarker for our molecular subtype which most based on the DNA mutation and RNA expression among different subtypes. By using the Chi-square test for DNA mutation and fold change with adjust FDR value for RNA expression, then, we detected SOX9 was a significant gene in the CS3 subtype. Through the IHC experiment, we found SOX9 was higher in the 50 tumor tissue than the 50 tumor adjacent tissue. IHC showed the represent sample in adjacent and tumor samples (Figures 8A, B). TCGA-COAD public cohort also confirmed that SOX9 was higher in tumor tissue than that in adjacent tissue (Figures 8C, D). SOX9 mainly located in the nucleoplasm of cell in A-431, U-2 OS and U-251 MG multi cell lines by immunofluorescence with HPA001758 antibody in the Human Protein Atlas (HPA) (Figure 8E).

Discussion

The high incidence and mortality of CRC have brought a huge burden on patients. How to effectively judge the prognosis of CRC patients and correctly evaluate the severity of the disease of CRC patients are the main objectives of the study on the prognosis of CRC. The prognosis of patients based on traditional tumor typing is often very different. Molecular typing of tumors can better reflect the differences in internal molecular characteristics of tumors, which is the basis for the realization of precision medicine. Accurate identification of patients' molecular subtypes will help to accurately predict patient prognosis and develop personalized treatment plans.

Currently, tumor molecular subtype studies are mainly based on single omics data, such as transcriptomics, proteomics, genomics, etc (56–61). Bhattacharjee et al. divided lung adenocarcinoma into 4 subtypes by analyzing gene expression profile data from lung adenocarcinoma patients, and found that abnormal expression profile can be used to
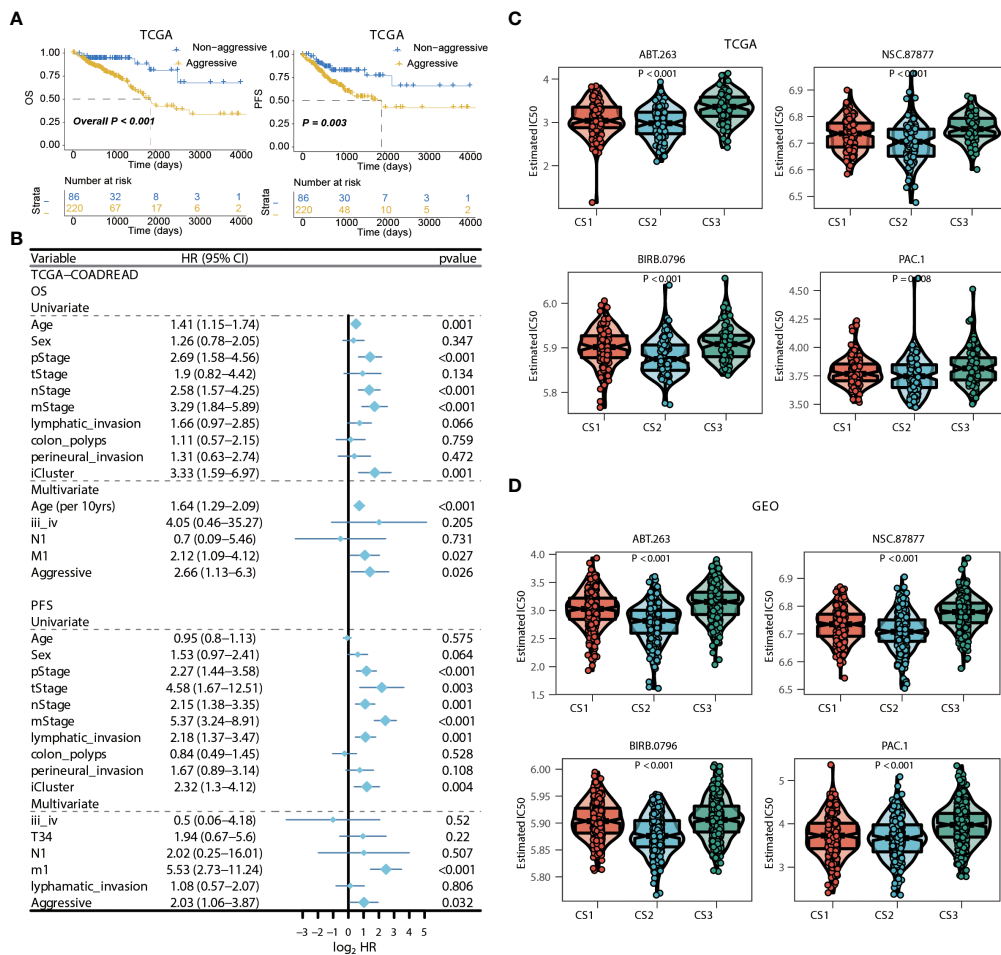
FIGURE 7

Identification of transcriptome biomarkers for colorectal cancer subtype. **(A)** KM of os using movics agreesiveness in coadread of tcga. **(B)** Forest plot showing the hazard ratio (95% CI) in univariate and multivariate Cox regressions with the corresponding *P* values. **(C, D)** Boxplot showing the distribution of estimated IC50 among three cancer subtypes based on GDSC database, **(C)** TCGA, **(D)** GEO.

distinguish primary and metastatic adenocarcinoma of lung (58). Based on genomic CNV data, Shibata et al. divided lung adenocarcinoma into three subtypes by unsupervised clustering analysis, and found that patients with EGFR mutations had shorter disease-free survival times (60). As for CRC, Roepman et al. conducted unsupervised classification of genome-wide data of CRC patients based on EMT, microsatellite instability caused by mismatch repair gene defects, and high mutation frequency associated with cell proliferation (62). Meanwhile, Lai et al. proposed the co-ordinate immune response cluster (CIRC), and identified four patient groups by this method (63). Zhang et al. identified two molecular subtypes, C1 and C2, based on cell cycle-related genes. PIK3CA, RYR2 and FBXW7 mutations were more frequent in C1, and the clinical characteristics and prognosis of patients were relatively poor (64). In addition to the above genotyping based on gene mutations and cytogenetic changes in the genome (10, 21, 65–68), CRC was also classified based on differences in gene expression profiles and proteomic

(69–72)biomarkers. Therefore, molecular typing based on omics data can effectively identify clinically relevant tumor subtypes, which plays a very important role in judging patient prognosis and guiding clinical treatment.

Nevertheless, any single omics data can only reflect the intrinsic molecular characteristics of tumors from a single perspective, and the contribution of single-omics analysis to tumor typing is one-sided. Therefore, the integration of multi-omics information can simultaneously capture the heterogeneity of tumors in different omics and integrate the information from multiple perspectives to identify more accurate tumor molecular typing. As the high heterogeneity of tumors is determined by multiple omics, such as genome, epigenome, transcriptome, and proteome, the analysis of data from different omics sources is expected to better reveal the mechanism of tumor genesis and development. For the first time, Matan Hofree et al. integrated genomic mutations and protein interaction networks for molecular typing of tumors to identify subtypes significantly
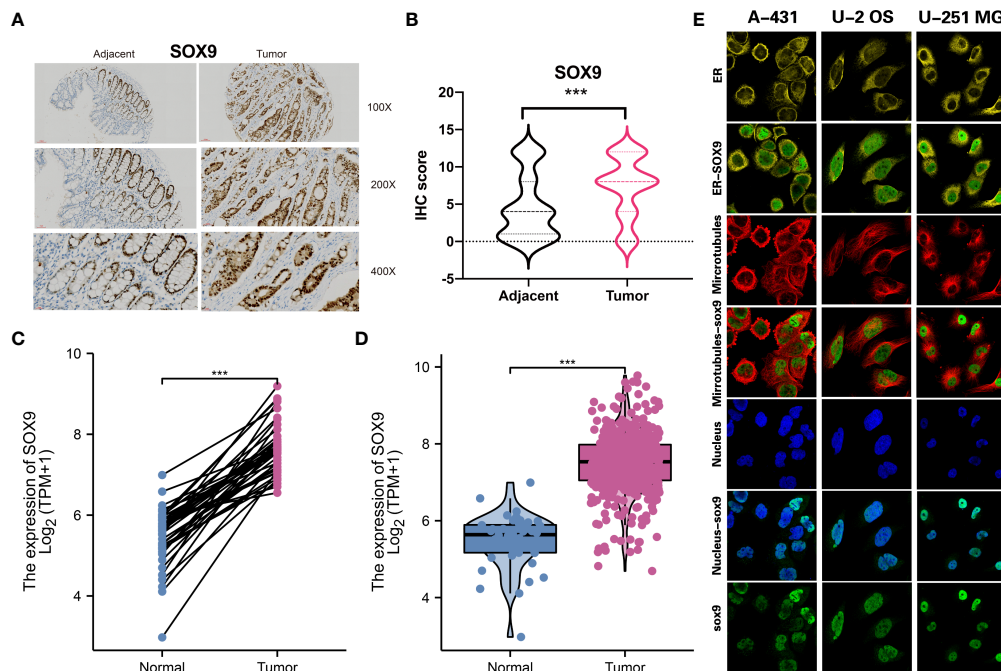
FIGURE 8

Molecular subtype biomarkers validated by wet experiment. **(A)** SOX9 protein expression of the represent sample in adjacent and tumor samples by IHC. **(B)** Pair-test for SOX9 protein expression between 50 tumor tissue and 50 tumor adjacent tissue by IHC. **(C)** and **(D)** SOX9 gene expression between tumor tissue and adjacent tissue in TCGA-COAD public cohort. **(E)** The location of SOX9 in A-431, U-2 OS and U-251 MG in the Human Protein Atlas(HPA). ***p<0.001.

associated with clinical features (73). Ronglai Shen et al. integrated genomic mutations, CNV and transcriptome expression profiles to obtain tumor classification based on iCluster (74). Herein, using transcriptome, DNA methylation, and driver mutations of CRC, we developed a classifier based on multi-omics integration for the prognosis prediction of CRC for the first time. At present, many studies have proved that CRC is the result of accumulation of multiple gene mutations and epigenetic modifications, and DNA hypermethylation or hypomethylation can be used as epigenetic biomarkers to predict the occurrence and prognostic effects of CRC (75–77). Driver mutations in the genome can be viewed as responsible for molecular changes associated with CRC progression, so targeting such genes for the elimination of multiple CRC gene dependencies could significantly improve efficacy (78). In conclusion, CRC can be comprehensively understood from multiple omics based on transcriptome, DNA methylation, and driver mutation levels to predict prognosis and guide clinical medication.

In the medical field, prognostic models need to undergo extensive and rigorous validation before they can be used in practice, and they also need to be continuously evaluated by feedback. At present, due to the different data standards and coding systems used by different sources, the output platforms

and schemes of omics data also have certain heterogeneity. Therefore, the current integrated prognostic models are often internally verified by resampling or cross-validation. The few externally validated integrated prognostic models often involve only one type of omics data and have been externally validated in only a few open data sets, making it difficult for the current integrated prognostic models to be applied in clinical practice. In order to verify the reproducibility of the colorectal molecular subtypes we identified, we combined four external datasets from GEO cohort. We removed batch effects across different datasets and predicted the identified molecular subtypes in the GEO cohort using NTP algorithm. CS3 presented with the most favourable clinical outcome out of the three subtypes, indicating the accuracy of the subtype system.

Beyond that, there are several new findings and notable advantages to our study. TME and tumor cells interact and co-evolve to drive tumor growth and progression, and also play an important role in regulating tumor sensitivity to treatment (79). The results showed that CS1 was highly immune-infiltrated, CS3 showed relatively higher immunocyte infiltration, while CS2 was generally immune-depleted, explaining the difference in prognosis. Immunotherapy is an important treatment for CRC. We compared the responses of the three subtypes to immune checkpoint inhibitors. Abnormal metabolism is

closely related to the occurrence, development, recurrence, metastasis, and prognosis of CRC. We found that the enrichment level of carbohydrate, amino acid, and lipid metabolism-relevant pathways in CS3 was higher. Our results showed that CS2 tended to loss methylations while CS3 tended to gain methylations. ABT.263 is a small molecule Bcl-2 inhibitor that can induce cell apoptosis (80). BIRB.0796 is one of the most potent compounds of (81) p38 inhibitors. PAC.1 (Caspase activator) is an effective procaspase-3 activator, which acts on primary cancer cells and induces apoptosis (82). Our findings showed that ABT.263, NSC.87877, BIRB.0796, and PAC.1 were discovered to be potentially effective in treating patients with CS2 phenotype.

Nonetheless, some limitations of the current study should not be ignored. Hence, the cases of CRC patients were relatively small; more cases are needed to confirm our conclusions. The molecular subtypes of CRC were based on retrospective cohorts. Therefore, prospective studies are needed in the future. Even though we developed molecular subtypes based on integrated multi-omics, the metabolomics and proteomics data were missing because the relevant omics information was not available in the TCGA database. With the development of information technology and genetic testing technology, more and more clinical data in the form of accessible electronic medical records and shared omics data are available. The rapid development of artificial intelligence technology can further mine the correlation and interaction between different scales of data and more effectively use different scales of data for information complementarity to achieve a more accurate prediction model. Therefore, it is of great significance to further improve and optimize the multi-omics analysis based on this study, realize the multi-center collaborative multi-omics integrated analysis, and apply it to the prognostic analysis of CRC.

Conclusion

Taken together, we carried out multi-omics analysis of transcriptome mRNA expression, DNA methylation, and colorectal cancerous driver mutations. Three molecular subtypes were constructed and clinical significances, such as prognosis, mechanisms, and clinical therapeutic targets were observed among them. Besides, the subtypes were independent prognostic factors.

Data availability statement

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/Supplementary Material.

Author contributions

CZ, TX, LC, XZ, XL, KC, and XC designed the study. All authors contributed to the article and approved the submitted version.

Funding

Conflict of interest

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

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2022.983636/full#supplementary-material

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin* (2019) 69:7–34. doi: 10.3322/caac.21551

2. Li J, Guo BC, Sun LR, Wang JW, Fu XH, Zhang SZ, et al. TNM staging of colorectal cancer should be reconsidered by T stage weighting. *World J Gastroenterol* (2014) 20:5104–12. doi: 10.3748/wjg.v20.i17.5104

3. Amin MB, Greene FL, Edge SB, Compton CC, Gershenwald JE, Brookland RK, et al. The eighth edition AJCC cancer staging manual: Continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. *CA Cancer J Clin* (2017) 67:93–9. doi: 10.3322/caac.21388

4. Roswall P, Bocci M, Bartoschek M, Li H, Kristiansen G, Jansson S, et al. Microenvironmental control of breast cancer subtype elicited through paracrine platelet-derived growth factor-CC signaling. *Nat Med* (2018) 24:463–73. doi: 10.1038/nm.4494

5. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LJ, Kinzler KW. Cancer genome landscapes. *SCIENCE* (2013) 339:1546–58. doi: 10.1126/science.1235122

6. Weiser MR. AJCC 8th edition: Colorectal cancer. *Ann Surg Oncol* (2018) 25:1454–5. doi: 10.1245/s10434-018-6462-1

7. Hudson TJ, Anderson W, Artez A, Barker AD, Bell C, Bernabé RR, et al. International network of cancer genome projects. *NATURE* (2010) 464:993–8. doi: 10.1038/nature08987

8. Hasin Y, Seldin M, Lusis A. Multi-omics approaches to disease. *Genome Biol* (2017) 18:83. doi: 10.1186/s13059-017-1215-1

9. Ideker T, Dutkowski J, Hood L. Boosting signal-to-noise in complex biology: prior knowledge is power. *CELL* (2011) 144:860–3. doi: 10.1016/j.cell.2011.03.007

10. Guinney J, Dienstmann R, Wang X, de Reynies A, Schlicker A, Soneson C, et al. The consensus molecular subtypes of colorectal cancer. *Nat Med* (2015) 21:1350–6. doi: 10.1038/nm.3967

11. Kang KJ, Min BH, Ryu KJ, Kim KM, Chang DK, Kim JJ, et al. The role of the CpG island methylator phenotype on survival outcome in colon cancer. *GUT LIVER* (2015) 9:202–7. doi: 10.5009/gnl13352

12. Wang B, Mezlini AM, Demir F, Fiume M, Tu Z, Brudno M, et al. Similarity network fusion for aggregating data types on a genomic scale. *Nat Methods* (2014) 11:333–7. doi: 10.1038/nmeth.2810

13. Muzny DM, Bainbridge MN, Chang K, Dinh HH, Drummond JA, Fowler G, et al. Comprehensive molecular characterization of human colon and rectal cancer. *NATURE* (2012) 487:330–7. doi: 10.1038/nature11252

14. Koboldt DC, Fulton RS, McLellan MD, Schmidt H, Veizer JK, McMichael JF, et al. Comprehensive molecular portraits of human breast tumours. *NATURE* (2012) 490:61–70. doi: 10.1038/nature11412

15. Hoadley KA, Yau C, Wolf DM, Cherniack AD, Tamborero D, Ng S, et al. Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *CELL* (2014) 158:929–44. doi: 10.1016/j.cell.2014.06.049

16. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* (2013) 6:l1. doi: 10.1126/scisignal.2004088

17. Chabanais J, Labrousse F, Chaunavel A, Germot A, Maftah A. POFUT1 as a promising novel biomarker of colorectal cancer. *Cancers (Basel)* (2018) 10(11):411. doi: 10.dio.10.3390/cancers10110411

18. Jorissen RN, Gibbs P, Christie M, Prakash S, Lipton L, Desai J, et al. Metastasis-associated gene expression changes predict poor outcomes in patients with dukes stage b and c colorectal cancer. *Clin Cancer Res* (2009) 15:7642–51. doi: 10.1158/1078-0432.CCR-09-1431

19. Freeman TJ, Smith JJ, Chen X, Washington MK, Roland JT, Means AL, et al. Smad4-mediated signaling inhibits intestinal neoplasia by inhibiting expression of beta-catenin. *GASTROENTEROLOGY* (2012) 142:562–71. doi: 10.1053/j.gastro.2011.11.026

20. Tripathi MK, Deane NG, Zhu J, An H, Mima S, Wang X, et al. Nuclear factor of activated T-cell activity is associated with metastatic capacity in colon cancer. *Cancer Res* (2014) 74:6947–57. doi: 10.1158/0008-5472.CAN-14-1592

21. Marisa L, de Reynies A, Duval A, Selves J, Gaub MP, Vescovo L, et al. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. *PloS Med* (2013) 10:e1001453. doi: 10.1371/journal.pmed.1001453

22. Katz S, Irizarry RA, Lin X, Tripputi M, Porter MW. A summarization approach for affymetrix GeneChip data using a reference training set from a large, biologically diverse database. *BMC Bioinf* (2006) 7:464. doi: 10.1186/1471-2105-7-464

23. Wagner GP, Kin K, Lynch VJ. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci* (2012) 131:281–5. doi: 10.1007/s12064-012-0162-3

24. Gautier L, Cope L, Bolstad BM, Irizarry RA. Affy–analysis of affymetrix GeneChip data at the probe level. *BIOINFORMATICS* (2004) 20:307–15. doi: 10.1093/bioinformatics/btg405

25. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *BIOINFORMATICS* (2012) 28:882–3. doi: 10.1093/bioinformatics/bts034

26. Tian Y, Morris TJ, Webster AP, Yang Z, Beck S, Feber A, et al. ChAMP: updated methylation analysis pipeline for illumina BeadChips. *BIOINFORMATICS* (2017) 33:3982–4. doi: 10.1093/bioinformatics/btx513

27. Zhou W, Laird PW, Shen H. Comprehensive characterization, annotation and innovative use of infinium DNA methylation BeadChip probes. *Nucleic Acids Res* (2017) 45:e22. doi: 10.1093/nar/gkw967

28. Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A, et al. Comprehensive characterization of cancer driver genes and mutations. *CELL* (2018) 173:371–85. doi: 10.1016/j.cell.2018.02.060

29. Lu X, Meng J, Zhou Y, Jiang L, Yan F. MOVICS: an r package for multi-omics integration and visualization in cancer subtyping. *BIOINFORMATICS* (2020) 14:btaa1018. doi: 10.1093/bioinformatics/btaa1018

30. Mo Q, Shen R, Guo C, Vannucci M, Chan KS, Hilsenbeck SG. A fully Bayesian latent variable model for integrative clustering analysis of multi-type omics data. *BIOSTATISTICS* (2018) 19:71–86. doi: 10.1093/biostatistics/kxx017

31. Ramasamy MN, Minassian AM, Ewer KJ, Flaxman AL, Folegatti PM, Owens DR, et al. Safety and immunogenicity of ChAdOx1 nCoV-19 vaccine administered in a prime-boost regimen in young and old adults (COV002): a single-blind, randomised, controlled, phase 2/3 trial. *LANCET* (2021) 396:1979–93. doi: 10.1016/S0140-6736(20)32466-1

32. Xiao Y, Ma D, Zhao S, Suo C, Shi J, Xue MZ, et al. Multi-omics profiling reveals distinct microenvironment characterization and suggests immune escape mechanisms of triple-negative breast cancer. *Clin Cancer Res* (2019) 25:5002–14. doi: 10.1158/1078-0432.CCR-18-3524

33. Hanzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinf* (2013) 14:7. doi: 10.1186/1471-2105-14-7

34. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol* (2016) 17:218. doi: 10.1186/s13059-016-1070-5

35. Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* (2013) 4:2612. doi: 10.1038/ncomms3612

36. Jeschke J, Bizet M, Desmedt C, Calonne E, Dedeurwaerder S, Garaud S, et al. DNA Methylation-based immune response signature improves patient diagnosis in multiple cancers. *J Clin Invest* (2017) 127:3090–102. doi: 10.1172/JCI91095

37. Rosario SR, Long MD, Affronti HC, Rowsam AM, Eng KH, Smiraglia DJ. Pan-cancer analysis of transcriptional metabolic dysregulation using the cancer genome atlas. *Nat Commun* (2018) 9:5330. doi: 10.1038/s41467-018-07232-8

38. Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *NATURE* (2009) 462:108–12. doi: 10.1038/nature08460

39. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* (2015) 43:e47. doi: 10.1093/nar/gkv007

40. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb)* (2021) 2:100141. doi: 10.1016/j.xinn.2021.100141

41. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst* (2015) 1:417–25. doi: 10.1016/j.cels.2015.12.004

42. Xie Z, Bailey A, Kuleshov MV, Clarke D, Evangelista JE, Jenkins SL, et al. Gene set knowledge discovery with enrichr. *Curr Protoc* (2021) 1:e90. doi: 10.1002/cpz1.90

43. Silva TC, Coetzee SG, Gull N, Yao L, Hazelett DJ, Noushmehr H, et al. ELMER v.2: an R/Bioconductor package to reconstruct gene regulatory networks from DNA methylation and transcriptome profiles. *BIOINFORMATICS* (2019) 35:1974–7. doi: 10.1093/bioinformatics/bty902

44. Hoshida Y. Nearest template prediction: A single-sample-based flexible class prediction with confidence assessment. *PloS One* (2010) 5:e15543. doi: 10.1371/journal.pone.0015543

45. Eide PW, Bruun J, Lothe RA, Sveen A. CMScaller: an r package for consensus molecular subtyping of colorectal cancer pre-clinical models. *Sci REP-UK* (2017) 7:16618. doi: 10.1038/s41598-017-16747-x

46. Audia JE, Campbell RM. Histone modifications and cancer. *Cold Spring Harb Perspect Biol* (2016) 8:a19521. doi: 10.1101/cshperspect.a019521

47. Castro MA, de Santiago I, Campbell TM, Vaughn C, Hickey TE, Ross E, et al. Regulators of genetic risk of breast cancer identified by integrative network analysis. *Nat Genet* (2016) 48:12–21. doi: 10.1038/ng.3458

48. Geeleher P, Cox NJ, Huang RS. Clinical drug response can be predicted using baseline gene expression levels and *in vitro* drug sensitivity in cell lines. *Genome Biol* (2014) 15:R47. doi: 10.1186/gb-2014-15-3-r47

49. Geeleher P, Cox N, Huang RS. pRRophetic: an r package for prediction of clinical chemotherapeutic response from tumor gene expression levels. *PloS One* (2014) 9:e107468. doi: 10.1371/journal.pone.0107468

50. McGranahan N, Furness AJ, Rosenthal R, Ramskov S, Lyngaa R, Saini SK, et al. Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *SCIENCE* (2016) 351:1463–9. doi: 10.1126/science.aaf1490

51. Lu X, Jiang L, Zhang L, Zhu Y, Hu W, Wang J, et al. Immune signature-based subtypes of cervical squamous cell carcinoma tightly associated with human papillomavirus type 16 expression, molecular features, and clinical outcome. *Neoplasia (New York N.Y.)* (2019) 21:591–601. doi: 10.1016/j.neo.2019.04.003

52. Matkowski R, Gisterek I, Halon A, Lacko A, Szewczyk K, Staszek U, et al. The prognostic role of tumor-infiltrating CD4 and CD8 T lymphocytes in breast cancer. *Anticancer Res* (2009) 29:2445–51.

53. Shah W, Yan X, Jing L, Zhou Y, Chen H, Wang Y. A reversed CD4/CD8 ratio of tumor-infiltrating lymphocytes and a high percentage of CD4(+)FOXP3(+) regulatory T cells are significantly associated with clinical outcome in squamous cell carcinoma of the cervix. *Cell Mol Immunol* (2011) 8:59–66. doi: 10.1038/cmi.2010.56

54. Wu X, Nelson M, Basu M, Srinivasan P, Lazarski C, Zhang P, et al. MYC oncogene is associated with suppression of tumor immunity and targeting myc induces tumor cell immunogenicity for therapeutic whole cell vaccination. *J Immunother Cancer* (2021) 9(3):e001388. doi: 10.1136/jitc-2020-001388

55. Dongre A, Weinberg RA. New insights into the mechanisms of epithelial-mesenchymal transition and implications for cancer. *Nat Rev Mol Cell Biol* (2019) 20:69–84. doi: 10.1038/s41580-018-0080-4

56. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *P Natl Acad Sci USA* (2003) 100:8418–23. doi: 10.1073/pnas.0932692100

57. Sun Y, Zhang W, Chen D, Lv Y, Zheng J, Lilljebjorn H, et al. A glioma classification scheme based on coexpression modules of EGFR and PDGFRA. *Proc Natl Acad Sci U.S.A.* (2014) 111:3538–43. doi: 10.1073/pnas.1313814111

58. Bhattacharjee A, Richards WG, Staunton J, Li C, Monti S, Vasa P, et al. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc Natl Acad Sci U.S.A.* (2001) 98:13790–5. doi: 10.1073/pnas.191502998

59. Pena-Llopis S, Vega-Rubin-de-Celis S, Liao A, Leng N, Pavia-Jimenez A, Wang S, et al. BAP1 loss defines a new class of renal cell carcinoma. *Nat Genet* (2012) 44:751–9. doi: 10.1038/ng.2323

60. Shibata T, Uryu S, Kokubu A, Hosoda F, Ohki M, Sakiyama T, et al. Genetic classification of lung adenocarcinoma based on array-based comparative genomic hybridization analysis: its association with clinicopathologic features. *Clin Cancer Res* (2005) 11:6177–85. doi: 10.1158/1078-0432.CCR-05-0293

61. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. *NATURE* (2000) 406:747–52. doi: 10.1038/35021093

62. Roepman P, Schlicker A, Tabernero J, Majewski I, Tian S, Moreno V, et al. Colorectal cancer intrinsic subtypes predict chemotherapy benefit, deficient mismatch repair and epithelial-to-mesenchymal transition. *Int J Cancer* (2014) 134:552–62. doi: 10.1002/ijc.28387

63. Lal N, Beggs AD, Willcox BE, Middleton GW. An immunogenomic stratification of colorectal cancer: Implications for development of targeted immunotherapy. *ONCOIMMUNOLOGY* (2015) 4:e976052. doi: 10.4161/2162402X.2014.976052

64. Zhang Z, Ji M, Li J, Wu Q, Huang Y, He G, et al. Molecular classification based on prognostic and cell cycle-associated genes in patients with colon cancer. *Front Oncol* (2021) 11:636591. doi: 10.3389/fonc.2021.636591

65. Budinska E, Popovici V, Tejpar S, D'Ario G, Lapique N, Sikora KO, et al. Gene expression patterns unveil a new level of molecular heterogeneity in colorectal cancer. *J Pathol* (2013) 231:63–76. doi: 10.1002/path.4212

66. Schlicker A, Beran G, Chresta CM, McWalter G, Pritchard A, Weston S, et al. Subtypes of primary colorectal tumors correlate with response to targeted treatment in colorectal cell lines. *BMC Med Genomics* (2012) 5:66. doi: 10.1186/1755-8794-5-66

67. Zhang J, Wang N, Wu J, Gao X, Zhao H, Liu Z, et al. 5-methylcytosine related LncRNAs reveal immune characteristics, predict prognosis and oncology treatment outcome in lower-grade gliomas. *Front Immunol* (2022) 13:844778. doi: 10.3389/fimmu.2022.844778

68. Perez-Villamil B, Romera-Lopez A, Hernandez-Prieto S, Lopez-Campos G, Calles A, Lopez-Asenjo JA, et al. Colon cancer molecular subtypes identified by expression profiling and associated to stroma, mucinous type and different clinical behavior. *BMC Cancer* (2012) 12:260. doi: 10.1186/1471-2407-12-260

69. Zhang B, Wang J, Wang X, Zhu J, Liu Q, Shi Z, et al. Proteogenomic characterization of human colon and rectal cancer. *NATURE* (2014) 513:382–7. doi: 10.1038/nature13438

70. Wang J, Mouradov D, Wang X, Jorissen RN, Chambers MC, Zimmerman LJ, et al. Colorectal cancer cell line proteomes are representative of primary tumors and predict drug sensitivity. *GASTROENTEROLOGY* (2017) 153:1082–95. doi: 10.1053/j.gastro.2017.06.008

71. Mischak H, Ioannidis JP, Argiles A, Attwood TK, Bongcam-Rudloff E, Broenstrup M, et al. Implementation of proteomic biomarkers: making it work. *Eur J Clin Invest* (2012) 42:1027–36. doi: 10.1111/j.1365-2362.2012.02674.x

72. Sveen A, Kopetz S, Lothe RA. Biomarker-guided therapy for colorectal cancer: strength in complexity. *Nat Rev Clin Oncol* (2020) 17:11–32. doi: 10.1038/s41571-019-0241-1

73. Hofree M, Shen JP, Carter H, Gross A, Ideker T. Network-based stratification of tumor mutations. *Nat Methods* (2013) 10:1108–15. doi: 10.1038/nmeth.2651

74. Shen R, Olshen AB, Ladanyi M. Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis. *Bioinf (Oxford England)* (2009) 25:2906–12. doi: 10.1093/bioinformatics/btp543

75. Akimoto N, Zhao M, Ugai T, Zhong R, Lau MC, Fujiyoshi K, et al. Tumor long interspersed nucleotide element-1 (LINE-1) hypomethylation in relation to age of colorectal cancer diagnosis and prognosis. *Cancers (Basel)* (2021) 13(9):2016. doi: 10.3390/cancers13092016

76. Hagg S, Jylhava J. Should we invest in biological age predictors to treat colorectal cancer in older adults? *Eur J Surg Oncol* (2020) 46:316–20. doi: 10.1016/j.ejso.2019.11.003

77. Liu J, Liu Z, Zhang X, Yan Y, Shao S, Yao D, et al. Aberrant methylation and microRNA-target regulation are associated with downregulated NEURL1B: a diagnostic and prognostic target in colon cancer. *Cancer Cell Int* (2020) 20:342. doi: 10.1186/s12935-020-01379-5

78. Raskov H, Soby JH, Troelsen J, Bojesen RD, Gogenur I. Driver gene mutations and epigenetics in colorectal cancer. *Ann Surg* (2020) 271:75–85. doi: 10.1097/SLA.0000000000003393

79. Jin K, Ren C, Liu Y, Lan H, Wang Z. An update on colorectal cancer microenvironment, epigenetic and immunotherapy. *Int IMMUNOPHARMACOL* (2020) 89:107041. doi: 10.1016/j.intimp.2020.107041

80. Chang J, Wang Y, Shao L, Laberge RM, Demaria M, Campisi J, et al. Clearance of senescent cells by ABT263 rejuvenates aged hematopoietic stem cells in mice. *Nat Med* (2016) 22:78–83. doi: 10.1038/nm.4010

81. Pargellis C, Tong L, Churchill L, Cirillo PF, Gilmore T, Graham AG, et al. Inhibition of p38 MAP kinase by utilizing a novel allosteric binding site. *Nat Struct Biol* (2002) 9:268–72. doi: 10.1038/nsb770

82. Putt KS, Chen GW, Pearson JM, Sandhorst JS, Hoagland MS, Kwon JT, et al. Small-molecule activation of procaspase-3 to caspase-3 as a personalized anticancer strategy. *Nat Chem Biol* (2006) 2:543–50. doi: 10.1038/nchembio814