Coupling of serum CK20 and hyper-methylated CLIP4 as promising biomarker for colorectal cancer diagnosis: from bioinformatics screening to clinical validation

Zhongjian Liu1,2,*, Hui Tang1,2,*, Wen Zhang1,2, Jinli Wang1,2, Lilan Wan1,2, Xisha Li1,2, Yuping Ji3, Na Kong3, Yanfang Zhang3, Jiangang Wang3, Zhang Fan3, Qiang Guo1,2

1Department of Gastroenterology, The First People’s Hospital of Yunnan Province, Kunming, China
2The Affiliated Hospital of Kunming University of Science and Technology, Kunming, China
3Department of Gastroenterology, The Third People’s Hospital of Yunnan Province, Kunming, China

*Equal contribution

Correspondence to: Qiang Guo, Zhang Fan; email: gqkj003@sina.com, https://orcid.org/0000-0003-0636-5548; 2873144767@qq.com, https://orcid.org/0000-0002-1528-7353

Keywords: colorectal cancer, serum diagnostic biomarkers, DNA methylation, CK20, CLIP4

Received: September 7, 2021    Accepted: December 13, 2021    Published: December 29, 2021

Copyright: © 2021 Liu et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Colorectal cancer (CRC) is one of the most common and lethal malignancies. The identification of minimally invasive and precise biomarkers is an urgent need for the early diagnosis of CRC. Through bioinformatics analysis of 395 CRC tissues and 63 CRC cell lines, CK18, CK20, de-methylated HPDL and hyper-methylated CLIP4 were identified as candidate serum biomarkers. Then, a training cohort consisting of 60 CRC, 30 colorectal adenomas (CA) and 33 healthy controls and a validation cohort consisting of 60 CRC, 30 CA and 30 healthy controls were enrolled. In the training cohort, enzyme-linked immunosorbent assay (ELISA) showed that CK18 and CK20 were all significantly higher in CRC and CA. CK18 diagnosed CRC with 46.67% sensitivity and 87.3% specificity; CK20 diagnosed CRC with 28.33% sensitivity and 90.47% specificity. Methylation-specific PCR (MSP) indicated that de-methylated HPDL and hyper-methylated CLIP4 were significantly detected in CRC and CA. De-methylated HPDL diagnosed CRC with 36.67% sensitivity and 93.65% specificity and hyper-methylated CLIP4 with 73.33% sensitivity and 84.13% specificity. Random combined analysis suggested that CK20/hyper-methylated CLIP4 diagnosed CRC with 91.67% sensitivity and 82.54% specificity. In the validation cohort, CK20 diagnosed CRC with 36.7% sensitivity and 88.3% specificity and hyper-methylated CLIP4 with 80% sensitivity and 85% specificity. CK20/hyper-methylated CLIP4 diagnosed CRC with 95% sensitivity and 81.7% specificity. Compared with serum biomarkers reported before, CK20/hyper-methylated CLIP4 possessed the potential to be a new effective and precise diagnostic biomarker for CRC.

INTRODUCTION

Colorectal cancer (CRC) is one of the most common malignancies worldwide and has high mortality rates. In recent years, the incidence of CRC has risen. CRC has become the third most common cancer among males and the second most common cancer among females [1]. CRC development is a complex multistep process that involves a gradual progression from adenomatous polyps to adenomas, and then to malignant carcinomas [2]. From a clinical perspective, CRC is difficult to diagnose early, as patients do not present with symptoms such as colorectal bleeding or anemia until later stages, and the survival rate decreases as the stage of diagnosis increases. Therefore, early detection and rapid diagnosis are important for CRC screening and treatment. Blood serum contains a certain amount of secretory proteins and cell-free DNA (cfDNA) derived...
from all cells in the body and could be a useful material for screening CRC.

Currently, several serum markers such as carcinoembryonic antigen (CEA), carbohydrate antigen 19-9 (CA199), carbohydrate antigen 125 (CA125), carbohydrate antigen 242 (CA242) and alpha fetoprotein (AFP) have been applied for diagnosing and monitoring CRC in the clinic [3, 4]. These biomarkers achieved 10.39~46.59% sensitivity and 80~95% specificity in diagnosing CRC [5, 6].

Aberrant DNA methylation changes have previously been shown to be an early event in the development of CRC [7] can be detected in cfDNA, making it an ideal and useful biomarker for the early detection of CRC [8, 9]. Currently, various tumor suppressor genes have emerged as potential blood-based methylation markers for CRC including APC, MGMT, hMLH1, HLF, ALX4, NGFR, TMEFF2, NEUROG1, SERP2, VIM, RASSF2A, WIFI1, RUNX3 and SEPT9 with sensitivities spanning from 34% to 90% and specificities ranging from 69% to 100% [10, 11].

With the vast amounts of CRC transcriptomics and DNA methylation data that are continuously generated and easily accessed from published sources, it is possible to use bioinformatics to screen biomarkers for CRC diagnosis, specifically and systematically. In this study, The Cancer Genome Atlas (TCGA), Genotype-Tissue Expression (GTEx) [12], Cancer Cell Line Encyclopedia (CCLE) [13], Gene Expression Profiling Interactive (GEPIA) [14], Human Protein Atlas (HPA) [15], UCSC [16], UALCN [17] and MEXPRESS [18] were used to screen specific secretory protein-encoding genes, de-methylated overexpressed genes and hypermethylated underexpressed genes in CRC tissues and cell lines. Then, these candidate biomarkers in CRC cell lines and clinical serum samples including CRC, colorectal adenoma (CA) patients and healthy controls, were detected and the relationship with clinicopathologic parameters and their value as CRC diagnostic markers were analyzed.

MATERIALS AND METHODS

Bioinformatics analysis

mRNA data of 395 CRC patients were downloaded from the TCGA database. The “limma” package was used to calculate the DEGs between CRC tissues and normal colorectal tissues, and the filter was applied according to the thresholds |log₂FC|>1 and P value <0.01. Specifically overexpressed or underexpressed genes in CRC tissues were verified by GEPIA. Overexpressed genes in CRC cell lines were selected by CCLE. Genes that encoded secretory proteins were screened according to HPA. The methylation status in CRC tissues and the CpG island locations of candidate genes were checked by UALCAN, MEXPRESS and UCSC.

Clinical specimens

Serum and tissue samples were obtained from the First People’s Hospital of Yunnan Province and the Third People’s Hospital of Yunnan Province with informed consent, comprising a training cohort (60 CRC, 30 colorectal CA and 33 healthy controls) and a validation cohort (60 CRC, 30 CA and 30 healthy controls). The diagnosis of CRC was verified by endoscopy and pathological biopsy. None of the patients had received prior radiotherapy, chemotherapy or surgery treatment when blood samples were collected. In addition, 1 placental sample was used as a control to test the methylation status of HPDL and CLIP4.

Cell culture and treatment

Seven human CRC cell lines (HT29, HCT116, SW480, SW620, RKO, DLD-1 and LOVO) and one normal colon cell line (CCD841CON) were obtained from the cell bank of the Chinese Academy of Sciences (Shanghai, China). All of cell lines were cultured in DMEM medium containing 10% fetal bovine serum (BI) and 100 IU/ml penicillin and streptomycin (Gibco) and maintained in 37°C in a humidified incubator with 5% CO₂. For de-methylation treatment, cultured cells were incubated with 10 μM 5-aza-2'-deoxycytidine (Sigma, USA) for 3 days with medium changed every day.

Quantitative PCR (Q-PCR) and Real-time PCR (RT-PCR)

The mRNA expression of candidate genes was analyzed by Q-PCR and RT-PCR. Total RNA was extracted with a Tissue Total RNA Isolation Kit (TSINGKE, China) and cDNA was obtained with a PrimerScript™ RT Reagent Kit (TSINGKE, China). Q-PCR was performed with 2 × Taq PCR Master Mix (TIANGEN, China). Real-time PCR was performed with EvaGreen 2 × qPCR MasterMix (Takara, Japan) in a CFX96TM Real-Time PCR System (BioRad, USA). The PCR reaction conditions were listed as follows: pre-denaturation at 94°C for 1.5 min, 30 cycles of pre-denaturation at 94°C for 10 s, annealing at 60°C for 20 s, extension at 72°C of 30 s, and ultimate extension at 72°C of 1 min. Primer sequences (10 μM concentration), annealing temperatures, and product sizes are listed in Table 1. The expression of the assayed genes was normalized to GAPDH.
Table 1. Primer sequences and product length.

| Gene          | Primer sequence (5′–3′) | Annealing temp (°C) | Amplification size (bp) |
|---------------|-------------------------|---------------------|------------------------|
| ADHFE1        | F: GTGAGAGTGGAAACCAACGGATTC | R: AGCACCTTACAGGTTGTCCATG | 60 | 120 |
| ASCL2         | F: CGCTCTACCTGGAGCCACGAG | R: CCCGCTATGCGTGGCTTCCG | 60 | 140 |
| B3GNT3        | F: AGGCAAGACATTCCAGGAGCAT | R: GTTGAGACAGCAGAGCAGTT | 60 | 128 |
| CCL24         | F: TGAAGAACCGTGCTCAGCTC | R: TTCTGTCTGGGCTCAGGTTC | 60 | 153 |
| CDX1          | F: GAGGAGGAAGTCTTACACACC | R: GTTCACCTGTGGCTCTTTGCC | 60 | 132 |
| CDX2          | F: ACAGTGCTCAGATCATCCACC | R: CTCCTCTTTTGGCTCTCAGTTC | 60 | 102 |
| CEACAM5       | F: GCCCTCAATAGGACACAGTCAC | R: CAGTGTAAGGTCACGATCCCTC | 60 | 115 |
| CHRD1         | F: GGTCTTTTGCAGAATCGCAAC | R: AGAGACTGGAAACGACAGTT | 60 | 113 |
| CLIP4         | F: CTTGGAAGTGCTGCTGGAGCA | R: GCTTGATTTCTTAGGTGGCCT | 60 | 141 |
| CPXM2         | F: CAGAGGATCCAGAAGATCTCCC | R: CATCGACGCTACTGGCTTGC | 60 | 119 |
| CST1          | F: TGTGGCTCTAGGACACAGCA | R: CTTGCCAGATCTTGGAGTTT | 60 | 130 |
| CYP2S1        | F: GATGGAGCGTCTAGGAGAAGAT | R: GGAGAAGCTGGTTGAAGGTAG | 60 | 126 |
| DEF5A         | F: FCTCAGACGAAGGCGTCTTGC | R: TGGCAAGATAGCTGGCTTCC | 60 | 141 |
| DEF6A         | F: ATGAGCAAGACTTTGCGCTGTC | R: CATGACAGTGAGGTCCTCACATG | 60 | 140 |
| EPCAM         | F: GCCAGAGTGAAGCTTCGCTGC | R: CTCCTAGCTGTTGCTCTCCC | 60 | 122 |
| EPHB2         | F: CGCACATCTGATGCTTACAGTG | R: GATGAGTGCAACCTTCCTCGG | 60 | 130 |
| FERM1         | F: CCAACTCTTTGAGACAAGCAA | R: GCTGCTTTGGACAGACACAG | 60 | 128 |
| FTRA1         | F: CATGACACTCCAGAAGGCTGCT | R: GTCACATCCAGACAGTCGCCA | 60 | 151 |
| GMDS          | F: TGGATTTCTGGCAGAAAGGC | R: CAAAGGAGATCTGGCTAGTGA | 60 | 161 |
| GSTM2         | F: AGATACACCGAGCAAAGCCCAT | R: GGGCTTCCATAAACTGGTCTTC | 60 | 117 |
| HPDL          | F: AGCCAGAAGAGGAGGAGCAG | R: GGAGATCTGGGAGCAGCTG | 60 | 119 |
| IHH           | F: GGAGCAGCTTGAAGGCAAGATCG | R: CAGCGAGTTGCCGCTCTT | 60 | 150 |
| KCNE3         | F: GCCGCTGATGAACTTCTCTACATG | R: CACTAGCCTGTTTCACCTGCG | 60 | 114 |
| KRT8          | F: ACAAGGTAGAGCTGGAGTCTCG | R: AGGACCAAGAGTGCTGGAGA | 60 | 121 |
| KRT18         | F: GCTGGAGAAAGGAGGAGATTCTT | R: GGGATCTGAGACAGAGAC | 60 | 119 |
| KRT20         | F: CTTGAGTTTCAACAAAGGAGCTG | R: AACAGCGGAGCTGAGTTGCTA | 60 | 151 |
| LGALS4        | F: GGAACAGGCCCTCTGAGAAAGGC | R: CAACTGCTGTTGAAAGAGGCC | 60 | 130 |
| LGR5          | F: CCTGTGTTGACTTTGAGGAGACCC | R: CACCGCACTAAACAGGCTGTC | 60 | 100 |
| MUC3A         | F: TCTTAGTCTACATTTGGAAGGAG | R: TGGGGAGGTTGTTTGTAAG | 60 | 262 |
| MUC5B         | F: CTGCTACGACAAGGAGCCAAAC | R: AAGGCTTGAGGCAGACTGAT | 60 | 112 |
| MUC13         | F: TGGCTGTAACAGAAGTGGGAT | R: GCATCGAGGACTTGAAGACT | 60 | 123 |
| NFE2L3        | F: CCAGTGTTCTTCTATTCAAGCTGCTG | R: CAACCTGCTAACTGAGCGTAG | 60 | 142 |
| OLFM4         | F: GCCAAGAGGAAATGTGTGAGG | R: CACTCGATGAGCTTACGGAC | 60 | 138 |
| PDLM4         | F: TGATAGCAGGAAACGGAGAGCA | R: AGGCTTTGCTCTGCCATCCTC | 60 | 123 |
| PRR15         | F: CCTGATGACTTCTTTGGCGGCAAC | R: GCGCTGATGAGTGTTGGTTG | 60 | 146 |
| SCLC1A2       | F: CCTCTACAAGAAGCCAGCTC | R: CTTGAGTTTGGACAGCTGTC | 60 | 124 |
| SPINK4        | F: TGGAGGACGAGACTGAAAGGAGG | R: CAAACGAGGAGCGACACATTCACT | 60 | 144 |
| SRMS          | F: CCTCTACTAAAGATGGAAGACCC | R: GGTGAGGACTTCTCTCGGTTCA | 60 | 197 |
| UCH1          | F: CAGTTGGAAGAGAGCAACCTGCTG | R: CCACAGACATTTAGGCGTCTT | 60 | 122 |
| GAPDH         | F: GTCTGTCTGCTGTAACACACGG | R: ACCACCTTGTTGGCTGCAAC | 60 | 131 |

Methylation-specific PCR

| Gene          | Primer sequence (5′–3′) | Annealing temp (°C) | Amplification size (bp) |
|---------------|-------------------------|---------------------|------------------------|
| HPDL          | F: ATTAGTTAGGGAGTGGAGTTTGA | R: AGCAAGACGTAAAAACAGT | 60 | 137 |
| CLIP4         | F: AGAGGCCTTAAAGATTAGGTTTGG | R: ACTAAACAGTCTACAAGAATATCG | 56 | 143 |

Abbreviations: F: forward primer; R: reverse primer; M: methylation; U: unmethylation; bps: base pairs.

Enzyme-linked immunosorbent assay (ELISA)

Commercial ELISA kits were used to measure CEA, CK18, CK20, MUC13, CK8 and EPCAM (CUSABIO, China). Experiments were performed according to the manufacturers’ instructions. Optical density (OD) values were read at a wavelength of 450 nm using a 96-well microplate. All determinations were performed in duplicate.
DNA extraction and bisulfite conversion

Genomic DNA from tissues and cells was extracted by using a TIANamp Genomic DNA Kit (TIANGEN, China). Genomic DNA from serum samples was extracted by the Axy Prep Body Fluid Viral DNA/RNA Miniprep Kit (Axy Prep, China). Complete bisulfite conversion of GC-rich DNA was performed by using the EZ DNA Methylation-Gold™ Kit (Zymo Research, USA).

Methylation-specific PCR (MSP)

The methylation status of HPDL and CLIP4 was detected by methylation-specific PCR assay utilizing the abovementioned bisulfite-modified DNA as templates, according to the previously mentioned protocols [19]. The methylated and de-methylated specific primer sequences (10 μM concentration), annealing temperatures, and product sizes are listed in Table 1. PCR products were evaluated by electrophoresis on ethidium bromide (EB)-stained 2% agarose gels. The sample was considered de-methylated HPDL when only a visible band was detected in unmethylation primer allele. The sample was considered hyper-methylated CLIP4 when a visible band was detected in the methylation primer allele. All of the samples were amplified twice to check the accuracy of the results.

Statistical analysis

The differences in CEA, CK18, CK20, MUC13, CK8 and EPCAM among the study groups were compared via nonparametric analysis. The correlations between CK18, CK20, de-methylated HPDL, hyper-methylated CLIP4 and clinicopathologic parameters were evaluated by the chi-square test or Fisher’s exact test. To evaluate the validity of each studied parameter, sensitivity and specificity were used. All statistical analyses were performed using SPSS 19.0 (SPSS Inc., USA).

RESULTS

High levels of serum CK18 and CK20 were detected in CRC and CA patients

TCGA and GTEx analysis revealed 2658 genes highly expressed in CRC tissues (log₂-fold >1, P < 0.01). Compared with human normal tissues, 100 genes were specifically overexpressed in CRC tissues (fold log₂ >3, P < 0.01) (Supplementary Table 1). Among them, 74 genes were overexpressed in CRC cell lines (rank Top 3) (Supplementary Table 2), and 16 genes encoded secretory proteins (Supplementary Table 3; Figure 1A). Then, Q-PCR was used to detect the expression of 16 genes in 7 CRC cell lines (HT29, HCT116, SW480, SW620, RKO, DLD-1, LOVO) and 1 normal colon cell line (CCD841CON). It was found that CEACAM5, KRT8, KRT18, KRT20, MUC13 and EPCAM were significantly overexpressed in CRC cell lines (Figure 1B). With ELISAs to test serum CEA (encoded by CEACAM5), CK8 (encoded by KRT8), CK18 (encoded by KRT18), CK20 (encoded by KRT20), MUC13 and EPCAM in CRC, CA patients and healthy controls, the results showed that CEA, CK18 and CK20 were significantly higher in CRC and CA patients than in healthy controls (all P < 0.05) (Figure 1C). GEPIA, HPA and CCLE also verified CEA, CK18 and CK20 overexpressed in CRC tissue; CEA and CK20 specifically increased in CRC cell lines (Figure 2).

De-methylated HPDL was observed in CRC and CA serum

Normally, DNA de-methylation can lead to genome instability and high expression of oncogenes. Based on the previous bioinformatics analysis results, among 74 specifically overexpressed genes, UCSC showed that 19 genes possessed CpG islands in their promoters or the first exon region (Supplementary Table 4; Figure 3A). Detecting the expression of 19 genes in 7 CRC cell lines and CCD841CON revealed that HPDL, LGR5, ASCL2, KCNE3, HNF4G, KRT8, KRT18, SLC12A2 and FERMT1 were significantly overexpressed in CRC cell lines (Figure 3B). To test the relationship between DNA methylation status and the expression of these genes in CRC, the expression of 9 genes in 7 CRC cell lines treated with 5′-aza-2′-deoxycytiding (DAC) were detected. As shown in Figure 3C and 3D, HPDL, KRT8, KRT18, FERMT1 and SLC12A2 were increased in CRC cell lines in response to DAC treatment. According to the CpG island region, MSP primers for these genes were designed and the methylation status of 5 genes in CRC cell lines and a normal colon cell line were tested. The results revealed that only HPDL presented more de-methylation status in CRC cell lines (especially SW620) than CCD841CON (Figure 4A and 4B). GEPIA and CCLE demonstrated that HPDL was highly expressed in CRC tissues and CRC cell lines; MEXPRESS also revealed that CRC tissues possessed HPDL de-methylated regions (Probes ID: cg13951491 and cg16593917) compared with normal tissue (Figure 5). These results indicated that HPDL overexpressed in CRC may be upregulated by DNA de-methylation.

Because serum contains a certain amount of DNA derived from lysed tumor cells, the methylation status of HPDL was detected in the serum of 60 CRC patients, 30 CA patients and 33 healthy controls. As shown in Figure 4C, HPDL de-methylation was detectable in
CRC and CA patients but not in healthy controls. Statistical analysis showed that the de-methylated frequency of serum HPDL was 36.7% (22/60) in CRC patients and 13.3% (4/30) in CA patients (Figure 4D). Additionally, representative cases consisting of 20 CRC, 10 CA patients and 10 healthy controls were selected to detect HPDL methylation status in serum and colorectal normal or tumor tissue from the same patient. The results indicated that the HPDL methylation status in serum was almost consistent with that in CRC tissues (Figure 4C).

Hyper-methylated CLIP4 was identified in CRC and CA serum

DNA hyper-methylation is associated with tumor suppressor gene silencing and defects in cell cycle regulation, resulting in tumor development and progression. Comparing the top 250 underexpressed genes in TCGA (Supplementary Table 5) with the top 250 promoter hyper-methylated genes in UALCAN (Supplementary Table 6), we found that 9 genes showed underexpression and promoter hyper-methylation in...
CRC tissue. UCSC exhibited that 8 genes possessed CpG islands located in promoters (Figure 6A). Detecting the expression of 8 genes in 7 CRC cell lines and CCD841CON revealed that CLIP4, GARA1 and UCHL1 were underexpressed in CRC cell lines and overexpressed in a normal colon cell line (Figure 6B). As determined by Q-PCR and RT-PCR, after DAC treatment, CLIP4 and UCHL1 were upregulated in CRC cell lines (Figure 6C and 6D). According to the CpG islands located in the promoter, MSP primers were designed and tested the methylation status of 2 genes in CRC cell lines and a normal colon cell line. The results showed that CLIP4 presented significant hyper-methylation in CRC cell lines and total de-methylation in a normal colon cell line (Figure 7A and 7B). GEPIA and UALCAN also indicated that CLIP4 was underexpressed and hyper-methylated in CRC tissue (Figure 8). By detecting the methylation status of CLIP4 in serum from 30 CRC, 20 CA patients and 33 healthy controls, it was found that CLIP4 hyper-methylation was detectable in CRC and CA but not in healthy serum.

![Figure 2](image2.png)

**Figure 2. The expression of CEA, CK18 and CK20 in tumor tissues and cancer cell lines.** The mRNA expression of CEA (A), CK18 (B) and CK20 (C) in certain tumor tissues and CRC tissues was analyzed by GEPIA. The protein expression of CEA (A), CK18 (B) and CK20 (C) in CRC tissues was stained by immunohistochemistry (IHC) and analyzed by HPA. The mRNA expression of CEA (A), CK18 (B) and CK20 (C) in cancer cell lines was analyzed by CCLE. Abbreviations: COAD: colon adenocarcinoma; READ: rectum adenocarcinoma.
(Figure 7C). By statistical analysis, the hyper-methylation frequency of serum CLIP4 was 73.3% (44/60) in CRC and 33.3% (10/30) in CA patients (Figure 7D). Furthermore, representative cases consisting of 20 CRC patients, 10 CA patients and 10 healthy controls were chosen to detect the CLIP4 methylation status in serum and colorectal normal or tumor tissue from the same patient. The results illustrated that the CLIP4 methylation status in serum was completely consistent with that in CRC tissue (Figure 7C).

Clinical values of serum CK18 and CK20 and de-methylated HPDL and hyper-methylated CLIP4 for CRC diagnosis

A training cohort consisting of 60 patients with CRC (age range: 46–87 years), 30 patients with CA (age range: 26–77 years), and 33 healthy controls (age range: 33–75 years), and a validation cohort consisting of 60 CRC (age range: 43–88 years), 30 CA (age range: 35–82 years), and 30 healthy controls (age range: 31–72 years) were enrolled in this study. The baseline and clinical values of serum CK18 and CK20 and de-methylated HPDL and hyper-methylated CLIP4 for CRC diagnosis were obtained. The expression of de-methylated genes was tested using Q-PCR (C) and RT-PCR (D) in CRC cell lines after treatment with DAC.

Figure 3. Specific overexpressed and de-methylated genes in CRC tissues were screened by bioinformatics and verified in CRC cell lines by Q-PCR or RT-PCR. (A) Screening specific overexpressed and de-methylated genes in CRC tissues by bioinformatics. (B) Testing the expression of de-methylated genes by Q-PCR in CRC cell lines and a normal colon cell line. The expression of de-methylated genes was tested using Q-PCR (C) and RT-PCR (D) in CRC cell lines after treatment with DAC.
characteristics of the patients and controls are summarized in Table 2.

The relationships between CK18, CK20 or HPDL, CLIP4 methylation status and various clinicopathologic parameters in CRC patients are summarized in Table 3. According to the results, in the training cohort, CK18 was significantly correlated with TNM stage, differentiation grade, CEA and CA19-9 (all $P < 0.05$). CK20 was closely correlated with tumor size and CA199 ($P < 0.05$). De-methylated HPDL was apparently associated with tumor size, CEA and CA199 ($P < 0.05$). Hyper-methylated CLIP4 was markedly associated with differentiation grade and CEA ($P < 0.05$) in CRC patients. In the validation cohort, CK20 was significantly correlated with tumor location and CA199 ($P < 0.05$). Hyper-methylated CLIP4 was closely associated with age, TNM stage, differentiation grade, lymphovascular invasion and CEA (all $P < 0.05$).

Further analysis suggested that under the best cutoff values defined by the tertiles method, in the training cohort, CK18 detected CRC with 46.67% sensitivity and 87.3% specificity; CK20 with 28.33% sensitivity and 90.47% specificity; de-methylated HPDL with 36.67% sensitivity and 93.65% specificity; and hyper-methylated CLIP4 with 73.33% sensitivity and 84.13% specificity. Random combined analysis suggested CK20/hyper-methylated CLIP4 with 91.67% sensitivity and 82.54% specificity. In the validation cohort, CK20 detected CRC with 36.7% sensitivity and 85% specificity; hyper-methylated CLIP4 with 80% sensitivity and 85% specificity; and CK20/hyper-methylated CLIP4 with 95% sensitivity and 81.7% specificity (Table 4). Considering sensitivity and specificity, CK20/hyper-methylated CLIP4 was a potential diagnostic biomarker for CRC.

**DISCUSSION**

Cytokeratin is a conserved group of proteins that form the cytoplasmic structure of epithelial cells and tissues. Cytokeratin 20 (CK20) is a type I cytokeratin. It is a prominent component of the intestinal epithelium. CK20 expression is confined to astrointestinal epithelium, urothelium, and Merkel cells of the epidermis, as well as malignancies that originate from the aforementioned sites [20]. According to previous studies, Y Imai indicated that CK20 expression in tumor tissues was an independent prognostic factor of poorly differentiated adenocarcinoma of the colon and rectum [21]. As one of the most investigated markers for the

---

**Figure 4.** Serum de-methylated HPDL in CRC, CA patients and healthy controls. (A) Schematic illustration of the gene structure of HPDL, the CpG island region and the position of MSP primers. (B) Detected HPDL methylation status in 8 cell lines. Placental DNA (or treated by SSSI) represented a positive control of de-methylated or methylated status. Abbreviations: M: methylation; U: un-methylation. (C) Representative serum and tissue methylation status of HPDL in CRC, CA patients and healthy controls. (D) Frequency of serum HPDL methylation status in 60 CRC, 30 CA patients and 33 healthy controls.
detection of circulating CRC cells, CK20 mRNA in serum is widely tested by RT-PCR for predicting recurrence and poor prognosis of CRC [22–29]. However, the efficacy of CK20 protein in serum as a biomarker for early CRC screening and diagnosis is not clear. In this study, we offered a precise value of serum CK20 protein in CRC diagnosis with 28.33% sensitivity and 90.47% specificity in the training cohort and 36.7% sensitivity and 88.3% specificity in the validation cohort. We also detected that CK20 presented higher levels in CA patients with a rate of 16.67% in the training cohort. This result indicated that CK20 possessed diagnostic potential for early CRC screening.

CLIP4, as a member of the CAP-Gly domain containing linker protein (CLIP) family, which is involved in plus-end binding of microtubules, has been implicated in immune response-related biological processes, cell migration and viability in certain cancer metastases [30]. Hyper-methylation of CLIP4 has been shown diagnostic potential for CRC in serum [31]. S.O. Jensen reported that hyper-methylated CLIP4 was capable of

Figure 5. The expression of HPDL in tumor tissues and cancer cell lines and the methylation status of HPDL in CRC tissues. The mRNA expression of HPDL in certain tumor tissues and CRC tissues was analyzed by GEPIA. The mRNA expression of HPDL in cancer cell lines was analyzed by CCLE. The relationship between the expression and promoter methylation level of HPDL in CRC tissues was analyzed by MEXPRESS. The red frame showed HPDL methylation status in CRC and normal tissues (Probes ID: cg13951491 and cg16593917). Abbreviations: COAD: colon adenocarcinoma; READ: rectum adenocarcinoma.
distinguishing serum from CRC patients and healthy controls (the area under the curve was 0.88) [32]. By testing the methylation status in CRC serum, we found that serum hyper-methylated CLIP4 detected CRC with a sensitivity of 73.33% and specificity of 84.13% in the training cohort and 80% sensitivity and 85% specificity in the validation cohort. We also detected hyper-methylated CLIP4 in CA patients at a rate of 33.3% but not in healthy controls. This implied that serum CLIP4 hyper-methylation could be used for early CRC screening.

Due to the highly heterogeneous nature of CRC, a single tumor marker is unlikely to become a stand-alone diagnostic test as the commonly insufficient sensitivity and/or specificity. Using a panel of tumor markers and testing with different methods for CRC diagnosis has the potential to be an effective approach. With systematic bioinformatics screening and clinical verification, our study showed that a combination of serum CK20 and hyper-methylated CLIP4 was a novel and effective biomarker for CRC diagnosis with 91.67% sensitivity and 82.54% specificity in the training cohort; and 95% sensitivity and 81.7% specificity in the validation cohort. It was more sensitive than CLIP4 hyper-methylated alone in stool specimens (90.3% sensitive, 88.4% specificity) [33]. Comparing with previous serum CRC biomarkers, CK20/hyper-methylated CLIP4 was more effective than CEA/MMP-7/TIMP-1 (sensitivity: 70.3%, specificity: 91.3%) [34], RUNX3/SFRP1/CEA (sensitivity 84.71%) [35], LRG1/EGFR/ITIH4/ HPX/SOD3 (sensitivity: over 70%, specificity: 89%) [36], anti-SLP2/-p53/-SEC61B/-PLSCR1 (sensitivity: 64.1%, specificity: 80%) [37], miR-203a-3p/miR-145-5p/miR-375-3p/miR-200c-3p (sensitivity: 81.52%, specificity: 73.33%) [38], miR-144-3p/miR-425-5p/miR-1260b (sensitivity: 93.8%, specificity: 91.3%) [39], and less than CCL20/IL-17A (sensitivity: 96.1%, specificity: 96.5%) [40]. Elevated CCL20 and IL-17A levels may reflect inflammatory condition, which can increase the false-positive fraction (FPF) of CRC detection [40]. In comparison, CRC cells

Figure 6. Underexpressed and hyper-methylated genes were screened in CRC tissues by bioinformatics and verified in CRC cell lines by Q-PCR or RT-PCR. (A) Screening underexpressed and hyper-methylated genes in CRC tissues by bioinformatics. (B) Testing the expression of hyper-methylated genes by Q-PCR in CRC cell lines and a normal colon cell line. The expression of hyper-methylated genes was tested using Q-PCR (C) and RT-PCR (D) in CRC cell lines after treatment with DAC.
Figure 7. The serum hyper-methylated status of CLIP4 in CRC, CA patients and healthy controls. (A) Schematic illustration of the gene structure of CLIP4, the position of CpG islands and MSP primers. (B) Detecting CLIP4 methylation status in 8 cell lines. Placental DNA (or treated by SSSI) represented a positive control for de-methylation or methylation. Abbreviations: M: methylation; U: un-methylation. (C) Representative serum and tissue methylation status of CLIP4 in CRC, CA patients and healthy controls. (D) Frequency of serum CLIP4 methylation status in 60 CRC, 30 CA patients and 33 healthy controls.

Figure 8. The expression of CLIP4 in tumor tissues and cancer cell lines and the methylation status of CLIP4 in CRC tissues. The mRNA expression of CLIP4 in certain tumor tissues and CRC tissues was analyzed by GEPIA. The methylation level of CLIP4 in CRC and normal tissues was analyzed by UALCAN. Abbreviations: COAD: colon adenocarcinoma; READ: rectum adenocarcinoma.
Table 2. Demographic and clinicopathologic characteristics of clinical cohorts.

| Characteristics       | Training cohort (N = 123) | Validation cohort (N = 120) |
|-----------------------|---------------------------|-----------------------------|
|                       | Normal | CA     | CRC | Normal | CA | CRC |
| Age                   |        |        |     |        |    |     |
| ≤50                   | 7 (21.2%) | 5 (16.7%) | 9 (15.0%) | 14 (46.7%) | 8 (26.7%) | 7 (11.7%) |
| >50                   | 26 (78.8%) | 25 (83.3%) | 51 (85.0%) | 16 (53.3%) | 22 (73.3%) | 53 (88.3%) |
| Sex                   |        |        |     |        |    |     |
| Male                  | 13 (39.4%) | 19 (63.3%) | 39 (65.0%) | 17 (56.7%) | 23 (76.7%) | 41 (68.3%) |
| Female                | 20 (60.6%) | 11 (36.3%) | 21 (35.0%) | 13 (43.3%) | 7 (23.3%) | 19 (31.7%) |
| Tumor location        |        |        |     |        |    |     |
| Colon                 |        | 22 (36.7%) | 19 (33.9%) |        |    |     |
| Rectum                |        | 38 (63.3%) | 37 (66.1%) |        |    |     |
| Tumor Size (cm)       |        |        |     |        |    |     |
| ≤4                    |        | 30 (50.0%) | 22 (39.3%) |        |    |     |
| >4                    |        | 30 (50.0%) | 34 (60.7%) |        |    |     |
| TNM stage             |        |        |     |        |    |     |
| I + II                |        | 34 (56.7%) | 16 (28.6%) |        |    |     |
| III + IV              |        | 26 (23.3%) | 40 (71.4%) |        |    |     |
| Differentiation       |        |        |     |        |    |     |
| Well                  |        | 7 (11.7%) | 5 (8.9%) |        |    |     |
| Moderate              |        | 48 (80%) | 42 (75.0%) |        |    |     |
| Poor                  |        | 5 (8.3%) | 9 (16.1%) |        |    |     |
| Lymphovascular invasion |      |        |     |        |    |     |
| Absent                |        | 36 (60.0%) | 21 (37.5%) |        |    |     |
| Present               |        | 24 (40.0%) | 35 (62.5%) |        |    |     |
| CEA                   |        |        |     |        |    |     |
| <5 ng/ml              |        | 28 (46.6%) | 21 (35%) |        |    |     |
| ≥5 ng/ml              |        | 32 (53.3%) | 39 (65%) |        |    |     |
| CA199                 |        |        |     |        |    |     |
| <37 U/ml              |        | 46 (76.7%) | 38 (73.1%) |        |    |     |
| ≥37 U/ml              |        | 14 (23.3%) | 14 (26.9%) |        |    |     |
| CA125                 |        |        |     |        |    |     |
| <35 U/ml              |        | 54 (90.0%) | 48 (92.3%) |        |    |     |
| ≥35 U/ml              |        | 6 (10.0%) | 4 (7.7%) |        |    |     |

Abbreviations: CA: colorectal adenomas; CRC: colorectal cancer.

overexpressed CK20 and showed hyper-methylated CLIP4. Serum CK20/hyper-methylated CLIP4 represented the tumor status of patients. The combination of serum CK20/hyper-methylated CLIP4 could decrease FPF of CRC detection.

In this study, we found several limitations, which should be regarded as preliminary research, and upcoming surveys should focus on several issues. First, CRCs can be characterized by their primary tumor location. Left-sided colon cancer (LCC), including rectum and right-sided colon cancer (RCC), is different in pathogeneses, molecular characteristics, incidences and prognoses. In LCC, chromosomal instability has been detected in approximately 75% more than 30% of RCCs [41]. With increased chromosomal instability, LCC has been associated with more frequent overexpression of the epidermal growth factor receptor (EGFR) ligands, EGFR, EREG, AREG, HERS, VEGF-1 and COX-2 [42]. In RCC, Hypermutation is more prevalent. RCC has been shown to be associated with an increase in RAS and phosphoinositide 3-kinase pathway
### Table 3. Correlation of serum biomarkers level or methylation status with clinicopathological characteristics.

| Characteristics          | CK18 | CK20 | HPDL | CLIP4 | CK20 | CLIP4 |
|--------------------------|------|------|------|-------|------|-------|
|                          | Hemi-methylated | Un-methylated | Hemi-methylated | Un-methylated | Hemi-methylated | Un-methylated |
| Age                      | Positive | Negative | Positive | Negative | Positive | Negative |
| ≤50                      | 5       | 4      | 8      | 1      | 7      | 2      |
| >50                      | 27      | 24     | 36     | 16     | 31     | 20     |
|                         | 0.588   | 0.205  | 0.281  | 0.449  | 0.405  | 0.025  |
| Tumor Size (cm)          | Present | Absent | Present | Absent | Present | Absent |
| ≤4                       | 18      | 12     | 25     | 5      | 24     | 6      |
| >4                       | 14      | 16     | 18     | 12     | 14     | 16     |
|                         | 0.219   | 0.042  | 0.007  | 0.614  | 0.577  | 0.167  |
| TNM stage                | I + II  | III + IV | II | I     | III | III + IV |
| Tumor location           | Are     | Rectum | Present | Male | Tumor location | Sex |
| Colon                    | 11      | 11     | 13     | 9      | 12     | 10     |
|                         | 0.45    | 0.212  | 0.057  | 0.04   | 0.24   | 0.000  |
| Tumor Size (cm)          | ≤4      | >4     | 18     | 12     | 22     | 12     |
|                         | 0.013   | 0.203  | 0.017  | 0.009  | 0.407  | 0.001  |
| Differentiation          | Well    | Moderate | Poor | Absent | Present | CEA |
|                         | 6       | 1      | 6      | 1      | 7      | 0      |
|                         | 0.112   | 0.434  | 0.35   | 0.128  | 0.582  | 0.000  |
| Lymphovascular invasion | Absent  | Present | CK20 | CA199  | <5 ng/ml | ≥5 ng/ml |
|                         | 22      | 14     | 25     | 11     | 24     | 12     |
|                         | 0.112   | 0.434  | 0.35   | 0.128  | 0.582  | 0.000  |
|                         | 0.000   | 0.206  | 0.001  | 0.038  | 0.325  | 0.000  |
|                         | 0.001   | 0.046  | 0.018  | 0.2    | 0.2    | 0.001  |
|                         | 0.001   | 0.046  | 0.018  | 0.2    | 0.2    | 0.001  |
|                         | 0.07    | 0.216  | 0.276  | 0.512  | 0.285  | 0.456  |

*P < 0.05 is considered statistically significant.*

### Table 4. Evaluation of serum biomarkers level or methylation status in detection of CRC.

| Markers | Training cohort | Validation cohort |
|---------|-----------------|-------------------|
|         | Sensitivity     | Specificity       | Sensitivity | Specificity |
| CEA     | 53.33%          | 85.71%            | 65.00%      | 83.30%      |
| CA199   | 23.33%          | 92.06%            | 26.90%      | 91.70%      |
| CK18    | 46.67%          | 87.30%            |             |             |
| CK20    | 28.33%          | 90.47%            | 36.70%      | 88.30%      |
| HPDL    | 36.67%          | 93.65%            |             |             |
| CLIP4   | 73.33%          | 84.13%            | 80.00%      | 85.00%      |
| CEA or CLIP4 | 81.67% | 73.02% |
| CA199 or CLIP4 | 80.00% | 77.78% |
| CK20 or CLIP4 | 91.67% | 82.54% |
| HPDL or CLIP4 | 81.67% | 80.95% |
| CK20 or HPDL or CLIP4 | 93.33% | 76.19% |
mutations, \textit{BRAF} mutations, and \textit{TGF\betaR2} mutations. CpG island methylator phenotype (CIMP)-high and microsatellite-high subtype (MSI) have also been detected in RCC [43]. According to our study, in the validation cohort, elevated levels of CK20 were significantly correlated with the tumor location of the colon, not the rectum. Therefore, whether the expression of CK20 in tumor tissues and the serum level of CK20 are different between LCC and RCC and whether serum CK20 could distinguish LCC from RCC need to be further studied. Second, serum CK20 mRNA is a biomarker of circulating CRC cells. Serum CK20 protein originates from circulating CRC cells or CRC tumor tissue, which urgently needs to be determined. Therefore, for serum CK20 protein-positive patients, serum CK20 mRNA should be detected, and CK20 protein in CRC tumor tissues should be examined by IHC. Third, bioinformatics and DNA methlyolics showed that breast and gastric cancer tissues presented hyper-methylated \textit{CLIP4} [44–46]. The diagnostic value of hyper-methylated \textit{CLIP4} in serum for breast cancer and gastric cancer has not yet been reported. Thus, a study involving several cancer types should be conducted to verify the specificity of hyper-methylated \textit{CLIP4} and CK20/hyper-methylated \textit{CLIP4} for CRC diagnosis. Fourth, through clinical serum sample validation, we found that only the combination of CK20 and hyper-methylated \textit{CLIP4} displayed high sensitivity and specificity for CRC diagnosis. The reason is unclear. Therefore, the biological function of CK20 and CLIP4 in CRC and the relationship between them should be further explored. In addition, our study was performed on a limited number of CRC individuals (Only 120 patients were enrolled) from two centers. In the future, a study involving several hospitals/clinics from different regions covering a large population should be conducted to avoid overestimation of the sensitivity and specificity of serum CK20/hyper-methylated \textit{CLIP4}. Finally, although none of the CRC patients had received radiotherapy, chemotherapy or surgery treatment prior to blood collection, they had already been clinically diagnosed by endoscopy and pathological biopsy. Serum biomarkers would be more likely detectable in clinical patients than subclinical patients. Therefore, a large number of blood samples from a health examination center should be collected and serum CK20/hyper-methylated \textit{CLIP4} should be detected. Then, for patients positive for serum CK20 or hyper-methylated \textit{CLIP4} should be examined by endoscopy and pathological biopsy to verify the ability of serum CK20/hyper-methylated \textit{CLIP4} to diagnose CRC.

CONCLUSIONS

In summary, from systematical bioinformatics screening to clinical serum sample validation, this study shows that the combination of serum CK20 and hyper-methylated \textit{CLIP4} is a novel effective biomarker for CRC diagnosis.

Abbreviations

\textit{ASCL2}: Achaete-Scute Family BHLH Transcription Factor 2; \textit{CA}: Colorectal adenomas; \textit{CA125}: Carbohydrate antigen 125; \textit{CA19-9}: Carbohydrate antigen 19-9; \textit{CCLE}: Cancer cell line encyclopedia; \textit{CEA} (\textit{CEACAM5}): Carcinoembryonic antigen; \textit{cfDNA}: Cell free DNA; \textit{CK8} (\textit{KRT8}): Cytokeratin 8; \textit{CK18} (\textit{KRT18}): Cytokeratin 18; \textit{CK20} (\textit{KRT20}): Cytokeratin 20; \textit{CLIP4}: CAP-Gly Domain Containing Linker Protein Family Member 4; \textit{CRC}: Colorectal cancer; \textit{DAC}: 5'-Aza-2'-deoxycytiding; \textit{ELISA}: Enzyme-linked immunosorbent assay; \textit{EPCAM}: Epithelial Cell Adhesion Molecule; \textit{FERMT1}: FERM Domain Containing Kindlin 1; \textit{GAPDH}: Glyceraldehyde-3-Phosphate Dehydrogenase; \textit{GEPIA}: Gene Expression Profiling Interactive; \textit{GTEx}: Genotype-Tissue Expression; \textit{HNF4G}: Hepatocyte Nuclear Factor 4 Gamma; \textit{HPA}: Human Protein Atlas; \textit{HPDL}: 4-Hydroxyphenylpyruvate Dioxygenase Like; \textit{KCN3}: Potassium Voltage-Gated Channel Subfamily E Regulatory Subunit 3; \textit{LGR5}: Leucine Rich Repeat Containing G Protein-Coupled Receptor 5; \textit{MSP}: Methylation-specific PCR; \textit{MUC13}: Mucin 13; \textit{SLC12A2}: Solute Carrier Family 12 Member 2; \textit{TCGA}: The Cancer Genome Atlas.

AUTHOR CONTRIBUTIONS

QG, HT and ZL provided the conception. ZL performed the computational analyses. FZ, YJ, NK, YZ and JW provided the clinical samples. ZL, WZ, JW, LW and XL did the experiments and drafted the paper. QG and HT reviewed the paper. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

We thank all patients for participation in this study.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest related to this study.

FUNDING

This work was funded by the following grants: 2019LCZXXF-XH13, 2020LCZXKF-XH01 and 2021LCZXXF-XH01 from Yunnan Digestive Endoscopy Clinical Medical Center Foundation [2X2019-01-02] for Health Commission of Yunnan
Province; KHBSH-2020-001 from Post-doctor Foundation of The First People’s Hospital of Yunnan Province; 2019FFE001(-173) from Yunnan Provincial Foundation of Science and Technology Department and Kunming Medical University; 2018NS0264 from Yunnan Institute of Gastroenterology; 2018FF001(-049) from Yunnan Provincial Science and Technology Department; 2017-1-S-16759 from Kunming Science and Technology Bureau Program.

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA Cancer J Clin. 2019; 69:7–34. https://doi.org/10.3322/caac.21551 PMID:30620402

2. Grady WM, Carethers JM. Genomic and epigenetic instability in colorectal cancer pathogenesis. Gastroenterology. 2008; 135:1079–99. https://doi.org/10.1053/j.gastro.2008.07.076 PMID:18773902

3. Yu H. Reference intervals for gastrointestinal tumor markers (AFP, CEA, CA199 and CA724) in healthy adults of Han nationality in Chongqing by Roche ECLIA system. Scand J Clin Lab Invest. 2019; 79:484–90. https://doi.org/10.1080/00365513.2019.1661008 PMID:31599689

4. Dong D, Zhang L, Jia L, Ji W, Wang Z, Ren L, Niu R, Zhou Y. Identification of Serum Periostin as a Potential Diagnostic and Prognostic Marker for Colorectal Cancer. Clin Lab. 2018; 64:973–81. https://doi.org/10.7754/Clin.Lab.2018.171225 PMID:29945311

5. Huo YR, Huang Y, Liauw W, Zhao J, Morris DL. Prognostic Value of Carcinoembryonic Antigen (CEA), AFP, CA19-9 and CA125 for Patients with Colorectal Cancer with Peritoneal Carcinomatosis Treated by Cytoreductive Surgery and Intrapерitoneal Chemotherapy. Anticancer Res. 2016; 36:1041–9. PMID:26976996

6. Gao Y, Wang J, Zhou Y, Sheng S, Qian SY, Huo X. Evaluation of Serum CEA, CA19-9, CA72-4, CA125 and Ferritin as Diagnostic Markers and Factors of Clinical Parameters for Colorectal Cancer. Sci Rep. 2018; 8:2732. https://doi.org/10.1038/s41598-018-21048-y PMID:29426902

7. Esteller M, Herman JG. Cancer as an epigenetic disease: DNA methylation and chromatin alterations in human tumours. J Pathol. 2002; 196:1–7. https://doi.org/10.1002/path.1024 PMID:11748635

8. Rex DK, Boland CR, Dominitz JA, Giardiello FM, Johnson DA, Kaltenbach T, Levin TR, Lieberman D, Robertson DJ. Colorectal Cancer Screening: Recommendations for Physicians and Patients from the U.S. Multi-Society Task Force on Colorectal Cancer. Am J Gastroenterol. 2017; 112:1016–30. https://doi.org/10.1038/ajg.2017.14 PMID:28555630

9. Danese E, Montagnana M, Lippi G. Circulating molecular biomarkers for screening or early diagnosis of colorectal cancer: which is ready for prime time? Ann Transl Med. 2019; 7:610. https://doi.org/10.21037/atm.2019.08.97 PMID:32047771

10. Ng JM, Yu J. Promoter hypermethylation of tumour suppressor genes as potential biomarkers in colorectal cancer. Int J Mol Sci. 2015; 16:2472–96. https://doi.org/10.3390/ijms16022472 PMID:25622259

11. Okugawa Y, Grady WM, Goel A. Epigenetic Alterations in Colorectal Cancer: Emerging Biomarkers. Gastroenterology. 2015; 149:1204–25.e12. https://doi.org/10.1053/j.gastro.2015.07.011 PMID:26216839

12. Lee JS. Exploring cancer genomic data from the cancer genome atlas project. BMB Rep. 2016; 49:607–11. https://doi.org/10.5483/bmbrep.2016.49.11.145 PMID:27530686

13. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, Wilson CJ, Lehár J, Kryukov GV, Sonkin D, Reddy A, Liu M, Murray L, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature. 2012; 483:603–7. https://doi.org/10.1038/nature10903 PMID:22460905

14. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res. 2017; 45:W98–102. https://doi.org/10.1093/nar/gkw247 PMID:28407145

15. Colwill K, Gräslund S and Renewable Protein Binder Working Group. A roadmap to generate renewable protein binders to the human proteome. Nat Methods. 2011; 8:551–8. https://doi.org/10.1038/nmeth.1607 PMID:21572409

16. Zweig AS, Karolchik D, Kuhn RM, Haussler D, Kent WJ. UCSC genome browser tutorial. Genomics. 2008; 92:75–84. https://doi.org/10.1016/j.ygeno.2008.02.003 PMID:18514479
17. Chandrashekar DS, Bashel B, Balasubramanyan SAH, Creighton CJ, Ponce-Rodriguez I, Chakravarti BS, Varambally S. UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. Neoplasia. 2017; 19:649–58. https://doi.org/10.1016/j.neo.2017.05.002 PMID:28732212

18. Koch A, De Meyer T, Jeschke J, Van Criekinge W. MEXPRESS: visualizing expression, DNA methylation and clinical TCGA data. BMC Genomics. 2015; 16:636. https://doi.org/10.1186/s12864-015-1847-z PMID:26306699

19. Chen MM, Zhao RC, Chen KF, Huang Y, Liu ZJ, Wei YG, Jian Y, Sun AM, Qin L, Li B, Qin Y. Hypomethylation of CTCFL promoters as a noninvasive biomarker in plasma from patients with hepatocellular carcinoma. Neoplasma. 2020; 67:909–15. https://doi.org/10.4149/neo_2020_190819N789 PMID:32386482

20. Moll R, Zimbelmann R, Goldschmidt MD, Keith M, Laufer J, Kasper M, Koch PJ, Franke WW. The human gene encoding cytokeratin 20 and its expression during fetal development and in gastrointestinal carcinomas. Differentiation. 1993; 53:75–93. https://doi.org/10.1111/j.1432-0436.1993.tb00648.x PMID:8359955

21. Imai Y, Yamagishi H, Fukuda K, Okamura T, Ono Y, Ban S, Inoue T, Ueda Y. Expression of cytokeratin 20 indicates invasive histological phenotype in poorly differentiated colorectal adenocarcinoma. Anticancer Res. 2014; 34:159–67. PMID:24403457

22. Kust D, Šamija I, Kirac I, Kovačević D, Kusić Z. Cytokeratin 20 positive cells in blood of colorectal cancer patients as an unfavorable prognostic marker. Acta Clin Belg. 2016; 71:235–43. https://doi.org/10.1080/17843286.2016.1177264 PMID:27144776

23. Šamija I, Lukač J, Mubrin MK, Kirac I, Kovačević D, Kusić Z. Detection of cytokeratin-20-positive cells in preoperative and postoperative blood samples from colorectal cancer patients by real-time RT-PCR. Int J Biol Markers. 2013; 28:174–81. https://doi.org/10.5301/jbm.5000003 PMID:23558939

24. Molnár B, Galamb O, Kalmár A, Barták BK, Nagy ZB, Tóth K, Tulassay Z, Igaz P, Dank M. Circulating cell-free nucleic acids as biomarkers in colorectal cancer screening and diagnosis - an update. Expert Rev Mol Diagn. 2019; 19:477–98. https://doi.org/10.1080/14737159.2019.1613891 PMID:31046485

25. Auf FA, El-Eibdy GK, Mosaad YM, El-Ghannam DM, El-Masry EE. Cytokeratin 20 and vascular endothelial growth factor as molecular markers in Egyptian patients with colorectal cancer. J Oncol Pharm Pract. 2011; 17:160–7. https://doi.org/10.11177/1078155210365006 PMID:20395351

26. Wong CS, Cheung MT, Ma BB, Pun Hui E, Chan AC, Chan CK, Lee KC, Cheuk W, Lam MY, Wong MC, Chan CM, Chan JK, Chan AT. Isolated tumor cells and circulating CK20 mRNA in pN0 colorectal cancer patients. Int J Surg Pathol. 2008; 16:119–26. https://doi.org/10.11177/1066896907311901 PMID:18387990

27. Tsouma A, Aggeli C, Lembessis P, Zografos GN, Korkolis DP, Pectasides D, Skondra M, Pissimisis N, Tzonou A, Koutras M. Multiplex RT-PCR-based detections of CEA, CK20 and EGFR in colorectal cancer patients. World J Gastroenterol. 2010; 16:5965–74. PMID:21157973

28. Xu D, Li XF, Zheng S, Jiang WZ. Quantitative real-time RT-PCR detection for CEA, CK20 and CK19 mRNA in peripheral blood of colorectal cancer patients. J Zhejiang Univ Sci B. 2006; 7:445–51. https://doi.org/10.1631/jzus.2006.B0445 PMID:16691638

29. Shen C, Hu L, Xia L, Li Y. Quantitative real-time RT-PCR detection for survivin, CEA and CEA in peripheral blood of colorectal cancer patients. Jpn J Clin Oncol. 2008; 38:770–6. https://doi.org/10.1093/jjco/hyn105 PMID:18845519

30. Gaudent P, Livstone MS, Lewis SE, Thomas PD. Phylogenetic-based propagation of functional annotations within the Gene Ontology consortium. Brief Bioinform. 2011; 12:449–62. https://doi.org/10.1093/bib/bbr042 PMID:21873635

31. Jensen SØ, Øgaard N, Nielsen HJ, Bramsen JB, Andersen CL. Enhanced Performance of DNA Methylation Markers by Simultaneous Measurement of Sense and Antisense DNA Strands after Cytosine Conversion. Clin Chem. 2020; 66:925–34. PMID:32058146

32. Jensen SØ, Øgaard N, Ørntoft MW, Rasmussen MH, Bramsen JB, Kristensen H, Mouritzen P, Madsen MR, Madsen AH, Sunesen KG, Iversen LH, Laurberg S, Christensen JJ, et al. Novel DNA methylation biomarkers show high sensitivity and specificity for blood-based detection of colorectal cancer—a clinical biomarker discovery and validation study. Clin Epigenetics. 2019; 11:158.
33. Cao Y, Zhao G, Cao Y, Chen Z, Liu X, Yuan M, Yang J, Wang X, Ma Y, Liu Z, Xiong S, Zheng M, Fei S. Feasibility of Methylated CLIP4 in Stool for Early Detection of Colorectal Cancer: A Training Study in Chinese Population. Front Oncol. 2021; 11:647066. https://doi.org/10.3389/fonc.2021.647066 PMID: 33968748

34. Huang X, Lan Y, Li E, Li J, Deng Q, Deng X. Diagnostic values of MMP-7, MMP-9, MMP-11, TIMP-1, TIMP-2, CEA, and CA19-9 in patients with colorectal cancer. J Int Med Res. 2021; 49:3000605211012570. https://doi.org/10.1177/03000605211012570 PMID: 33942633

35. Pasha HF, Radwan MI, Yehia AM, Toam MM. Circulating methylated RUNX3 and SFRP1 genes as a noninvasive panel for early detection of colorectal cancer. Eur J Gastroenterol Hepatol. 2019; 31:1342–9. https://doi.org/10.1097/MEG.0000000000001532 PMID: 31524773

36. Ivancic MM, Megna BW, Sverchkov Y, Craven M, Reichelderfer M, Pickhardt PJ, Sussman MR, Kennedy GD. Noninvasive Detection of Colorectal Carcinomas Using Serum Protein Biomarkers. J Surg Res. 2020; 246:160–9. https://doi.org/10.1016/j.jss.2019.08.004 PMID: 31586988

37. Fan CW, Kuo YB, Lin GP, Chen SM, Chang SH, Li BA, Chan EC. Development of a multiplexed tumor-associated autoantibody-based blood test for the detection of colorectal cancer. Clin Chim Acta. 2017; 475:157–63. https://doi.org/10.1016/j.cca.2017.10.022 PMID: 29074220

38. Huang G, Wei B, Chen Z, Wang J, Zhao L, Peng X, Liu K, Lai Y, Ni L. Identification of a four-microRNA panel in serum as promising biomarker for colorectal carcinoma detection. Biomark Med. 2020; 14:749–60. https://doi.org/10.2217/bmm-2019-0605 PMID: 32672054

39. Tan Y, Lin JJ, Yang X, Gou DM, Fu L, Li FR, Yu XF. A panel of three plasma microRNAs for colorectal cancer diagnosis. Cancer Epidemiol. 2019; 60:67–76. https://doi.org/10.1016/j.canep.2019.01.015 PMID: 30925282

40. Wang D, Yuan W, Wang Y, Wu Q, Yang L, Li F, Chen X, Zhang Z, Yu W, Maimela NR, Cao L, Wang D, Wang J, et al. Serum CCL20 combined with IL-17A as early diagnostic and prognostic biomarkers for human colorectal cancer. J Transl Med. 2019; 17:253. https://doi.org/10.1186/s12976-019-2008-y PMID: 31387598

41. Shen H, Yang J, Huang Q, Jiang MJ, Tan YN, Fu JF, Zhu LZ, Fang XF, Yuan Y. Different treatment strategies and molecular features between right-sided and left-sided colon cancers. World J Gastroenterol. 2015; 21:6470–8. https://doi.org/10.3748/wjg.v21.i21.6470 PMID: 26074686

42. Huang X, Liu J, Mo X, Liu H, Wei C, Huang L, Chen J, Tian C, Meng Y, Wu G, Xie W, P C FJ, Liu Z, Tang W. Systematic profiling of alternative splicing events and splicing factors in left- and right-sided colon cancer. Aging (Albany NY). 2019; 11:8270–93. https://doi.org/10.18632/aging.102319 PMID: 31586988

43. Stintzing S, Tejpar S, Gibbs P, Thiebach L, Lenz HJ. Understanding the role of primary tumour localisation in colorectal cancer treatment and outcomes. Eur J Cancer. 2017; 84:69–80. https://doi.org/10.1016/j.ejca.2017.07.016 PMID: 28787661

44. Hu C, Zhou Y, Liu C, Kang Y. A novel scoring system for gastric cancer risk assessment based on the expression of three CLIP4 DNA methylation-associated genes. Int J Oncol. 2018; 53:633–43. https://doi.org/10.3892/ijo.2018.4433 PMID: 29901187

45. Chong Y, Mia-Jan K, Ryu H, Abdul-Ghafar J, Munkhdelger J, Lkhagvadorj S, Jung SY, Lee M, Ji SY, Choi E, Cho MY. DNA methylation status of a distinctively different subset of genes is associated with each histologic Lauren classification subtype in early gastric carcinogenesis. Oncol Rep. 2014; 31:2535–44. https://doi.org/10.3892/or.2014.3133 PMID: 24737029

46. Fan Y, He L, Wang Y, Fu S, Han Y, Fan J, Wen Q. CLIP4 Shows Putative Tumor Suppressor Characteristics in Breast Cancer: An Integrated Analysis. Front Mol Biosci. 2021; 7:616190. https://doi.org/10.3389/fmolb.2020.616190 PMID: 33575272
**SUPPLEMENTARY MATERIALS**

**Supplementary Tables**

**Supplementary Table 1. 100 genes specifically overexpressed in CRC tissue.**

| Gene       | Description |
|------------|-------------|
| CEACAM5    |             |
| GPX2       |             |
| RNF43      |             |
| KRT18      |             |
| UBD        |             |
| CST1       |             |
| CCL20      |             |
| ASCL2      |             |
| FERMT1     |             |
| NOX1       |             |
| OLFM4      |             |
| CYP2S1     |             |
| EPCAM      |             |
| VIL1       |             |
| MMP1       |             |
| B3GNT3     |             |
| IGHG4      |             |
| CCL24      |             |
| MISP       |             |
| MUC13      |             |
| SPINK4     |             |
| CDX2       |             |
| CDX1       |             |
| CDH17      |             |
| IHH        |             |
| RP11-357H14.17 |       |
| ATP10B     |             |
| MMP12      |             |
| TSPAN8     |             |
| PRR15      | C6orf223    |
| KRT8       | UBE2SP2     |
| MUC3A      |             |
| EPHB2      |             |
| GUCY2C     |             |
| PL51       |             |
| NFE2L3     |             |
| LG5R       |             |
| CTD-2377D24.6 |           |
| COL10A1    |             |
| GMDS       |             |
| AP000349.2 |             |
| KRT20      | C021218.2   |
| GPA33      |             |
| PPP1R14D   |             |
| ACSL5      |             |
| MEP1A      |             |
| LINC01559  |             |
| BCL2L15    |             |
| RP11-150012.3 |          |
| C6orf222   |             |
| GALNT4     |             |
| HDPI       |             |
| TRPM2-AS   |             |
| RP5-881L22.5 |           |
| SLC12A2    |             |
| AP0003774.1 |             |
| CASC9      |             |
| KCNE3      |             |
| HNF4G      |             |
| C2CD4A     |             |
| TRIM15     |             |
| APOBEC1    |             |
| FAM111B    |             |
| LINC01207  |             |
| GPR35      |             |
| LINC00483  |             |
| RP11-304L19.1 |          |
| GAL5S4     |             |
| RP11-474D1.3 |           |
| ABHD11-AS1 |             |
| EFA6       |             |
| CALML4     |             |
| AP000439.3 |             |
| CASC21     |             |
| EPSS8L3    |             |
| AC005355.2 |             |
| HTR1D      |             |
| FUT4       |             |
| AL3ST2     |             |
| TMEM211    |             |
| HOXB-AS4   |             |
| UC5B       |             |
| CXC3       |             |
| RP11-234B24.2 |           |
| DEFA5      |             |
| ATF2       |             |
| KRT8P11    |             |
| FEZF1-AS1  |             |
| RPI11-187E13.1 |          |
| RN7SKP54   |             |
| POUSF1B    |             |
| SRMS       |             |
| AC016735.1 |             |
| TLX1       |             |
| XFP4       |             |

**Supplementary Table 2. 74 genes specifically overexpressed in CRC cell lines.**

| Gene       | Description |
|------------|-------------|
| CEACAM5    |             |
| GPX2       |             |
| RNF43      |             |
| KRT18      |             |
| UBD        |             |
| CST1       |             |
| ASCL2      |             |
| FERMT1     |             |
| NOX1       |             |
| OLFM4      |             |
| CYP2S1     |             |
| VIL1       |             |
| B3GNT3     |             |
| IGHG4      |             |
| CCL24      |             |
| MISP       |             |
| MUC3A      |             |
| EPHB2      |             |
| GUCY2C     |             |
| PL51       |             |
| NFE2L3     |             |
| LG5R       |             |
| CTD-2377D24.6 |           |
| COL10A1    |             |
| GMDS       |             |
| AP000349.2 |             |
| KRT20      | C021218.2   |
| GPA33      |             |
| PPP1R14D   |             |
| ACSL5      |             |
| MEP1A      |             |
| BCL2L15    |             |
| RP11-150012.3 |          |
| C6orf222   |             |
| GALNT4     |             |
| HDPI       |             |
| TRPM2-AS   |             |
| RP5-881L22.5 |           |
| SLC12A2    |             |
| AP0003774.1 |             |
| CASC9      |             |
| KCNE3      |             |
| HNF4G      |             |
| C2CD4A     |             |
| TRIM15     |             |
| APOBEC1    |             |
| FAM111B    |             |
| LINC01207  |             |
| GPR35      |             |
| LINC00483  |             |
| RP11-304L19.1 |          |
| GAL5S4     |             |
| RP11-474D1.3 |           |
| ABHD11-AS1 |             |
| EFA6       |             |
| CALML4     |             |
| AP000439.3 |             |
| CASC21     |             |
| EPSS8L3    |             |
| AC005355.2 |             |
| HTR1D      |             |
| FUT4       |             |
| AL3ST2     |             |
| TMEM211    |             |
| HOXB-AS4   |             |
| UC5B       |             |
| CXC3       |             |
| RP11-234B24.2 |           |
| DEFA5      |             |
| BATF2      |             |
| KRT8P11    |             |
| RP11-187E13.1 |          |
| RN7SKP54   |             |
| POUSF1B    |             |
| SRMS       |             |
| AC016735.1 |             |
| TLX1       |             |
| XFP4       |             |

**Supplementary Table 3. 16 CRC specific overexpressed genes which encoding secreted proteins.**

| Gene       | Description |
|------------|-------------|
| KRT18      |             |
| CST1       |             |
| OLFM4      |             |
| EPCAM      |             |
| CCL24      |             |
| SPINK4     |             |
| IHH        |             |
| KRT8       |             |
| MUC3A      |             |
| DEFA6      |             |
| MUC5B      |             |
| DEFA5      |             |
| CEACAM5    |             |
| MUC13      |             |
| KRT20      |             |
| LGALS4     |             |

**Supplementary Table 4. 19 CRC specific overexpressed genes possessed CpG islands in their promoters or the first exon region.**

| Gene       | Description |
|------------|-------------|
| KRT18      |             |
| CYP2S1     |             |
| B3GNT3     |             |
| CDX2       |             |
| CDX1       |             |
| IHH        |             |
| PRR15      |             |
| KRT8       |             |
| EPHB2      |             |
| NFE2L3     |             |
| GMDS       |             |
| SRMS       |             |
| ASCL2      |             |
| FERMT1     |             |
| LG5R       |             |
| HPDL       |             |
| SLC12A2    |             |
| KCNE3      |             |
| HNF4G      |             |
### Supplementary Table 5. Top 250 underexpressed genes in CRC tissue.

CRC vs. Colon and Rectal normal tissue (log2 fold < -1, $P$ value < 0.01)

| Gene          | Description            |
|---------------|------------------------|
| TMBIM4        | ARHGEF17 COL6A2 GFRAI1 TXLNGY MAOB MBNL1-AS1 CES1 PODN GOLGA8 ACACB CA4 HMCN2 SFRRP1 ITGA5 FBXL2 SLC22A7 CILP PHYHD1 PCSK1N CLDN5 CLIP4 IL11RA GREM2 GABBR1 PTGSI MAP6 CLMP REEP1 ZEB1 GPRASp1 RP11-602.3 STON1 MRV11 ZG16 WISP2 CHGA EPHA7 C15orf52 SGCN APOD MFAP4 EBF4 MIR4697G PALM MEG3 CSPG4 STAR9 TN52 PDLIM4 PRRT2 ABI3BP DDR2 BHM2T2 AKAP12 NKX2-3 PHLD2B ORSI AC002398.12 COL6A1 MEIS2 LINC01573 TGFBR11 KDM5D GSTM1 TSPY2L CCDC136 CAV1 S100B ARHGEF26 APBB1 PTGIS FGFRI MYOCD CSRP1 RERG DPP6 CLIP3 RP11-166D19.1 MAG2-AS3 UCHL1 CALD1 ANK2 PYGM DNGA TMOD1 USP32P1 PDZD4 PER1 PDLIM7 GRIK5 LICAM DTNA PLEKHO1 AGTR1 BOC TLN2A10 C20orf166-AS1 ADCY5 OGN RP11-286H15.1 ABCA8 SLC2A4 ADHFE1 SPARC1 CFD RAMP1 FLNA LDB3 LVE1 SCAR5 CDDC69 MAL PDK4 RPS4Y1 UBXN10-AS1 FAM129A MEIS1 NNT TXN2C7N7A KCNA11 RGMAM3 SRPX AP000892.6 THBS4 NCA14 CCL14 CHRD1L SCGA AO3 PRPH ASB2 MGP TPM2 MASP1 SORBS1 MRGPRF MYLK AL442127.1 GNAO1 JPH2 C2orf40 FXYD6 CASQ2 PLXNB3 CACNA1H GSTM2 MAMDC2 ANGPTL1 FNG FABP4 PLP1 TGLLN CA1 MORN5 KCNB1 ARHGEF25 TNS1 CORO6 ATP1A2 PRMA1 MIR143HG RNASSP216 HIF3A PLN TCEL2 CRYP LG14 PRELP LMO3 ITGA7 PPP1R1A4A GPX7 HSPB7 POPDC2 LIMS2 MAB21L2 MYL9 FXYD1 PLIN4 PGM5 HANDF PNCK RBFOX3 HAND2-AS1 PGM5-AS1 TACR2 C7 FHL1 DPT RBPMS2 HAND2 HSPB8 PSD SYNPO2 SPEG LMOD1 ADAM33 HSPB6 CNN1 RP11-39404.5 ADH1B SYNM ACTG2 MYH11 DES |

### Supplementary Table 6. Top 250 promoter hyper-methylated genes in CRC tissue.

CRC vs. Colon and Rectal normal tissue (Beta value: 0.7–0.5, $P$ value < 0.01)

| Gene          | Description            |
|---------------|------------------------|
| VWC2 DPP10 EFCC1 KDHRBS2 FG53 ADHFE1 PTPRT FAIM2 EYA4 DOK5 PREX2 GATA5 CMTM3 COL4A1 COL4A2 GFRAI1 SNAP1P1 DPY1L92 CMTM2 DKFZP434H168 SYT9 RALYL FBX2 FAM194A SPATA32 INA ITGA8 ADAMT5 FIGN PRDM14 CNKSR1 RAB6C IRX4 GALTN13 CFAP100 MARCH11 NALCN DYDC1 DYDC2 ST8SIA5 CHST10 LOCL283392 TRHDE GPR26 ZNF582 BARHL2 MIR129-2 OTX2-AS1 LRFN3 PCDH10 EID3 PTPN5 FGFI2 FOXE1 MIR137 DOK6 CPXM2 MMP32 SORCS3 FNDC1 CRHR2 CRFRP GALNT6 GSC2 GRIK4 ALC3L9A5 TME1M13C NPWR1 NR2E1 CCDC178 GRWD1 CHLI SOX17 CHRD1L ST6GALNAC5 SLC16A12 COL26A1 RYR2 WDR86 SALL1 PCDH8 FGFI8 GRK3 SORCS1 MMP16 GMC2 PTF1A DMRT3 ORPK1 UNC80 DBX2 LAMA1 DRD4 TFP2L SDC2 SIM1 CIDEA ADGRA1 COL23A1 PDLIM4 VIP2 CCA2 __BRIP1 SLTRK4 ZNF814 VAV3 ZNF132 LOCL15762 EXSI1 MXS SLC6A3 HCNI PLPP4 SHISAH BHLHE23 CLDN10 NPY GABBR2 VX51 ZNF75A SH3GL3 NR0B1 ADAM12 FOX12 GPR101 TCEA2 FOXD2 FOXD2-AS1 NRK MYOD1 ZNF549 ADRA1A VSTM2B C1QL2 CFAP46 ITIH5 PAX7 PHOX2A SALL3 OLG2 LINCO0472 ZNF334 HOX555 POU4F2 MSC FUT9 GAD2 EEFEMP1 FAM19A5 HNF4G MDFIC NXP2H2 SHISAH2 EFS DNH11 LDC14 TCERG1L GRIK2 ACDY4 ZNF585 SLC18A3 SLC4A11 SOX3 SLC32A1 HOXA2 TWIST1 TRPC6 NID1 FG3 ZNF667 FAM162B BTG4 ZNF415 NID2 CLIP4 LOCL100SSS949 GPR38 SLC6A2 CD01 GJD2 NELL2 MIR34 MTE198 CRISP5L1 ADCYP1 UTT1 PCDH17 CUX2 VEGFC STX1B HOXDL1 TSP2R2 GHSR CYP7B1 PTH2R RXRG SLC6A5 ADAMTS16 EXOC1L2 NEFH PNTX2 SIX6 SULT4A1 TME1M108 TBNX20 ANKRD20A8P GPR27 UCHL1 FBX1 GALR2 TBX5 DRD5 ZFEG4 JAM3 TLX3 ZNF542P AJP1 SYNDIG1 CCNA1 NCA2 CDHC3 KDR WNK3 ADCYS DGF1 PDNP EML3 HS3ST3A FREM3 DTX2 C2H3 KCNG1 FBXO39 NGF ACSS3 TME1M132D CLVS2 HTR1A TDRD10 SHE TMEM26 GRM7 CCDC181 GSTM2 STAC FAM218A SLC6A11 USP44 |