Identification of Methylation Driven Biomarkers for Diagnosis and Prognosis in Colon Cancer by Integrative Analysis of TCGA, GTEx, and GEO Data

Lichao Cao
Northwest University

Erfei Chen
Northwest University

Jin Yang (✉ yangjin@nwu.edu.cn)
Northwest University  https://orcid.org/0000-0001-5394-2366

Original Article

Keywords: colorectal cancer, methylation driven genes, diagnosis biomarkers, prognosis biomarkers

DOI: https://doi.org/10.21203/rs.3.rs-179363/v1

License: ☺️ ☕️ This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

**Background:** The intention of the present work was to investigate methylation driven biomarkers for diagnosis and prognosis in colorectal cancer (CRC) by integrative analysis of DNA methylation and gene expression data from The Cancer Genome Atlas (TCGA), The Genotype-Tissue Expression (GTEx), and Gene Expression Omnibus (GEO).

**Methods and Results:** The differentially expression genes (DEGs) and differentially methylated genes (DMGs) were screened using mRNA expression and DNA methylation data from TCGA, respectively. And the methylation driven genes (MDGs) of CRC were further identified using MethylMix R package. Subsequently, the MDGs were underwent Random Forest (RF) analyses to establish diagnosis prediction model using mRNA expression data from TCGA and GTEx, which were then validated by GSE39582 from GEO dataset. In addition, prognostic biomarkers that were used to establish the methylation related risk score model was generated by the univariate and multivariate Cox regression analysis.

9 out of 10 DMGs revealed excellent performance as independent diagnostic predictors, including **CLDN1, EPHX4, TCN1, ARHGAP20, LY6G6D, FAM150A, KCNJ12, KRT7, and STK33.** Furthermore, **STK33 and EPHX4** were found to be associated with Overall survival (OS).

**Conclusions:** Our findings suggest that the identified MDGs could be potential biomarkers for diagnosis and prognosis of CRC.

Introduction

CRC is one of the most prevalent malignant tumors and the second leading cause of mortality worldwide.[2] Despite the great progress made in diagnosis and therapy, CRC patients usually develop recurrence and metastasis, leading to the 5-year survival rate dramatically decreased.[23] Therefore, it is urgent to improve the CRC diagnosis, treatment and prognosis.

Molecular characterization is extensively utilized to predict tumor diagnosis, treatment and prognosis, which offers great potential for improving understanding of tumor development.[5, 27] Previous studies on epigenetics have indicated that aberrant DNA methylation could led to gene silencing,[14, 25] which may promoted oncogenesis.[6, 13] DNA methylation related biomarkers for diagnosis and prognosis have been researched in various tumors, such as endometrial cancer,[32] prostate cancer,[18] and hepatocellular carcinoma.[30] MethylMix, a R package has been verified to be able to identify MDGs by integrative analyze DNA methylation and gene expression data from normal and tumor samples,[3, 9] and the relationship between MDGs and the prognosis of CRC patients has been investigated.[4, 28] However, individualized diagnosis and prognosis models for CRC based on MDGs have not been reported.

On the basis of existing literature data, we carried out studies in an effort to identify a set of MDGs using DNA methylation and gene expression data, and construct independent diagnosis model using RF algorithm. Additionally, the MDGs, presumed as potential prognostic indicators, were analyzed using
univariate and multivariate Cox regression. Our findings may suggest that the potential methylation driven biomarkers could prompt a more individualized diagnosis and therapies for CRC.

**Materials And Methods**

**TCGA and GTEx Data acquisition**

The mRNA expression and DNA methylation data (Illumina Human Methylation 450) of CRC from TCGA and the mRNA expression of normal samples from GTEx were downloaded from the UCSC Xena (https://xenabrowser.net/datapages/). Subsequently, the samples from TCGA were divided into normal and tumor group. Finally, we selected 667 samples (51 normal and 616 tumor) for DEGs analysis, 433 samples (45 normal and 388 tumor) for MDGs analysis, detailed information was shown in Table 1. In addition, we selected mRNA expression data of colon tissue from GTEx to provide supplementary data for normal samples from TCGA, and 307 normal samples were obtained.
Table 1
Clinical information of TCGA and GEO dataset

| Sample details for DEGs | Sample details for DMGs | GSE39582 for validation |
|------------------------|------------------------|-------------------------|
| Tumor                  | Normal                 | Overall                 | Tumor | Normal | Overall | Tumor | Normal | Overall |
| (N = 616)              | (N = 51)               | (N = 667)               | (N = 388) | (N = 45) | (N = 433) | (N = 566) | (N = 19) | (N = 585) |

Gender
- female: 288 (Tumor), 28 (Normal), 316 (Overall), 179 (Tumor), 21 (Normal), 200 (Overall), 256 (Tumor), 7 (Normal), 263 (Overall)
- male: 328 (Tumor), 23 (Normal), 351 (Overall), 209 (Tumor), 24 (Normal), 233 (Overall), 310 (Tumor), 12 (Normal), 322 (Overall)

Age
- <=60: 192 (Tumor), 14 (Normal), 206 (Overall), 148 (Tumor), 11 (Normal), 159 (Overall), 157 (Tumor), 3 (Normal), 160 (Overall)
- >60: 424 (Tumor), 37 (Normal), 461 (Overall), 240 (Tumor), 34 (Normal), 274 (Overall), 409 (Tumor), 16 (Normal), 425 (Overall)

Pathologic Stage
- I: 103 (Tumor), 8 (Normal), 111 (Overall), 53 (Tumor), 5 (Normal), 58 (Overall), 33 (Tumor), 5 (Normal), 38 (Overall)
- II: 228 (Tumor), 24 (Normal), 252 (Overall), 144 (Tumor), 21 (Normal), 165 (Overall), 264 (Tumor), 7 (Normal), 271 (Overall)
- III: 178 (Tumor), 9 (Normal), 187 (Overall), 119 (Tumor), 10 (Normal), 129 (Overall), 205 (Tumor), 5 (Normal), 210 (Overall)
- IV: 86 (Tumor), 9 (Normal), 95 (Overall), 52 (Tumor), 9 (Normal), 61 (Overall), 60 (Tumor), 0 (Normal), 60 (Overall)
- Missing: 21 (Tumor), 1 (Normal), 22 (Overall), 20 (Tumor), 0 (Normal), 20 (Overall), 4 (Tumor), 2 (Normal), 6 (Overall)

AJCC-T
- T1: 19 (Tumor), 2 (Normal), 21 (Overall), 10 (Tumor), 0 (Normal), 10 (Overall), 11 (Tumor), 1 (Normal), 12 (Overall)
- T2: 104 (Tumor), 7 (Normal), 111 (Overall), 54 (Tumor), 5 (Normal), 59 (Overall), 45 (Tumor), 4 (Normal), 49 (Overall)
- T3: 421 (Tumor), 36 (Normal), 457 (Overall), 271 (Tumor), 37 (Normal), 308 (Overall), 367 (Tumor), 12 (Normal), 379 (Overall)
- T4: 69 (Tumor), 6 (Normal), 75 (Overall), 50 (Tumor), 3 (Normal), 53 (Overall), 119 (Tumor), 0 (Normal), 119 (Overall)
- Missing: 3 (Tumor), 0 (Normal), 3 (Overall), 3 (Tumor), 0 (Normal), 3 (Overall), 24 (Tumor), 2 (Normal), 26 (Overall)

AJCC-M
- M0: 455 (Tumor), 35 (Normal), 490 (Overall), 263 (Tumor), 27 (Normal), 290 (Overall), 482 (Tumor), 17 (Normal), 499 (Overall)
- M1: 85 (Tumor), 9 (Normal), 94 (Overall), 51 (Tumor), 9 (Normal), 60 (Overall), 61 (Tumor), 0 (Normal), 61 (Overall)
- MX: 66 (Tumor), 6 (Normal), 72 (Overall), 66 (Tumor), 8 (Normal), 74 (Overall), 3 (Tumor), 0 (Normal), 3 (Overall)
- Missing: 10 (Tumor), 1 (Normal), 11 (Overall), 8 (Tumor), 1 (Normal), 9 (Overall), 20 (Tumor), 2 (Normal), 22 (Overall)

Abbreviation: AJCC, American Joint Committee on Cancer.
| Sample details for DEGs | Sample details for DMGs | GSE39582 for validation |
|-------------------------|-------------------------|--------------------------|
| Tumor                   | Normal                  | Overall                  |
| (N = 616)               | (N = 51)                | (N = 667)                |
| Tumor                   | Normal                  | Overall                  |
| (N = 388)               | (N = 45)                | (N = 433)                |
| Tumor                   | Normal                  | Overall                  |
| (N = 566)               | (N = 19)                | (N = 45)                 |
| Overall                 |                         |                         |
| (N = 667)               |                         | (N = 433)                |
| (N = 433)               |                         | (N = 566)                |
| (N = 566)               |                         | (N = 45)                 |

| AJCC-N                  |                         |                         |
| N0                      | 349                     | 34                       | 383                     |
| N0                      | 211                     | 28                       | 239                     |
| N0                      | 302                     | 12                       | 314                     |
| N1                      | 147                     | 8                        | 155                     |
| N1                      | 101                     | 10                       | 111                     |
| N1                      | 134                     | 3                        | 137                     |
| N2                      | 116                     | 9                        | 125                     |
| N2                      | 72                      | 7                        | 79                      |
| N2                      | 98                      | 2                        | 100                     |
| Missing                 | 4                       | 0                        | 4                       |
| Missing                 | 4                       | 0                        | 4                       |
| Missing                 | 32                      | 2                        | 34                      |

Abbreviation: AJCC, American Joint Committee on Cancer.

**Geo Data Acquisition**

We downloaded the gene expression matrix of GSE39582 based on platform GPL570 from GEO database (https://www.ncbi.nlm.nih.gov/geo/), with a total of 566 tumor and 19 normal samples (Table 1).

**Screening Degs, Dmgs And Mdgs**

The DEGs were identified by comparison tumor with normal groups using the limma R package,[22] with the cutoff criteria defined as $|\log_2 \text{fold change (FC)}| > 0.585$ and $P < 0.05$. In addition, $|\logFC| > 0.3$ and $P < 0.05$ was used to screen DMGs using ChAMP R package.[24]

Subsequently, the tumor methylation matrix, normal methylation matrix and tumor expression matrix with the overlapping genes of DEGs and DMGs were constructed to identify MDGs using R package MethylMix.[3, 9] The significant methylation events were filtered using the cutoff criteria set as correlation coefficient $< -0.3$ and $P < 0.05$, and the correlation of the MDGs were visualized by corrplot R package. Moreover, the expression differences of the MDGs between tumor and normal samples were statistically analyzed using unpaired $t$ test.

**Identification And Validation Of Diagnosis Biomarkers Using Rf Algorithm**

The mRNA expression datasets from TCGA and GTEx were integrated and randomly divided into training dataset and validation dataset with a 3:1 ratio in tumor and normal samples. The 10 MDGs were evaluated as independent indicators and combined indicator for diagnosis of CRC using RF algorithm. In addition, The GSE39582 dataset were analyzed as validation dataset combined with training dataset.
from TCGA and GTEx using 7 of the MDGs, owing to the absence of the other 3 MDGs in GSE39582. The receiver operating characteristic (ROC) curve was utilized to depict the sensitivity and specificity of diagnosis models using independent MDG or the combined MDGs panel.

### Identification And Evaluation Of Prognosis Biomarkers

The mRNA expression data with OS information were select to determine the prognostic biomarkers. The univariable Cox proportional hazards regression model was performed using Survival R package, using \( P \)-value cutoff of 0.05. Subsequently, the Kaplan-Meier survival analysis and multivariate Cox regression analysis were performed to determine independent prognostic factors. Accordingly, the coefficient of MDGs was obtained from the multivariate Cox results. The risk score based on the signature of each CRC patients was calculated using the following formula:

\[
\text{MDGs risk score} = \sum \text{Cox coefficient of gene } \chi_i \times \text{scale expression value of gene } \chi_i
\]

To further evaluate the predictive efficiency of the constructed MDGs risk score model, we performed the ROC curve to reflect the sensitivity and specificity of survival prediction, and quantified the area under the curve (AUC) using survivalROC R package. The optimal cutoff risk score was designated at the turning point of ROC curve, where the difference between true positive and false positive is the most significant. The patients above the cutoff value were in the high-risk group, while patients below it in the low-risk group. In addition, Kaplan-Meier curves were plotted to distinguish the two groups using survminer R package.

### Results

### Identification of MDGs as candidate diagnostic and prognostic biomarkers

The whole data generation and analysis workflow was shown in Fig. 1. Expression and DNA methylation data from TCGA were separately analyzed to screen DEGs and DMGs by comparison between normal and tumor samples. As a result, 529 DEGs (273 upregulated and 256 downregulated) were identified (Fig. 2a and Table S1). 91 overlapping genes of DEGs and DMGs were used for the MethylMix analysis with a filter criterion set as \( Cor < -0.3, \| \log2FC \| > 0.3 \) and \( P < 0.05 \), and 10 MDGs were identified as candidate diagnostic and prognostic biomarkers (Fig. S1). Pearson's correlation test was statistically analyzed the correlations between methylation degree and gene expression of the MDGs using the cor.test function of the R language (https://www.r-project.org/) (Fig. S2). It was found that the methylation degree of those 10 MDGs was negatively correlated with their expression. The higher the methylation degree was, the lower the gene expression, and the detailed information of MDGs was shown Table 2. To quantify the difference of each candidate biomarker between normal and tumor samples, the unpaired \( t \) test was conducted. As a result, significant difference was found in all the candidate
Evaluation Of The Mdgs As Candidate Diagnostic Biomarkers

We investigated the potential of using the MDGs for prediction of diagnosis in samples from TCGA and GTEx, 9 out 10 MDGs revealed excellent performance as independent diagnostic predictors, with the area under curve (AUC) of CLDN1 is 0.988, EPHX4 is 0.972, TCN1 is 0.962, ARHGAP20 is 0.929, LY6G6D is 0.888, FAM150A is 0.863, KCNJ12 is 0.858, KRT7 is 0.852, and STK33 is 0.847 (Fig. 3a-i). Additionally, the performance of the combined MDGs were further assessed in samples from TCGA-GTEx and GEO, and the outcome of AUC is 1.0 and 0.996, respectively (Fig. 3k, l). The importance of each predictor for the
combined predictive model was shown in Table 3. The larger the number of \textit{MeanDecreaseAccuracy} and \textit{MeanDecreaseGini}, the more important of the predictor for the model. It was found that the number of \textit{MeanDecreaseAccuracy} and \textit{MeanDecreaseGini} was positively correlated with the AUC of the predictive model of the predictor. Moreover, unsupervised hierarchical clustering of the combined panels with ten makers or seven makers indicated that the constructed models could distinguish CRC patients from normal controls accurately (Fig. 2d-g). Collectively, those 10 MDGs could be candidate biomarkers for diagnosis of CRC samples, and 9 of them performed well as independent indicator.

Table 3
The importance of diagnostic biomarkers

| Predict factor | Normal   | Tumor    | MeanDecreaseAccuracy | MeanDecreaseGini |
|----------------|----------|----------|-----------------------|------------------|
| CLDN1          | 0.290838842 | 0.104256821 | 0.172167458          | 115.7234309     |
| EPHX4          | 0.167970015 | 0.042270934 | 0.088001111          | 87.2102219      |
| TCN1           | 0.080559445 | 0.012116994 | 0.036887766          | 49.6403041      |
| ARHGAP20       | 0.124155382 | 0.032080488 | 0.065622355          | 41.5903612      |
| FAM150A        | 0.013360171 | 0.005650185 | 0.008431127          | 14.7855398      |
| LY6G6D         | 0.049046424 | 0.001993012 | 0.019240782          | 10.3220215      |
| KRT7           | 0.017936895 | 0.003898411 | 0.008980659          | 6.7563074       |
| STK33          | 0.037995918 | 0.005631629 | 0.017341714          | 6.6510197       |
| KCNJ12         | 0.059687348 | 0.002902918 | 0.0235857           | 6.5806269       |
| SPTBN5         | 0.001254881 | 0.000339465 | 0.000673382          | 0.7222647       |

The ‘MeanDecreaseAccuracy’ and ‘MeanDecreaseGini’ of each indicator was produced by the importance function in Random Forest model.

Construction and evaluation of MDGs related prognostic model in CRC patients

Univariate Cox regression analysis revealed that 2 of 10 MDGs were independent prognostic indicators of OS, with the hazard ratio of \textit{STK33} was 1.09 (95 % CI: 1.01–1.17, p = 0.021) and \textit{EPHX4} was 0.91 (CI: 0.83–0.99, p = 0.032) (Fig. 4a). Subsequently, Kaplan-Meier analysis with log-rank test using \textit{STK33} and \textit{EPHX4} as independent prognostic indicator, the results indicated that patients with high-risk scores suffered poor OS, with the \textit{P-value} was 0.008 and 0.047, respectively. Moreover, we calculated the risk score of each CRC patient using the formula as follows: (1.09 × \textit{STK33}) + (0.91 × \textit{EPHX4}). And then CRC patients (n=635) were categorized into high or low score group, according to the optimal cut-off value of the risk score obtained from survminer R package. And the results also indicated that high-risk score patients had a worse OS than those of low-risk score patients (\textit{P}< 0.0001) (Fig. 4d). The prognostic
accuracy of the risk score model was investigated as a continuous variable (Fig. 4e). The AUC of the prognostic model for OS was 0.569 at 3 years, 0.633 at 4 years, and 0.626 at 5 years.

**Discussion**

In present study, we identified 10 MDGs that could be considered as potential diagnostic and prognostic predictors for CRC samples, and seven of those MDGs have been reported to be associated with cancers as following:

1) **CLDN1** (Claudin 1) is a member of the claudin family and tight junction-related proteins, which were demonstrated to be associated with dysfunction or abnormal expression in various tumors.[1, 21] In addition, a previous study revealed that aberrant expression of **CLDN1** regulated AMPK/STAT1/ULK1 signaling pathway, lead to the promotion of proliferation and metastasis in esophageal squamous cancer.[29] Moreover, **CLDN1** was experimentally demonstrated remarkably upregulated in CRC patients and could be considered as methylated diagnostic biomarker in CRC and normal control groups. [17]

2) **EPHX4** (Epoxide Hydrolase 4) is a protein coding gene, which was significantly upregulated in rectal cancer confirmed by immunohistochemistry.[8] Additionally, a previous study performed exon array analysis from pseudomyxoma peritonei and normal colonic epithelia and found that **EPHX4** was up-regulated in pseudomyxoma peritonei. Our study also demonstrated that **EPHX4** was up-regulated in CRC tumor samples. However, no previous research reported that **EPHX4** was associated with CRC.

3) **TCN1** (Transcobalamin I) encoded a member of the vitamin B12-binding protein family that transports of cobalamin into cells. A growing number of studies have verified that the overexpression of **TCN1** obviously associated with tumor invasion and metastasis in CRC.[7, 19] Moreover, it was experimentally demonstrated that **TCN1** was significantly overexpressed in colon cancer tissues compared with normal controls at mRNA and protein level, and it could be considered as potential prognostic biomarker of colon cancer.[20]

4) **LY6G6D** (Lymphocyte Antigen 6 Family Member G6D) belongs to a cluster of leukocyte antigen-6 genes, which was conspicuously overexpressed (around 15-fold) and considered as a promising biomarker of immunotherapy for human Microsatellite stable CRC.[10]

5) **FAM150A** (family with sequence similarity 150A) can active lymphoma kinase (**ALK**) by binding to its extracellular domain,[11, 15] and **ALK** has been used as effective biomarkers in various human cancers, such as neuroblastoma and non-small cell lung cancer.[12] Additionally, the DNA methylation status of **FAM150A** was demonstrated to be a diagnostic and prognostic indicator for clear cell renal cell carcinoma (ccRCC) using high performance liquid chromatography method.[33] However, no studies found that **FAM150A** associated with CRC.

6) **KRT7** (Keratin 7) is a member of the keratin gene family. Previous studies showed that **KRT7** play a significant role in tumor metastasis and considered as prognostic biomarker and potential targets for
therapeutic prevention of metastasis. In addition, KRT7 was down-regulated and hypermethylated in CRC tissues compared with adjacent normal tissues and may lead to the occurrence of CRC.

7) STK33 (Serine/threonine kinase 33) was experimentally demonstrated that downregulation and hypermethylation of STK33 in CRC tissues compared with normal tissues with $P < 0.001$, and STK33 methylation was also remarkably associated with lymph node metastasis, tumor invasion, distant metastases and tumor stage. The precious study and my research work both suggested that the hypermethylation and downregulation of STK33 may be a promising biomarker for the diagnosis and prognosis of CRC.

To our best knowledge, there is still a lack of research about the relationship between the left three MDGs and oncogenesis, which may represent new predictive biomarkers.

**Conclusions**

In conclusion, we identified 10 MDGs that could be used as potential biomarkers for CRC, of which 9 performed well as independent diagnostic predictors and 2 could be used as prognostic indicators.

**Declarations**

**Funding information:** This work was supported by a grant from the Key Science and Technology Program of Shaanxi Province (2019ZDLSF02-05) and National Natural Scientific Foundation of China (81974378 and 82003115).

**Availability of data and material:** The data that support the findings of this study are openly available in TCGA and GTEx from the UCSC Xena (https://xenabrowser.net/datapages/) and GEO database (https://www.ncbi.nlm.nih.gov/geo/).

**Author Contributions:** Lichao Cao were involved in the study concept and design, and drafting the manuscript. Erfei Chen put forward some kind suggestions. Lichao Cao and Jin Yang helped analyze and interpret the data.

**Conflict of Interest:** The authors declare that they have no conflict of interest.

**References**

1. Ashikari D, Takayama K, Obinata D et al (2017) CLDN8, an androgen-regulated gene, promotes prostate cancer cell proliferation and migration. Cancer Sci 108:1386–1393
2. Bray F, Ferlay J, Soerjomataram I et al (2018) Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. Cancer J Clin 68:394–424
3. Cedoz PL, Prunello M, Brennan K et al (2018) MethylMix 2.0: an R package for identifying DNA methylation genes. Bioinformatics 34:3044–3046
4. Dai QX, Liao YH, Deng XH et al (2020) A novel epigenetic signature to predict recurrence-free survival in patients with colon cancer. Clin Chim Acta 508:54–60

5. Dexheimer GM, Alves J, Reckziegel L et al (2017) DNA Methylation Events as Markers for Diagnosis and Management of Acute Myeloid Leukemia and Myelodysplastic Syndrome. Disease markers 2017:5472893

6. Egger G, Liang G, Aparicio A et al (2004) Epigenetics in human disease and prospects for epigenetic therapy. Nature 429:457–463

7. Feodorova Y, Tashkova D, Koev I et al (2018) Novel insights into transcriptional dysregulation in colorectal cancer. Neoplasma 65:415–424

8. Flebbe H, Hamdan FH, Kari V et al (2019) Epigenome Mapping Identifies Tumor-Specific Gene Expression in Primary Rectal Cancer. Cancers (Basel) 11

9. Gevaert O (2015) MethylMix: an R package for identifying DNA methylation-driven genes. Bioinformatics 31:1839–1841

10. Giordano G, Parcesepe P, D’andrea MR et al (2019) JAK/Stat5-mediated subtype-specific lymphocyte antigen 6 complex, locus G6D (LY6G6D) expression drives mismatch repair proficient colorectal cancer. Journal of experimental clinical cancer research: CR 38:28

11. Guan J, Umapathy G, Yamazaki Y et al (2015) FAM150A and FAM150B are activating ligands for anaplastic lymphoma kinase. eLife 4:e09811

12. Hallberg B, Palmer RH (2013) Mechanistic insight into ALK receptor tyrosine kinase in human cancer biology. Nat Rev Cancer 13:685–700

13. Herman JG, Baylin SB (2003) Gene silencing in cancer in association with promoter hypermethylation. N Engl J Med 349:2042–2054

14. Jaenisch R, Bird A (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat Genet 33 Suppl:245–254

15. Janostiak R, Malvi P, Wajapeyee N (2019) Anaplastic Lymphoma Kinase Confers Resistance to BRAF Kinase Inhibitors in Melanoma. iScience 16:453–467

16. Jiang L, Tolani B, Yeh CC et al (2019) Differential gene expression identifies KRT7 and MUC1 as potential metastasis-specific targets in sarcoma. Cancer management research 11:8209–8218

17. Kerachian MA, Javadmanesh A, Azghandi M et al (2020) Crosstalk between DNA methylation and gene expression in colorectal cancer, a potential plasma biomarker for tracing this tumor. Scientific reports 10

18. Lesche R, Payne S, Model F et al (2007) DNA methylation markers for diagnosis and prognosis of prostate cancer. Tumor Biol 28:51–51

19. Li M, Zhao LM, Li SL et al (2018) Differentially expressed IncRNAs and mRNAs identified by NGS analysis in colorectal cancer patients. Cancer medicine 7:4650–4664

20. Liu GJ, Wang YJ, Yue M et al (2020) High expression of TCN1 is a negative prognostic biomarker and can predict neoadjuvant chemosensitivity of colon cancer. Scientific reports 10
21. Martinez C, Rodino-Janeiro BK, Lobo B et al (2017) miR-16 and miR-125b are involved in barrier function dysregulation through the modulation of claudin-2 and cingulin expression in the jejunum in IBS with diarrhoea. Gut 66:1537–1538

22. Ritchie ME, Phipson B, Wu D et al (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic acids research 43:e47

23. Siegel RL, Miller KD, Goding Sauer A et al (2020) Colorectal cancer statistics, 2020. Cancer J Clin 70:145–164

24. Tian Y, Morris TJ, Webster AP et al (2017) ChAMP: updated methylation analysis pipeline for Illumina BeadChips. Bioinformatics 33:3982–3984

25. Vaissiere T, Sawan C, Herceg Z (2008) Epigenetic interplay between histone modifications and DNA methylation in gene silencing. Mutat Res 659:40–48

26. Wang JY, Wang CL, Wang XM et al (2017) Comprehensive analysis of microRNA/mRNA signature in colon adenocarcinoma. Eur Rev Med Pharmacol Sci 21:2114–2129

27. Wang T, Birsoy K, Hughes NW et al (2015) Identification and characterization of essential genes in the human genome. Science 350:1096–1101

28. Wang X, Zhang D, Zhang C et al (2020) Identification of epigenetic methylation-driven signature and risk loci associated with survival for colon cancer. Annals of translational medicine 8:324

29. Wu J, Gao F, Xu T et al (2020) CLDN1 induces autophagy to promote proliferation and metastasis of esophageal squamous carcinoma through AMPK/STAT1/ULK1 signaling. Journal of cellular physiology 235:2245–2259

30. Xu RH, Wei W, Krawczyk M et al (2017) Circulating tumour DNA methylation markers for diagnosis and prognosis of hepatocellular carcinoma. Nat Mater 16:1155–1155+

31. Yin MD, Ma SP, Liu F et al (2018) Role of serine/threonine kinase 33 methylation in colorectal cancer and its clinical significance. Oncology letters 15:2153–2160

32. Ying JC, Xu T, Wang Q et al (2018) Exploration of DNA methylation markers for diagnosis and prognosis of patients with endometrial cancer. Epigenetics-Us 13:490–504

33. Yotani T, Yamada Y, Arai E et al (2018) Novel method for DNA methylation analysis using high-performance liquid chromatography and its clinical application. Cancer Sci 109:1690–1700