The Methylation of SDC2 and TFPI2 Defined Three Methylator Phenotypes of Colorectal Cancer

Ruixue Lei
The Fourth Affiliated Hospital of Henan University of Science and Technology, Anyang Tumor Hospital
Yanteng Zhao
The First Affiliated Hospital of Zhengzhou University
Kai Huang
The First Affiliated Hospital of Anhui Medical University
Qian Wang
Zhongnan Hospital of Wuhan University
Kangkang Wan
Wuhan Ammunition Life-tech Company, Ltd
Tingting Li
Wuhan Ammunition Life-tech Company, Ltd
Haijun Yang (yhj1972@126.com)
The Fourth Affiliated Hospital of Henan University of Science and Technology, Anyang Tumor Hospital
Xianping Lv
The First Affiliated Hospital of Zhengzhou University

Research Article

Keywords: colorectal cancer, methylation, phenotypes, SDC2, TFPI2

Posted Date: November 29th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1022929/v1

Version of Record: A version of this preprint was published at BMC Gastroenterology on February 28th, 2022. See the published version at https://doi.org/10.1186/s12876-022-02175-3.

Abstract

Backgroud

Methylated *SDC2* and *TFPI2* are applied frequently for the early detection of colorectal cancer (CRC). However, they often miss some positive samples, which directly affects their sensitivities, and the underlining mechanism is not well known.

Methods:

CRC samples from TCGA and GEO datasets were divided into three groups, Highmethylation/ High-methylation (HH), High-methylation/Low-methylation (HL), and Lowmethylation/Low-methylation (LL) according to the methylation status of SDC2 and TFPI2 promoters. Variations in age, tumor location, and microsatellite instable were then assessed between the three groups and verified in our custom cohort.

Results

Samples of HL group preferred to derive from left-sided CRCs ($P < 0.05$). HH samples showed the highest microsatellite instability and mutation load (mean nonsynonymous mutations for HH/HL/LL: 10.55/3.91/7.02, $P = 0.0055$). Almost all mutations of *BRAF*, one of the five typical CpG island methylator phenotype (CIMP) related genes, were observed in HH group (HH/HL/LL: 51/0/1, $P = 0.018$). Besides, older patients were frequently found in HH group. Expression analysis identified 37, 84, and 22 group-specific differentially expressed genes (DEGs) for HH, HL, and LL, respectively. Functional enrichment analysis revealed that HH-specific DEGs were mainly related to transcription regulation, while LL-specific DEGs were enriched in the biological processes of extracellular matrix interaction and cell migration.

Conclusions:

The three methylation phenotypes identified based on *SDC2* and *TFPI2* methylation status showed extensive variations in tumor location, patient age, MSI and ECM biology processes, suggesting that these respective sides should be considered when developing new methylation-based biomarkers for CRC detection.

Background

Colorectal cancer (CRC) is responsible for over 1 million new cases every year and around 700,000 deaths occurred worldwide, making it the third most frequently diagnosed cancer [1, 2]. In China, the incidence and mortality of CRC have been witnessed with an increasing trend of 12.8 in 2003 to 16.8 per 100,000 in 2011 and 5.9 in 2003 to 7.8 per 100,000 in 2011, respectively [3]. It is believed that CRC represents a heterogeneous group of tumors characterized by complex multifactorial phenotypes and multiple risk factors responses for the development of CRC [4, 5]. Many factors including diet, tobacco smoking, microbial, overweight and obesity, genetic factors, as well as metabolic and other exposures can alter the

risk of getting CRC [6-9]. Nearly half of CRCs were attributed to unhealthy diets such as low vegetable and fruit intake, high red and processed meat intake, and alcohol drinking, *etc.,* in China in 2012 [8].

Syndecan-2 (*SDC2*), as one of the syndecan family of heparan sulfate proteoglycan, has been demonstrated to play an important role in cancer progression through regulation of cell adhesion, proliferation, and migration in many studies [10-13]. Tissue factor pathway inhibitor-2 (*TFPI-2*), belongs to the Kunitz-type serine proteinase inhibitor family and is thought to be functional in the regulation of extracellular matrix digestion and re-modeling by inhibiting a broad spectrum of serine proteinases [14, 15]. Unlike the tumorigenic behaviors of *SDC2* in colon cancer cells, *TFPI2* has been shown as a tumor suppressor gene in several malignant tumors [16-19]. However, promoters of both were found with frequently hypermethylated status in colon cancer cells compared to normal tissue cells in several epi-genomics studies [20, 21]. The frequently aberrant DNA methylation of *SDC2* and *TFPI2* makes them promise biomarkers for the early detection of CRC [20, 22-25]

Several studies have suggested a better performance of combined multi-targets for CRC early detection than a single biomarker [26-28]. However, during clinical practice, some CRC samples were detected only by a single or no target, reflecting the preference of different targets in distinguishing CRCs from normal samples, which directly affects the target utility. Further investigations for these undetected samples will help us to have a comprehensive understanding of the performance of methylation-based markers for CRC early detection. In this study, we first classified CRC samples into three methylator groups, *SDC2*/ *TFPI2* double-positive group (HH, high-methylation/high-methylation), *SDC2*/ *TFPI2* single positive group (HL, high-methylation/low-methylation) and *SDC2*/ *TFPI2* double negative group (LL, low-methylation/low-methylation) according to the promoter methylation status of *SDC2* and *TFPI2*. The clinical-pathological parameters and molecular features were then evaluated by inner and outer cohorts including TCGA, GEO, and our D311 CRC dataset.

Methods

Data preparation

The level 3 methylation data, raw read-count of RNA-seq, and clinical information of colon and rectum adenocarcinoma patients were retrieved from The Cancer Genome Atlas (TCGA) data portal (https://portal.gdc.cancer.gov/) by using the TCGAbiolinks R package [29]. The platform of methylation data from TCGA is Illumina Human Methylation 450 Beadchip (450K array) and we also searched the Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) for eligible datasets that are generated by 450K array. Two GEO datasets, GSE48684 [30] and GSE79740 [31], were then downloaded because of their available clinical information. Empirical Bayes (EB) batch adjustment along with a two-step quantile normalization method [32] was conducted for batch effect removal before GSE48684 and GSE79740 datasets were merged as one set. Missing values of the 450k array were inferred and fulfilled by the Bayesian Network structure learning algorithm [33]. All samples without clinical information were removed. The information of preprocessed data used in this study was presented in **Table 1**.

Patient samples collection

Fresh-frozen colorectal cancer tissues (n=257) and colorectal mucosa (n=54) tissues were collected at Zhongnan Hospital of Wuhan University at the time of surgery. All objects recruited signed a written informed consent and their final diagnosis was determined based on colonoscopy or histological test. Participants who undertook any chemotherapy or radiotherapy, or had incomplete information were excluded. The collected information consists of age, gender, tumor size, tumor location, grade, and MSI status. Detailed demographic and clinical features of the subjects were listed in **Table 1**. We classified rectosigmoid, descending colon, and splenic flexure tumors as left-sided cancer, whereas hepatic flexure and ascending colon tumors were grouped as right-sided cancer [34]. This study was approved by the medical ethics committee of Zhongnan Hospital of Wuhan University (No.2019099).

Table 1 Clinical characteristics of subjects.

Subject characteristics	TCGA CRC	GSE48684/GSE79740	D311 CRC
Normal	45	51	54
Tumor	394	108	257
Age,no.(%)			
>=60	254(64.96%)	Not available	149(58.75%)
<60	137(35.04%)	Not available	106(41.25%)
Localization, no. (%)			
Left Colon	148(37.56%)	72(66.67%)	82(31.91%)
Right Colon	147(37.31%)	28(25.93%)	49(19.07%)
Rectum	46(11.68%)	6(5.56%)	90(35.02%)
Others	53(13.45%)	2(1.85%)	36(14.01%)
Gender,no.(%)			
Male	211(53.55%)	Not available	152(59.14%)
Female	183(46.45%)	Not available	105(40.86%)
MSI-H,no.(%)	65(16.75%)	Not available	15(8.06%)
MSS,no.(%)	323(83.25%)	Not available	171(91.94%)

Identification of target regions

For *SDC2* and *TFPI2*, there are 32 and 23 probes located in their 5'UTRs and gene bodies in Illumina Human Methylation 450k array, where 14 and 21 probes were within the 2 kilo-base upstream of

transcript start site (TSS), respectively (**Supplementary table 1**). To determine the appropriate target regions, we focused on the regions that showed the largest difference between normal and tumor samples. First, the normal and tumor methylation values for each probe (β_{Tumor} and β_{Normal}) were calculated using average β values of the normal and tumor samples in each dataset. The probes with Δβ (β_{Tumor}-β_{Normal}) >=0.3 were selected as the most varied probes. Then, the most varied regions were identified according to the probe genomic coordinates. Three probes of *SDC2* satisfied the criterial, cg16935295, cg04261408 and cg10292139, were found located in a contiguous genomic positions. Since adjacent CpG sites usually exhibited similar methylation status, also known as methylation block, the region that consisted of four probes, cg16935295, cg04261408, cg14625631 and cg10292139, was finally selected as the target. The mean β values of these four probes were calculated as the methylation levels of *SDC2*. Similarly, we determined the region where seven probes, cg12973591, cg22799321, cg24531255, cg17338208, cg26739865, cg22441533, and cg14377593, located in as the target region of *TFPI2*, and their average β values were also calculated as the methylation levels of *TFPI2*.

Identification of appropriate β threshold

In this study, the appropriate β threshold to define high-methylation or low-methylation was determined according to previous studies [35, 36]. Briefly, the distribution parameters of the methylation M-values were estimated by Gaussian mixed linear model to determine the optimal M-value as well as the corresponding β-value. β-value and M-value are two metrics that are often used for Infinium methylation array. β-value is more easily interpreted as an approximation of the overall percentage of methylation at a given CpG site in a sample, while M-value is more statistically valid for the differential analysis of methylation levels. Although M-value is difficult to infer the methylation levels directly, it provides an insight into the distribution of methylation across the genome. Therefore, we investigated the distribution patterns of β-value and M-value of 450K array on TCGA CRC dataset.

Methylation-specific PCR

The genomic DNA was extracted by UnigeneDx FFEE DNA extraction kit according to the manufacturer's instructions. Tissue-derived genomic DNA was chemically modified by sodium bisulfite to convert unmethylated cytosine to uracil while leaving methylated cytosine unchanged. Methylation-specific PCR (MSP) was used to determine the methylation status of *SDC2* and *TFPI2* in normal and tumor tissue DNA, β-actin [37] was used as an internal control. Specific primers and probes for the target region of *SDC2* and *TFPI2* was designed as showing in **Supplementary table 2**. We use the cycle threshold (Ct) value to determine the methylation status of these two genes and the values for tissue samples were considered "invalid" if the ACTB Ct was greater than 36.00 and methylated *SDC2*/ *TFPI2* were considered "detected" if the Ct values were less than 45.00. For samples with no amplification curve of the MSP occurred after 45 cycles, the Ct value was assigned 45.00. Three MSP replicates were conducted for each sample and the average Ct value was used for further analysis.

Identification of differentially expressed genes

Level 3 RNA-seq data of TCGA CRCs were preprocessed before expression analysis by removing low expressed genes whose expressions were zero among more than 90% samples. DESeq2 [38] (V1.30.0) package was used to perform a pairwise comparison between all the three methylator groups for the identification of differentially expressed genes (DEGs). Adjusted P values were calculated by the false discovery rate (FDR) method. Genes with adjusted P-value < 0.05 were selected as DEGs and used for further functional enrichment analysis. We used GeneCodis [39] for the enrichment analysis of GO biology process (BP) and KEGG pathway for identified DEGs

Statistical analysis

Comparisons for two paired or unpaired samples were performed for continuous variables using paired or unpaired student t-test. Kruskal-Wallis nonparametric analysis was used for multi-group comparisons of continuous variables. For categorical variables, fisher's exact test was applied to determine if there are nonrandom associations between SDC2/ *TFPI2* methylator groups and clinical characteristics, such as age, sex, and tumor location. In this study, we chose 0.2 as the optimal β threshold. If the β values of *SDC2* and *TFPI2* promoters were > 0.2, they were defined as high-methylation, and vice versa as low-methylation. For Ct values, 38 was set as the threshold according to the previous study [40]. High-methylated *SDC2* or *TFPI2* were determined when their Ct values were < 38, otherwise, they were determined low-methylated. All statistical analyses were performed using R software (version 3.6.0) and the source code was deposited to GitHub (https://github.com/amsinfor/methylator_group/blob/master/soure_code_v1.R).

Results

Methylation status of *SDC2* and *TFPI2* in CRC

We identified 4 and 7 probes in the promoters of *SDC2* and *TFPI2,* respectively (**Supplementary table 1**). The average β values of these filtered probes were used as the methylation level of *SDC2* and *TFPI2* (herein termed as SDC2_P and TFPI2_P). Correlation analysis indicated that the methylation levels of both genes showed a weak negative correlation with the expression levels (**Supplementary Figure 1**). A lower Δβ value was found for *TFPI2* than *SDC2* which might attribute to its higher background methylation level on normal controls (**Figure 1A&B**). The β values were severely compressed around 0.05 and 0.95, and no obvious bimodal was observed (**Figure 1C**). In contrast, the histogram of M values clearly showed a bimodal distribution (**Figure 1D**). We attempted to estimate the distribution parameters of M values by fitting a Gaussian mixed linear model, and the results indicated that when M equaled to -1.95 which corresponded to β value of 0.205, the optimal boundary was determined between the two peaks (methylated peak and unmethylated peak). These findings implied that 0.2 may be an appropriate β threshold, which is consistent with the studies of Pan Du et al. and Sarah Dedeurwaerder et al (see

methods). We observed that high-methylated *SDC2* and *TFPI2* occurred in more than 85% of CRC samples in three datasets (average 88.21%). Approximately 10% of CRCs had single high-methylated gene (*SDC2* or *TFPI2*) (average 9.24%), and less than 3% of CRCs harbored low-methylated *SDC2* and *TFPI2* (2.56% on average) (**Table 2**).

Table 2. Methylation status of *SDC2* and *TFPI2* in three datasets.

Methylation status	TCGA CRC	GSE48684/GSE79740	D311
High-M SDC2 and High-M TFPI2	353 (89.59%)	97 (89.81%)	219 (85.21%)
High-M SDC2 and Low-M TFPI2	9 (2.28%)	3 (2.78%)	6 (2.33%)
Low-M SDC2 and High-M TFPI2	25 (6.35%)	5 (4.63%)	24 (9.34%)
Low-M SDC2 and Low-M TFPI2	7 (1.78%)	3 (2.78%)	8 (3.11%)

The association of methylator groups with tumor location

For the purpose of early detection, HL and LL groups are very important because they can affect the sensitivity of a given biomarker directly. We first compared the three methylator groups with tumor location, and found that HL group CRCs were more frequently originated from left-sided colon (**supplementary table 3**). A small fraction of CRCs were from rectum, and very few were from right-sided colon (**Figure 2A-C**).

The association of methylator groups with genomic variations

Microsatellite instability and hypermutation have been regarded as important molecular characteristics of CRCs. Compared to other two groups, HH group CRCs presented the highest mutation load (**Figure 3A**, $P < 0.05$) in the TCGA CRC dataset. By using the MANTIS score [41], which is used to evaluate the MSI status, we grouped the TCGA CRCs into MSI-H and MSS if this score is > 0.4. The β values of *SDC2* and *TFPI2* showed a high concordance with MANTIS scores in the MSI-H group (**Figure 3B**, $P < 0.001$), which is consistent with the result that higher mutation load occurred in HH group. Meanwhile, the methylation levels of *MLH1* which has been demonstrated its methylation associated with microsatellite instability [42], also showed significant variations between the three groups. Results indicated that HH group exhibited the highest methylation levels, followed by HL group and the lowest on LL group (**Supplementary figure 2**). We studied the association of three methylator groups with the mutation of 5 typical CIMP-related genes including *BRAF*, *PIK3CA*, *KRAS*, *TP53*, and *APC*. Almost all *BRAF*-mutated CRCs were in HH group (**Figure 3C**, HH/HL/LL: 51/0/1, $P = 0.018$). We further compared the association between MSI status and tumor locations with TCGA CRCs and our D3111 CRCs. The MSI-H CRCs were more preferred in the right-sided colon (**Figure 3D&E**, $P < 0.001$), which possibly elucidated the potential mechanism that HL group CRCs were mainly in left-sided colon and MSI-H CRCs were less likely in HL

group. Additionally, a total of 111 mutated genes were significantly enriched between the three groups (**Supplementary table 4**), involved in 53 biological processes and 31 KEGG pathways. The top 5 enriched terms were mainly related to signal transduction, expression regulation and metabolic pathways (**Figure 4A&B**).

The association of methylator groups with patient age

Patient age is one of the risk factor for colon cancer and we found a significantly older age in HH group patients, while it was the youngest for LL group patients (**Figure 5A&B,** $P < 0.05$). Since the genomic DNA methylation is associated with patient age, we observed a positive correlation of the methylation levels of *SDC2* and *TFPI2* with patient age (**Figure 5C&D**). We then verified this result on the D311 dataset. Because of no direct correspondence between absolute Ct values and β values, we adopted a method similar to quantify gene expressions by quantitative PCR to determine the methylation levels of *SDC2* and *TFPI2*. The Ct values of internal reference gene, *ACTB*, was used as control to quantify the relative methylation levels of these two genes, also called ΔCt. The $2^{-\Delta Ct}$ was then calculated for correlation analysis. Similar results were observed on the D311 dataset, though no strong correlation was presented (**Supplementary figure 3**), indicating that young patients might be more likely to be miss-detected.

Identification of DEGs among the three methylator groups

We performed differential expression analysis by using the gene expression profile of TCGA CRCs to identify group-specific DEGs. A total of 37 HH-specific, 84 HL-specific and 22 LL-specific DEGs were identified according to their average expression values on the three groups (**Figure 6A**). Functional enrichment analysis indicated that HH specific DEGs were mainly related to the regulation of transcription and other processes (**Figure 6B**), while LL specific DEGs are enriched in the biological processes of extracellular matrix interaction (ECM) and cell migration (**Figure 6C**). These results might elucidate potential alterations in the biological processes of ECM and cell migration that are related to the different characteristics of these three groups.

Discussion

Quantifying aberrantly methylated genes was a feasible method for the early detection of CRCs. Several biomarkers have been demonstrated excellent performance in CRC early detection [25, 27, 28], however, they often suffer the limitation with some positive samples miss detected, which directly affects their efficiency for CRC detection. In this study, we defined three CRC methylator groups, HH, HL, and LL based on the methylation status of *SDC2* and *TFPI2* and then assessed their characteristics of genomic

instability, mutation load, patient age, and biological processes. These findings suggested that it would be reasonable and essential to define three methylator groups according to the methylation status of these targets and will benefit guiding the development of more effective methylated biomarkers.

Our results revealed that high-methylated *SDC2* or *TFPI2* occurred in more than 95% of CRCs, suggesting that their methylation status can perform well to discriminate CRCs from normal controls. Previous studies have demonstrated the good performance of methylated *SDC2* and *TFPI2* alone or in combination with other biomarkers for stool-based CRC detection [23, 43], however, the combination of these two targets has not been reported. The present study revealed to some extent that the dual-target showed huge potential for CRC detection, which will be helpful for the development of stool-based noninvasive or blood-based minimally invasive detection techniques in the future.

However, in terms of individual genes, nearly 10% of the samples had only one gene methylated, and here we defined them as the HL group. Further analysis showed that more samples harboring high-methylated *TFPI2* and low-methylated *SDC2* were from HL group (about 3 times than the samples with high-methylated *SDC2* and low-methylated *TFPI2*). These results suggested that methylated *TFPI2* occurs more frequently on CRCs, implying that combining *TFPI2* with *SDC2* would help improve the sensitivity of CRC detection. Previous studies have demonstrated that multi-target outperformed single target [44, 45], which was evidenced by this study.

In clinical practice, the HL and LL group CRCs will cut down the sensitivity of these biomarkers. Our findings indicated that HL group CRCs more likely originate from the left-sided colon. Many differences were observed between the proximal (right-sided) and distal (left-sided) CRCs. For example, right-sided cancers were reported an increased incidence of proximal migration, while it was inversed for rectosigmoid tumors [46]. Moreover, the incidence between proximal and distal CRCs also differs in age and gender [47]. These data reflect extensive distinctions in molecular pathogenesis between the two anatomical locations, which might generate a significant impact on tumorigenesis in these respective sides. Several studies have demonstrated that left-sided colon presents lower degrees of methylation than the right-sided colon, which was called the CpG island methylator phenotype, or CIMP, characterized by significant hyper-methylated CpG islands of tumor suppressors [48][49]. Therefore, tumor location would be an important factor of biological heterogeneity. In this study, we observed lower methylation levels of *SDC2* in left-sided CRCs. Given that HL group CRCs were dominated by high-methylated *TFPI2* and low-methylated *SDC2*, this might explain why they appeared more frequently in the left-sided colon. These results suggested a potential impact of tumor locations on the early detection of CRCs when adopting abnormal methylated DNA as biomarkers.

Additionally, we found a positive correlation between the methylation levels of *SDC2* and *TFPI2* and MSI scores in MSI-H CRCs, as well as lower mutation load and rare *BRFA* mutations in HL group CRCs. Approximately 10% of CRCs harbor mutated *BRAF*, and its mutations are associated not only with poor prognosis but also with less benefit in metastatic CRCs when treated by antibodies [50]. Therefore, the methylator phenotypes may serve as a stratification factor in clinical therapies. It's been reported

that CIMP tumors showed significant associations with *BRAF* mutations, MSI-H [49]. These results, on the other hand, confirmed that molecular events, such as epigenetic variations, instability, aberrant DNA mutations, and MSI, are coupled with each other.

Gene expression analysis identified methylator group-specific DEGs and functional annotation of LL-specific DEGs was suggested to focus on the biological process of ECM-receptor interaction, implying the potential alteration in molecular pathways in LL group CRCs. Interestingly, many studies have showed very important roles of *SDC2* and *TFPI2* in the interaction of extracellular matrix with cell plasma [15, 51]. Besides, we also observed a significant enrichment of 111 mutant genes on GO term of extracellular matrix organization, which might, from the other hand, imply the tight association between ECM-receptor interaction and the three *methylator groups*. Overall, these findings revealed the possible impact of ECM process on the performance of methylated *SDC2* and *TFPI2* in detecting CRCs.

Colorectal cancer is a disease with high heterogeneity, often divided into proximal (right sided) and distal (left-sided) cancer according to their anatomical locations. This classification is reasonable because of their distinctive embryonic derivation, which is the midgut and the hindgut for the proximal and distal colon, respectively [46, 52]. It might give rise to the altered methylations between the left- and right-sided colons, thus affecting the early detection of CRCs based on abnormally methylated DNA.

In conclusion, the current study demonstrated the possible association of CIMP phenotype, tumor location, and MSI with the dual-target in CRC early diagnosis, making us propose a possible diagram of patient characteristics between the three groups **(Figure 7)**. In terms of genomic features, HH group CRCs are characterized by more often microsatellite instable (MSI), high mutation load, and frequent BRAF-mutated. HL group CRCs prefer to originate from the left-sided colon. Our observations also suggested that it should be considered when developing new methylation-based biomarkers for CRC detection in these respective sides.

Declarations

Ethics approval and consent to participate

Approval for this study was obtained from the Ethics Committee of Zhongnan Hospital of Wuhan University (No.2019099).All experiments were performed in accordance with relevant guidelines and regulations. Written informed consent was obtained from individual or guardian participants.

Consent for publication

Not applicable

Availability of data and materials

The TCGA CRC 450k data and GEO datasets are publicly available online. The D311 data used and analyzed in study are available from the corresponding author on reasonable request.

Competing interest

Wuhan Ammunition Life-tech Company, Ltd. has applied for the patent relating to the dual targets of *SDC2* and *TFPI2*. The other authors declare no conflict of interest.

Funding

This study was supported by Medical Science and Technology Research Plan Joint Construction Project of Henan Province (Granted No. 2018020121).

Acknowledgments

We declare that submitted manuscript and other materials are not under consideration for publication elsewhere. All authors listed have read the complete manuscript and have approved submission of the paper.

Author Contributions

Study design: HJ Y.

Data collection and analysis: RX L, YT Z, KK W, TT L and Q W.

Manuscript writing: RX L, YT Z, K H and KK W.

Manuscript and results revising: HJ Y and XP L.

References

1. Navarro M, Nicolas A, Ferrandez A, Lanas A: **Colorectal cancer population screening programs worldwide in 2016: An update**. *WORLD J GASTROENTERO* 2017, **23**(20):3632-3642.
2. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F: **Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012**. *INT J CANCER* 2015, **136**(5):E359-E386.
3. Zhu J, Tan Z, Hollis-Hansen K, Zhang Y, Yu C, Li Y: **Epidemiological Trends in Colorectal Cancer in China: An Ecological Study**. *DIGEST DIS SCI* 2017, **62**(1):235-243.

4. Binefa G, Rodríguez-Moranta F, Teule A, Medina-Hayas M: **Colorectal cancer: from prevention to personalized medicine**. *World J Gastroenterol* 2014, **20**(22):6786-6808.
5. Ang PW, Loh M, Liem N, Lim PL, Grieu F, Vaithilingam A, Platell C, Yong WP, Iacopetta B, Soong R: **Comprehensive profiling of DNA methylation in colorectal cancer reveals subgroups with distinct clinicopathological and molecular features**. *BMC CANCER* 2010, **10**:227.
6. Azeem S, Gillani SW, Siddiqui A, Jandrajupalli SB, Poh V, Syed SS: **Diet and Colorectal Cancer Risk in Asia: a Systematic Review**. *Asian Pac J Cancer Prev* 2015, **16**(13):5389-5396.
7. Gao R, Gao Z, Huang L, Qin H: **Gut microbiota and colorectal cancer**. *EUR J CLIN MICROBIOL* 2017, **36**(5):757-769.
8. Gu M, Huang Q, Bao C, Li Y, Li X, Ye D, Ye Z, Chen K, Wang J: **Attributable causes of colorectal cancer in China**. *BMC CANCER* 2018, **18**(1):38.
9. Boland CR, Goel A: **Microsatellite instability in colorectal cancer**. *GASTROENTEROLOGY* 2010, **138**(6):2073-2087.
10. Choi S, Choi Y, Jun E, Kim IS, Kim SE, Jung SA, Oh ES: **Shed syndecan-2 enhances tumorigenic activities of colon cancer cells**. *Oncotarget* 2015, **6**(6):3874-3886.
11. Park H, Kim Y, Lim Y, Han I, Oh ES: **Syndecan-2 mediates adhesion and proliferation of colon carcinoma cells**. *J BIOL CHEM* 2002, **277**(33):29730-29736.
12. Vicente CM, Ricci R, Nader HB, Toma L: **Syndecan-2 is upregulated in colorectal cancer cells through interactions with extracellular matrix produced by stromal fibroblasts**. *BMC CELL BIOL* 2013, **14**:25.
13. Hua R, Yu J, Yan X, Ni Q, Zhi X, Li X, Jiang B, Zhu J: **Syndecan-2 in colorectal cancer plays oncogenic role via epithelial-mesenchymal transition and MAPK pathway**. *BIOMED PHARMACOTHER* 2020, **121**:109630.
14. Sierko E, Wojtukiewicz MZ, Kisiel W: **The role of tissue factor pathway inhibitor-2 in cancer biology**. *SEMIN THROMB HEMOST* 2007, **33**(7):653-659.
15. Chand HS, Foster DC, Kisiel W: **Structure, function and biology of tissue factor pathway inhibitor-2**. *Thromb Haemost* 2005, **94**(6):1122-1130.
16. Wang G, Huang W, Li W, Chen S, Chen W, Zhou Y, Peng P, Gu W: **TFPI-2 suppresses breast cancer cell proliferation and invasion through regulation of ERK signaling and interaction with actinin-4 and myosin-9**. *Sci Rep* 2018, **8**(1):14402.
17. Lavergne M, Jourdan ML, Blechet C, Guyetant S, Pape AL, Heuze-Vourc'H N, Courty Y, Lerondel S, Sobilo J, Iochmann S *et al*: **Beneficial role of overexpression of TFPI-2 on tumour progression in human small cell lung cancer**. *FEBS OPEN BIO* 2013, **3**:291-301.
18. Vaitkiene P, Skiriute D, Skauminas K, Tamasauskas A: **Associations between TFPI-2 methylation and poor prognosis in glioblastomas**. *Medicina (Kaunas)* 2012, **48**(7):345-349.
19. Xu Y, Qin X, Zhou J, Tu Z, Bi X, Li W, Fan X, Zhang Y: **Tissue factor pathway inhibitor-2 inhibits the growth and invasion of hepatocellular carcinoma cells and is inactivated in human hepatocellular carcinoma**. *ONCOL LETT* 2011, **2**(5):779-783.

20. Oh T, Kim N, Moon Y, Kim MS, Hoehn BD, Park CH, Kim TS, Kim NK, Chung HC, An S: **Genome-wide identification and validation of a novel methylation biomarker, SDC2, for blood-based detection of colorectal cancer**. *J MOL DIAGN* 2013, **15**(4):498-507.
21. Hibi K, Goto T, Kitamura YH, Yokomizo K, Sakuraba K, Shirahata A, Mizukami H, Saito M, Ishibashi K, Kigawa G *et al*: **Methylation of TFPI2 gene is frequently detected in advanced well-differentiated colorectal cancer**. *ANTICANCER RES* 2010, **30**(4):1205-1207.
22. Glöckner SC, Dhir M, Yi JM, McGarvey KE, Van Neste L, Louwagie J, Chan TA, Kleeberger W, de Bruïne AP, Smits KM *et al*: **Methylation of TFPI2 in stool DNA: a potential novel biomarker for the detection of colorectal cancer**. *CANCER RES* 2009, **69**(11):4691-4699.
23. Bagheri H, Mosallaei M, Bagherpour B, Khosravi S, Salehi AR, Salehi R: **TFPI2 and NDRG4 gene promoter methylation analysis in peripheral blood mononuclear cells are novel epigenetic noninvasive biomarkers for colorectal cancer diagnosis**. *J GENE MED* 2020, **22**(8):e3189.
24. Han YD, Oh TJ, Chung T, Jang HW, Kim YN, An S, Kim NK: **Early detection of colorectal cancer based on presence of methylated syndecan-2 (SDC2) in stool DNA**. *CLIN EPIGENETICS* 2019, **11**(1):51.
25. Hibi K, Goto T, Shirahata A, Saito M, Kigawa G, Nemoto H, Sanada Y: **Detection of TFPI2 methylation in the serum of colorectal cancer patients**. *CANCER LETT* 2011, **311**(1):96-100.
26. Chen J, Sun H, Tang W, Zhou L, Xie X, Qu Z, Chen M, Wang S, Yang T, Dai Y *et al*: **DNA methylation biomarkers in stool for early screening of colorectal cancer**. *J CANCER* 2019, **10**(21):5264-5271.
27. Sun M, Liu J, Hu H, Guo P, Shan Z, Yang H, Wang J, Xiao W, Zhou X: **A novel panel of stool-based DNA biomarkers for early screening of colorectal neoplasms in a Chinese population**. *J Cancer Res Clin Oncol* 2019, **145**(10):2423-2432.
28. Zhao G, Li H, Yang Z, Wang Z, Xu M, Xiong S, Li S, Wu X, Liu X, Wang Z *et al*: **Multiplex methylated DNA testing in plasma with high sensitivity and specificity for colorectal cancer screening**. *Cancer Med* 2019, **8**(12):5619-5628.
29. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, Sabedot TS, Malta TM, Pagnotta SM, Castiglioni I *et al*: **TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data**. *NUCLEIC ACIDS RES* 2016, **44**(8):e71.
30. Luo Y, Wong CJ, Kaz AM, Dzieciatkowski S, Carter KT, Morris SM, Wang J, Willis JE, Makar KW, Ulrich CM *et al*: **Differences in DNA methylation signatures reveal multiple pathways of progression from adenoma to colorectal cancer**. *GASTROENTEROLOGY* 2014, **147**(2):418-429.
31. Alvi MA, Loughrey MB, Dunne P, McQuaid S, Turkington R, Fuchs MA, McGready C, Bingham V, Pang B, Moore W *et al*: **Molecular profiling of signet ring cell colorectal cancer provides a strong rationale for genomic targeted and immune checkpoint inhibitor therapies**. *Br J Cancer* 2017, **117**(2):203-209.
32. Sun Z, Chai HS, Wu Y, White WM, Donkena KV, Klein CJ, Garovic VD, Therneau TM, Kocher JP: **Batch effect correction for genome-wide methylation data with Illumina Infinium platform**. *BMC MED GENOMICS* 2011, **4**:84.
33. Franzin A, Sambo F, Di Camillo B: **bnstruct: an R package for Bayesian Network structure learning in the presence of missing data**. *BIOINFORMATICS* 2017, **33**(8):1250-1252.

34. Gervaz P, Usel M, Rapiti E, Chappuis P, Neyroud-Kaspar I, Bouchardy C: **Right colon cancer: Left behind**. *Eur J Surg Oncol* 2016, **42**(9):1343-1349.

35. Du P, Zhang X, Huang CC, Jafari N, Kibbe WA, Hou L, Lin SM: **Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis**. *BMC BIOINFORMATICS* 2010, **11**:587.

36. Dedeurwaerder S, Defrance M, Calonne E, Denis H, Sotiriou C, Fuks F: **Evaluation of the Infinium Methylation 450K technology**. *EPIGENOMICS-UK* 2011, **3**(6):771-784.

37. Ooki A, Maleki Z, Tsay JJ, Goparaju C, Brait M, Turaga N, Nam HS, Rom WN, Pass HI, Sidransky D *et al*: **A Panel of Novel Detection and Prognostic Methylated DNA Markers in Primary Non-Small Cell Lung Cancer and Serum DNA**. *CLIN CANCER RES* 2017, **23**(22):7141-7152.

38. Love MI, Huber W, Anders S: **Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2**. *GENOME BIOL* 2014, **15**(12):550.

39. Tabas-Madrid D, Nogales-Cadenas R, Pascual-Montano A: **GeneCodis3: a non-redundant and modular enrichment analysis tool for functional genomics**. *NUCLEIC ACIDS RES* 2012, **40**(Web Server issue):W478-W483.

40. Li R, Qu B, Wan K, Lu C, Li T, Zhou F, Lin J: **Identification of two methylated fragments of an SDC2 CpG island using a sliding window technique for early detection of colorectal cancer**. *FEBS OPEN BIO* 2021.

41. Kautto EA, Bonneville R, Miya J, Yu L, Krook MA, Reeser JW, Roychowdhury S: **Performance evaluation for rapid detection of pan-cancer microsatellite instability with MANTIS**. *Oncotarget* 2017, **8**(5):7452-7463.

42. Toyota M, Ahuja N, Ohe-Toyota M, Herman JG, Baylin SB, Issa JJ: **CpG island methylator phenotype in colorectal cancer**. *Proceedings of the National Academy of Sciences* 1999, **96**(15):8681.

43. Oh TJ, Oh HI, Seo YY, Jeong D, Kim C, Kang HW, Han YD, Chung HC, Kim NK, An S: **Feasibility of quantifying SDC2 methylation in stool DNA for early detection of colorectal cancer**. *CLIN EPIGENETICS* 2017, **9**(1):126.

44. Imperiale TF, Ransohoff DF, Itzkowitz SH, Levin TR, Lavin P, Lidgard GP, Ahlquist DA, Berger BM: **Multitarget stool DNA testing for colorectal cancer screening**. *N Engl J Med* 2014, **370**(14):1287-1297.

45. Yang H, Xia BQ, Jiang B, Wang G, Yang YP, Chen H, Li BS, Xu AG, Huang YB, Wang XY: **Diagnostic value of stool DNA testing for multiple markers of colorectal cancer and advanced adenoma: a meta-analysis**. *CAN J GASTROENTEROL* 2013, **27**(8):467-475.

46. Missiaglia E, Jacobs B, D'Ario G, Di Narzo AF, Soneson C, Budinska E, Popovici V, Vecchione L, Gerster S, Yan P *et al*: **Distal and proximal colon cancers differ in terms of molecular, pathological, and clinical features**. *ANN ONCOL* 2014, **25**(10):1995-2001.

47. Bufill JA: **Colorectal cancer: evidence for distinct genetic categories based on proximal or distal tumor location**. *ANN INTERN MED* 1990, **113**(10):779-788.

48. Zong L, Abe M, Ji J, Zhu WG, Yu D: **Tracking the Correlation Between CpG Island Methylator Phenotype and Other Molecular Features and Clinicopathological Features in Human Colorectal Cancers: A Systematic Review and Meta-Analysis**. *Clin Transl Gastroenterol* 2016, **7**(3):e151.
49. Advani SM, Advani P, DeSantis SM, Brown D, VonVille HM, Lam M, Loree JM, Mehrvarz SA, Bressler J, Lopez DS *et al*: **Clinical, Pathological, and Molecular Characteristics of CpG Island Methylator Phenotype in Colorectal Cancer: A Systematic Review and Meta-analysis**. *TRANSL ONCOL* 2018, **11**(5):1188-1201.
50. Sanz-Garcia E, Argiles G, Elez E, Tabernero J: **BRAF mutant colorectal cancer: prognosis, treatment, and new perspectives**. *ANN ONCOL* 2017, **28**(11):2648-2657.
51. Xian X, Gopal S, Couchman JR: **Syndecans as receptors and organizers of the extracellular matrix**. *CELL TISSUE RES* 2010, **339**(1):31-46.
52. Yang J, Du XL, Li ST, Wang BY, Wu YY, Chen ZL, Lv M, Shen YW, Wang X, Dong DF *et al*: **Characteristics of Differently Located Colorectal Cancers Support Proximal and Distal Classification: A Population-Based Study of 57,847 Patients**. *PLOS ONE* 2016, **11**(12):e167540.

Figures

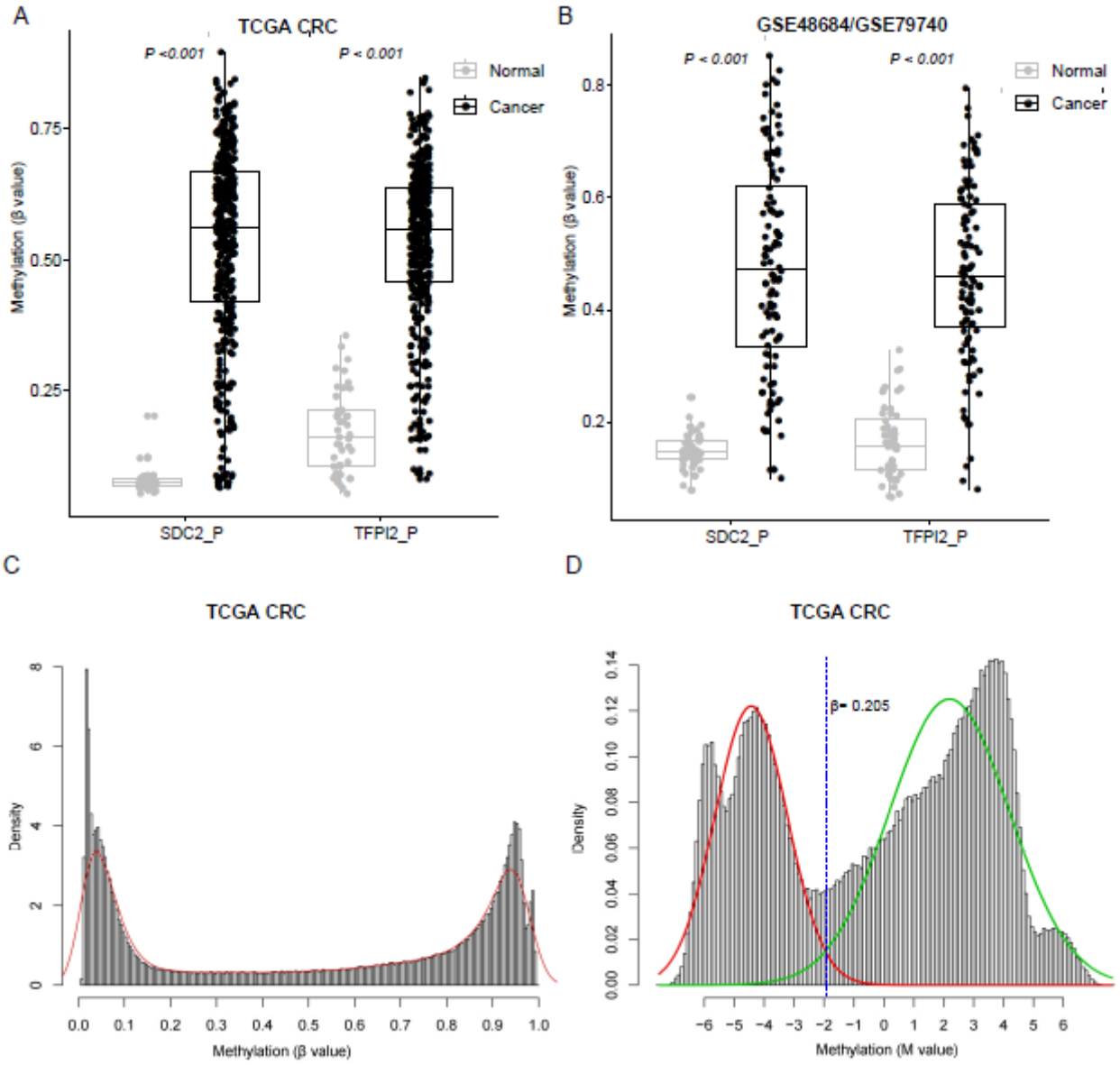

Figure 1

Methylation status of SDC2 and TFPI2 in CRC. A-B: β values of SDC2 and TFPI2 in CRCs and normal controls in TCGA dataset (A) and GEO dataset (B). C-D: The histograms and density distributions of methylation β-value (C) and M-value (D) on TCGA CRC. The red and green curves in F indicated the distribution of M values fitted by the Gaussian linear mixed model.

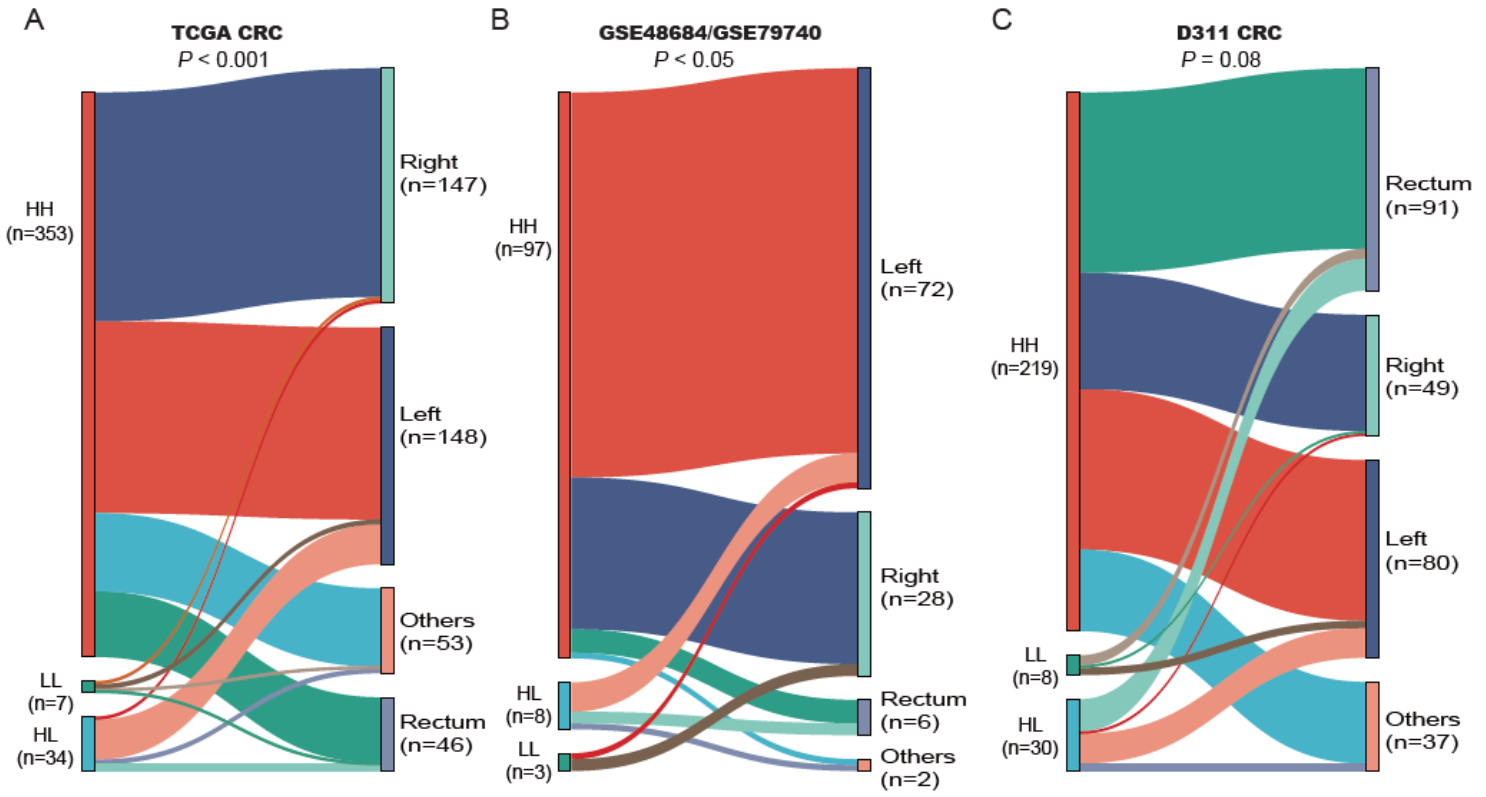

Figure 2

Sankey diagram showing the relationship of 3 methylator groups with tumor locations. Size of each rectangle and thickness of the connecting lines represents the number of samples on each group. P value was estimated by Fisher exact test.

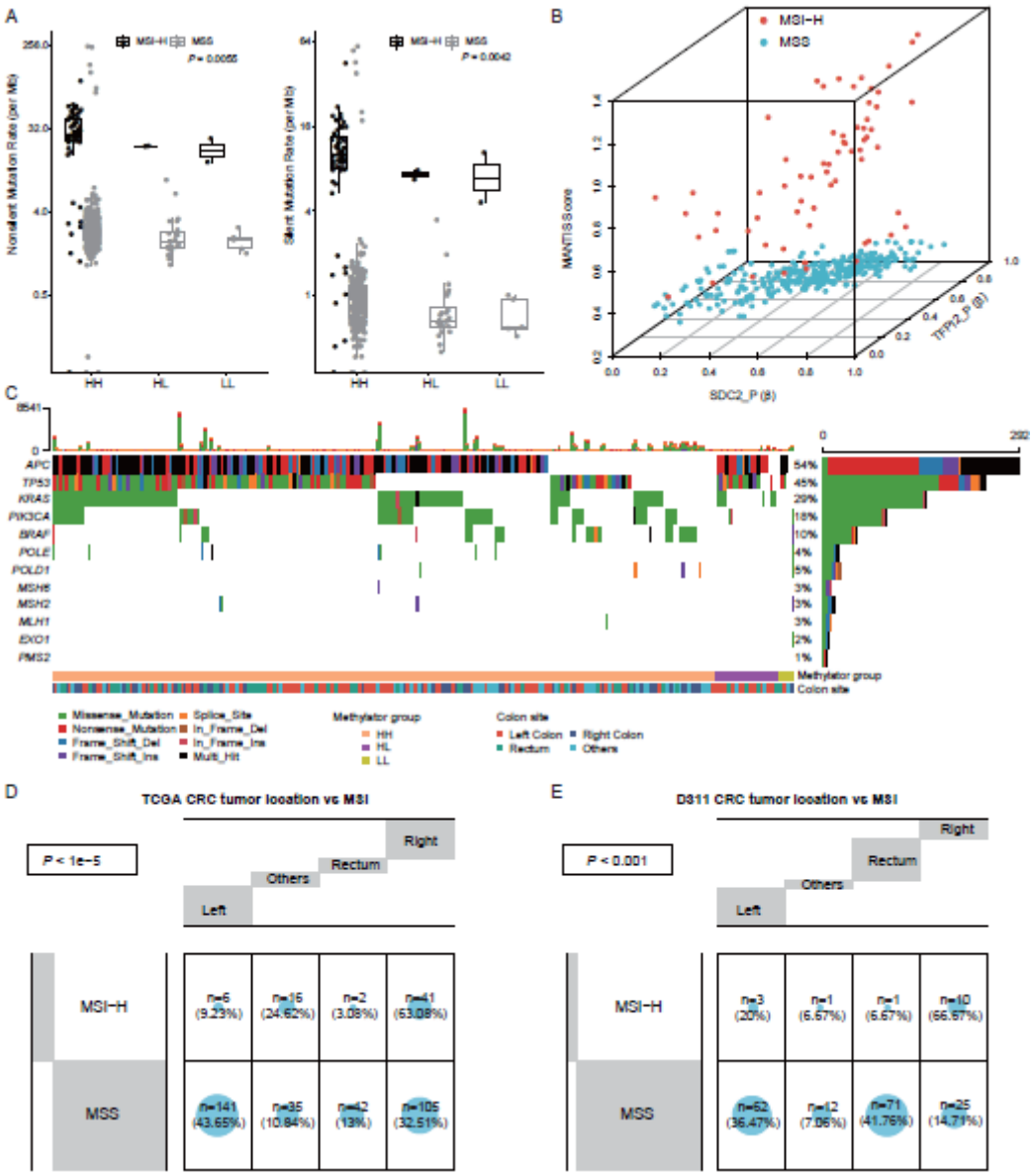

Figure 3

Genomic characteristics in three methylator groups. A: Nonsilent and silent mutation rates in HH, HL and LL groups. B: The correlation of MANTIS score with β values of SDC2 and TFPI2 in TCGA CRCs. C: Somatic mutation profile of five CIMP-related and MMR-related genes in HH, HL and LL groups. D&E: Comparison of tumor location with MSI in TCGA CRCs (D) and D311 CRCs (E), significant p value is calculated using fisher exact test.

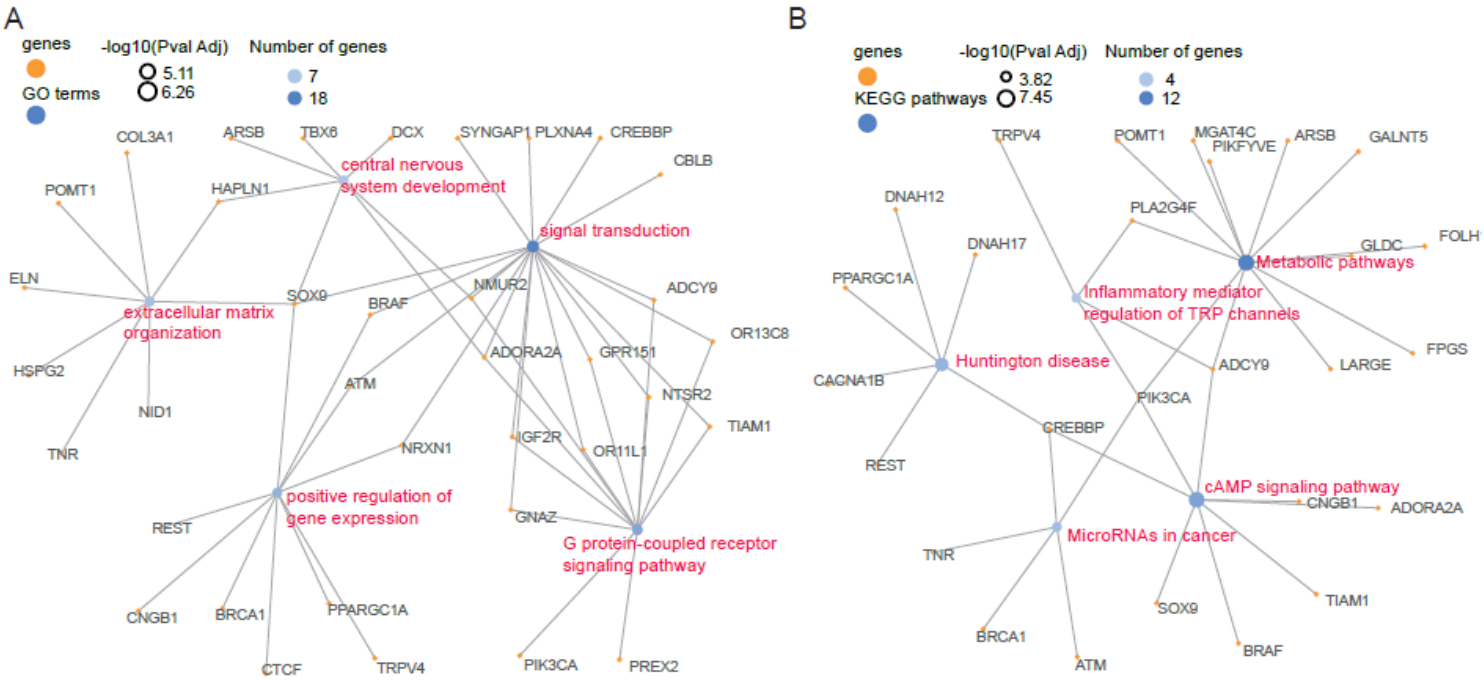

Figure 4

The top 5 enriched GO terms (A) and KEGG pathways (B). The yellow and blue circles indicated mutated genes and enriched terms, respectively. The circle size represented negative log10 of p values. The GO terms and KEGG pathways were showing in red color.

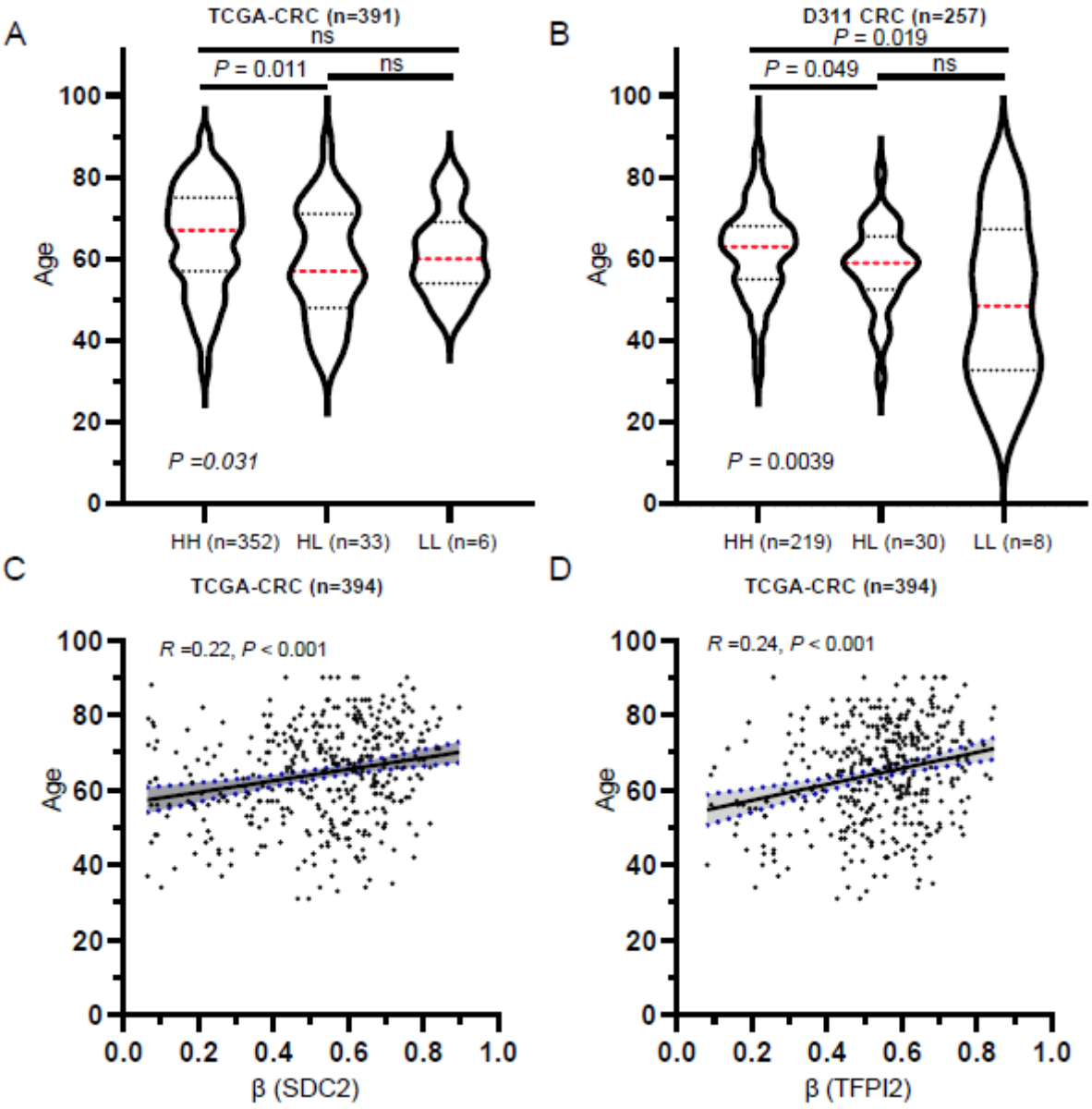

Figure 5

The association of patient age with methylation levels of SDC2 and TFPI2. A&B: Patient ages in HH, HL and LL groups in TCGA CRCs (A) and D311 CRCs (B). Patient age were not available for three samples in TCGA CRCs. The dashed red line indicate median age, while upper and lower dashed lines represent 75% quantile and 25% quantile. C&D: The correlation of patient age with β values of SDC2 and TFPI2 in TCGA CRCs. P value and correlation coefficient was calculated using pearsion' method.

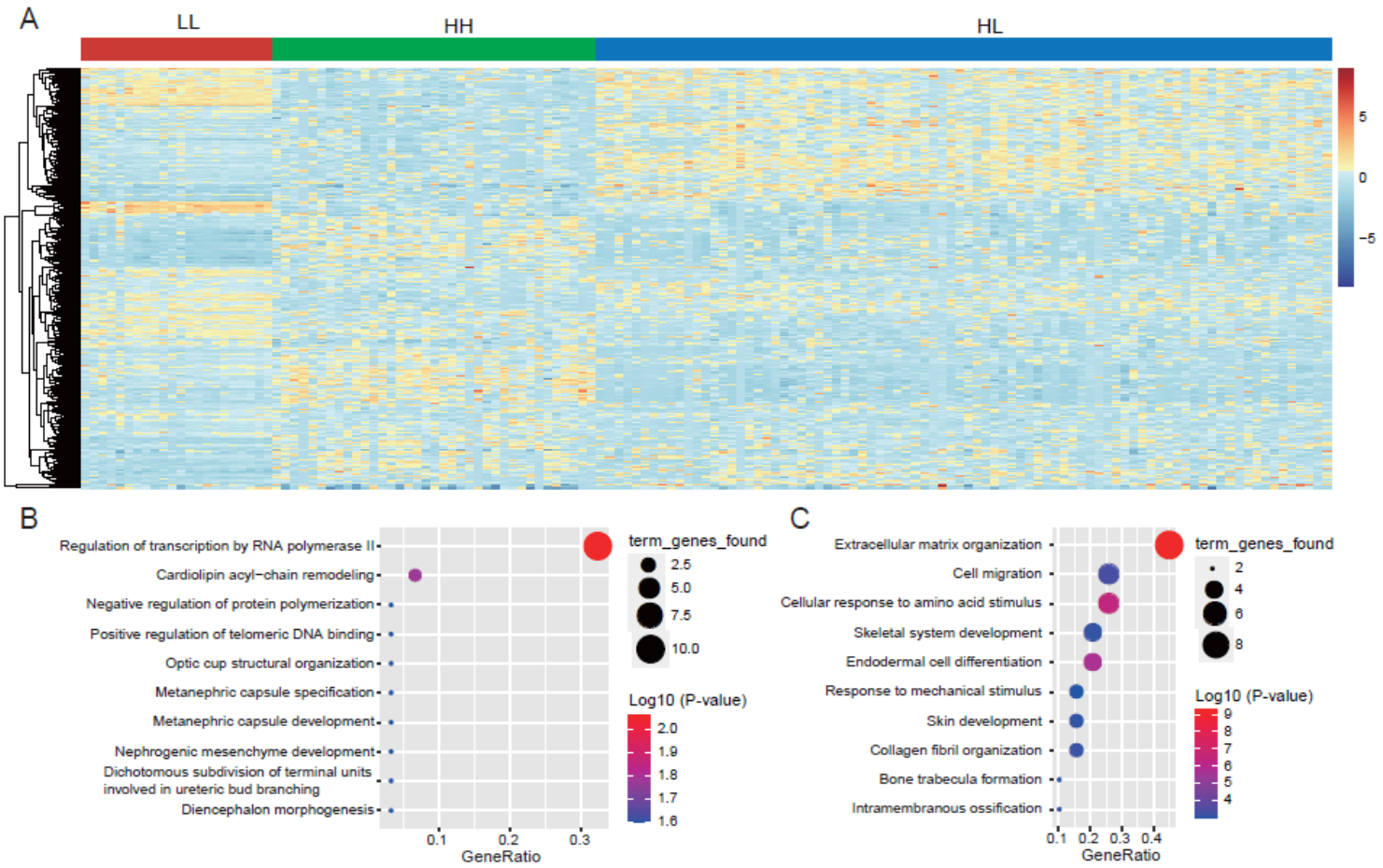

Figure 6

Identification of group-specific DEGs. A: Expression heat map for the HH-, HL- and LL-specific DEGs. The columns and rows represented group-specific DEGs and samples, respectively. B: GO functional enrichment analyses for HH-specific DEGs. C: GO functional enrichment analyses for LL-specific DEGs. The log10 (P-value) of each term is colored according to the legend.

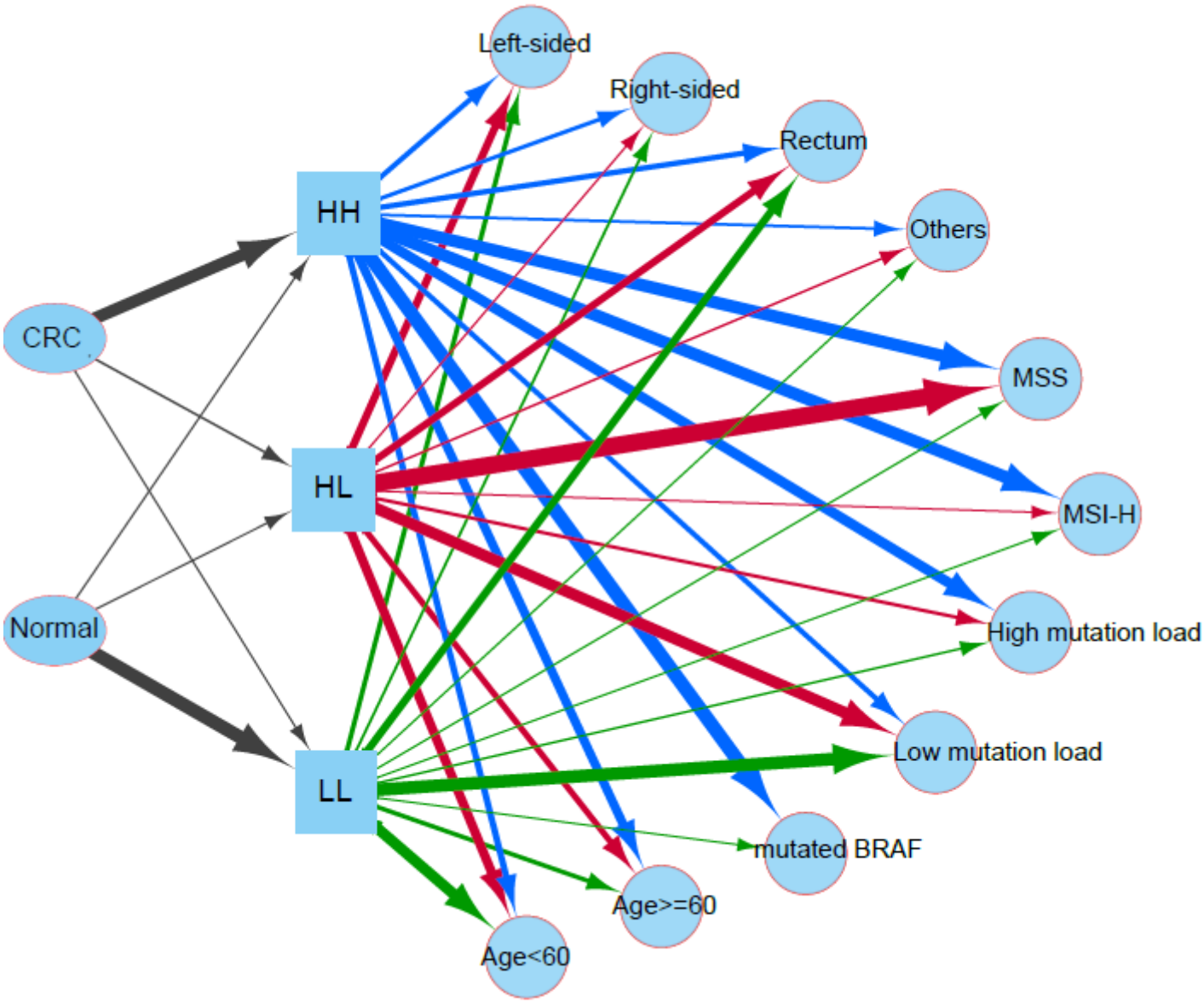

Figure 7

A diagram of possible associations of CIMP phenotype, tumor location, and MSI with the three methylator groups. The sample proportions of three methylator groups stratified by each clinical characteristic of CIMP phenotype, tumor location, and MSI were calculated separately and used to draw this diagram. Ellipses and squares indicated the disease status and methylator groups. Circles represented the patient characteristics. The line width reflected the proportion of samples in a group that has a particular characteristic.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarymaterials.docx