

RESEARCH

Open Access



# The methylation of *SDC2* and *TFPI2* defined three methylator phenotypes of colorectal cancer

Ruixue Lei<sup>1†</sup>, Yanteng Zhao<sup>2†</sup>, Kai Huang<sup>3†</sup>, Qian Wang<sup>4</sup>, Kangkang Wan<sup>5</sup>, Tingting Li<sup>5</sup>, Haijun Yang<sup>1\*</sup> and Xianping Lv<sup>2\*</sup>

## Abstract

**Background:** Methylated *SDC2* and *TFPI2* are widely used for colorectal cancer (CRC) detection. However, they often miss some CRCs, which directly diminishes the sensitivity. Further investigations of the underlying mechanisms leading to the missed samples will facilitate developing more eligible methylation markers.

**Methods:** CRC samples from TCGA and GEO datasets were divided into three groups, High-methylation/ High-methylation (HH), High-methylation/Low-methylation (HL), and Low-methylation/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 current study revealed that the performance of methylation-based markers might be affected by tumor location, patient age, mutation load and MSI, and these respective sides should be considered when developing new methylation markers for CRC detection.

**Keywords:** Colorectal cancer, Methylation, Phenotypes, *SDC2*, *TFPI2*

## Background

Colorectal cancer (CRC) is responsible for over 1 million new cases every year, and around 700,000 deaths occur 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

\*Correspondence: yhj1972@126.com; lvxianping@126.com

<sup>†</sup>Ruixue Lei, Yanteng Zhao and Kai Huang have contributed equally to this manuscript

<sup>1</sup> Department of Pathology, The Fourth Affiliated Hospital of Henan University of Science and Technology, Anyang Tumor Hospital, Anyang, Henan, China

<sup>2</sup> Department of Transfusion, The First Affiliated Hospital of Zhengzhou University, Zhengzhou 450052, China

Full list of author information is available at the end of the article



characterized by complex multifactorial phenotypes and multiple risk factors responses for the development of CRC [4, 5]. Many factors have been reported associated with CRC development, including unhealthy diets, smoking, overweight and obesity, inflammatory bowel disease and gut imbalanced microbiota [6–8]. 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 in China in 2012 [7].

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 [9–12]. Tissue factor pathway inhibitor-2 (*TFPI2*) belongs to the Kunitz-type serine proteinase inhibitor family and is thought to be functional in regulating extracellular matrix digestion and re-modeling by inhibiting a broad spectrum of serine proteinases [13, 14]. Unlike the tumorigenic behaviors of *SDC2* in colon cancer cells, *TFPI2* has been shown as a tumor suppressor gene in several malignant tumors [15–18]. Interestingly, both promoters were frequently found hypermethylated in colon cancer cells compared to normal tissue cells in several epi-genomics studies [19, 20].

The frequently aberrant DNA methylation of *SDC2* and *TFPI2* makes them promise biomarkers for the early detection of CRC, and a few testing kits based on the two targets were also developed [19, 21–24]. However, the sensitivity of any single target (*SDC2* alone or *TFPI2* alone) did not exceed 90%. Several studies have suggested a better performance of combined multi-targets for CRC early detection than a single biomarker [25–27]. We observed an improved detection sensitivity when combining the two targets in our previous studies [28, 29], but not all CRC samples can be detected. 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 the undetected CRC samples will help identify a small subset of tumors missed by the two targets and meanwhile benefit researchers to discover more eligible methylation markers. 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 TCGA Bioinformatics R package [30]. The platform of methylation data from TCGA is Illumina Human Methylation 450 Beadchip (450 K array), and we also searched the Gene Expression Omnibus (GEO; <http://www.ncbi.nlm.nih.gov/geo/>) for eligible datasets that are generated by 450 K array. Two GEO datasets, GSE48684 [31] and GSE79740 [32] were then downloaded because of their available clinical information. Empirical Bayes (EB) batch adjustment along with a two-step quantile normalization method [33] was conducted for batch effect removal before GSE48684 and GSE79740 datasets were merged as one set. Missing values of the 450 k array were inferred and fulfilled by the Bayesian Network structure learning algorithm [34]. All samples without clinical information were removed. The preprocessed data used in this study is 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 participants signed 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 includes age, gender, tumor size, tumor location, grade, and MSI status. Detailed demographic and clinical features of the participants are 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 [35]. This study was approved by the medical ethics committee of Zhongnan Hospital of Wuhan University (No. 2019099).

### 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 450 k array, where 14 and 21 probes are within the two kilo-base upstream of transcript start site (TSS), respectively (Additional file 1: Table S1). We focused on the regions that showed the largest difference between normal and tumor samples to determine the appropriate target regions. First, the normal and tumor methylation values for each probe ( $\beta_{\text{Tumor}}$  and  $\beta_{\text{Normal}}$ ) were calculated using average  $\beta$  values of the normal and

**Table 1** Clinical characteristics of subjects

Participant's 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%)

tumor samples in each dataset. Next, the probes with  $\Delta\beta$  ( $\beta_{\text{Tumor}} - \beta_{\text{Normal}}) \geq 0.3$  were selected as the most varied probes. The threshold of 0.3 was chosen in this study as only about 5% of the top-ranked probes showed  $\Delta\beta \geq 0.3$ . Then, the most varied regions were identified according to the probe genomic coordinates. Three probes of *SDC2* that satisfied the criteria, cg16935295, cg04261408, and cg10292139, were found in a contiguous genomic position. Since adjacent CpG sites usually exhibited similar methylation status, also known as methylation block, the region consisting of four probes, cg16935295, cg04261408, cg14625631, and cg10292139, was finally selected as the target. The mean  $\beta$  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  $\beta$  values were also calculated as the methylation levels of *TFPI2*.

**Identification of appropriate  $\beta$  threshold**

In this study, the appropriate  $\beta$  threshold to define high-methylation or low-methylation was determined according to previous studies [36, 37]. Briefly, the distribution parameters of the methylation M-values were estimated by a Gaussian mixed model to determine the optimal M-value and the corresponding  $\beta$ -value.  $\beta$ -value and M-value are two metrics often used for the Infinium methylation array.  $\beta$ -value is more easily interpreted as an approximation of the overall methylation percentage 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  $\beta$ -value and M-value of 450 K array on TCGA CRC dataset.

**Methylation-specific PCR**

The genomic DNA was extracted by the 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,  $\beta$ -actin [38] was used as an internal control. Specific primers and probes for the target region of *SDC2* and *TFPI2* were designed as shown in Additional file 1: Table S2. 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 CRC cohort were pre-processed by removing zero expressed genes in more than 90% of all samples before performing differential expression analysis. DESeq2 [39] (V1.30.0) R package was used to perform a pairwise comparison between

the three methylator groups to identify differentially expressed genes (DEGs). Genes with adjusted  $P$  values estimated by false discovery rate were recognized as candidate DEGs. Functional enrichment analysis of Gene Ontology and KEGG pathway was then conducted for the candidate DEGs using the web-based tool of GeneCodis [40].

### 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 the non-random 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  $\beta$  threshold. If the  $\beta$  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 [41]. 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/source\\_code\\_v1.R](https://github.com/amsinfor/methylator_group/blob/master/source_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 (Additional file 1: Table S1). The average  $\beta$  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 (Additional file 1: Figure S1). A lower  $\Delta\beta$  value was found for *TFPI2* than *SDC2*, which might attribute to its higher background methylation level on normal controls (Fig. 1A, B). The  $\beta$  values were severely compressed around 0.05 and 0.95, and no obvious bimodal was observed (Fig. 1C). In contrast, the histogram of M values clearly showed a bimodal distribution (Fig. 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 the  $\beta$  value of 0.205, the optimal boundary was determined between the two peaks (methylated peak and unmethylated peak). These findings implied that 0.2 could be an appropriate  $\beta$  threshold, which is consistent with the studies of Pan Du et al. and Sarah

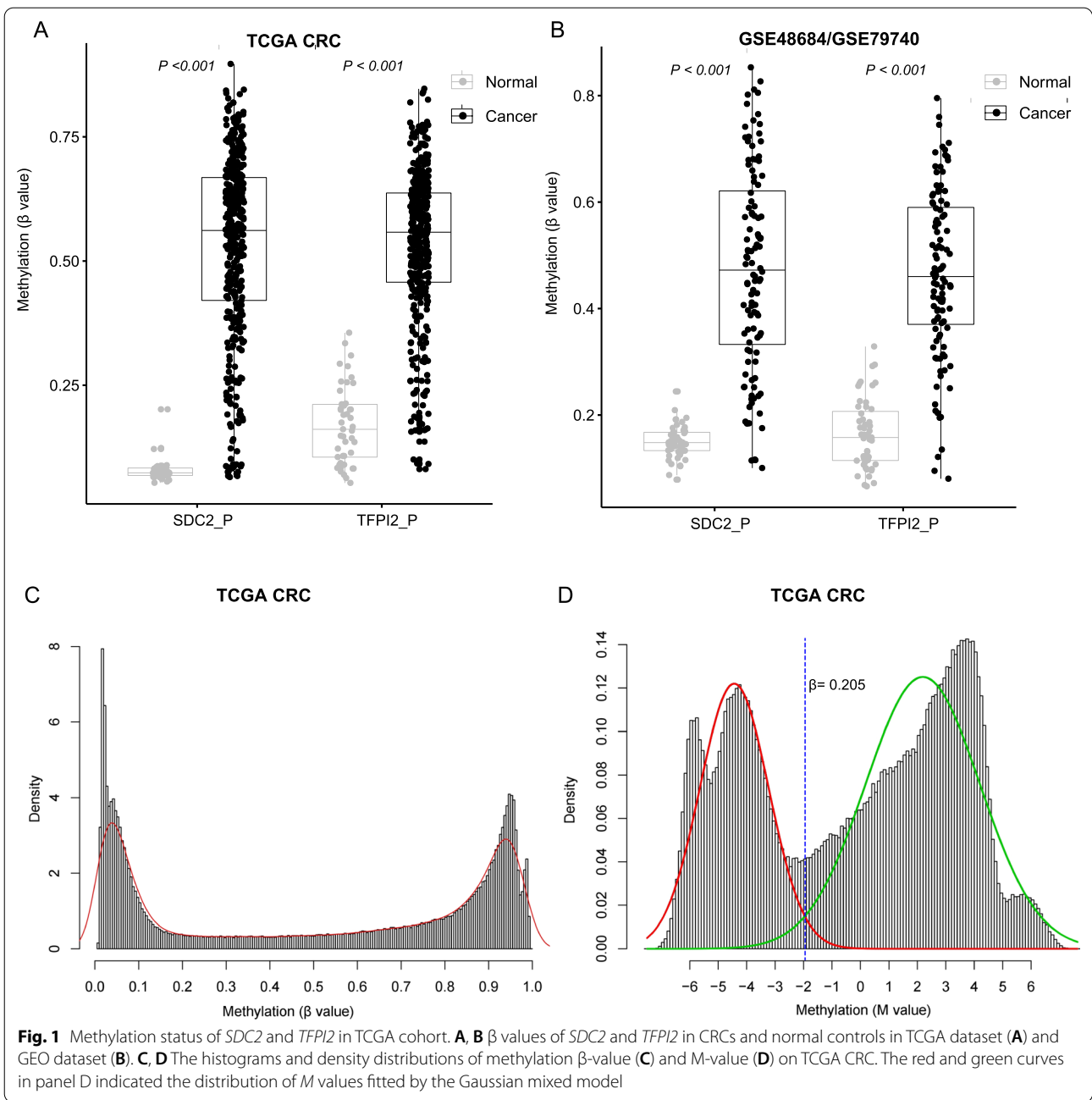
Dedeurwaerder et al. (see methods). We observed that high-methylated *SDC2* and *TFPI2* occurred in more than 85% of CRC samples in three datasets (average 88.21%). Approximately 10% of CRCs had single high-methylated gene (*SDC2* or *TFPI2*) (average 9.24%), and less than 3% of CRCs harbored low-methylated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

### The association of methylator groups with tumor location

For early detection, HL and LL groups are very important because they can directly affect a given biomarker's sensitivity. Therefore, we first compared the three methylator groups with tumor locations. Overall, no preference was observed for the distribution of HH group samples over different tumor locations in all three datasets, and this was similar for the LL group samples (Fig. 2A–C, Additional file 1: Table S3). Interestingly, the samples of HL group were more frequently originated from left-sided colon, while only a tiny fraction was from right-sided colon and rectum. These results indicated that the left-sided CRCs were more likely missed by the two targets, as the HL group samples preferred to locate in the left-sided colon.

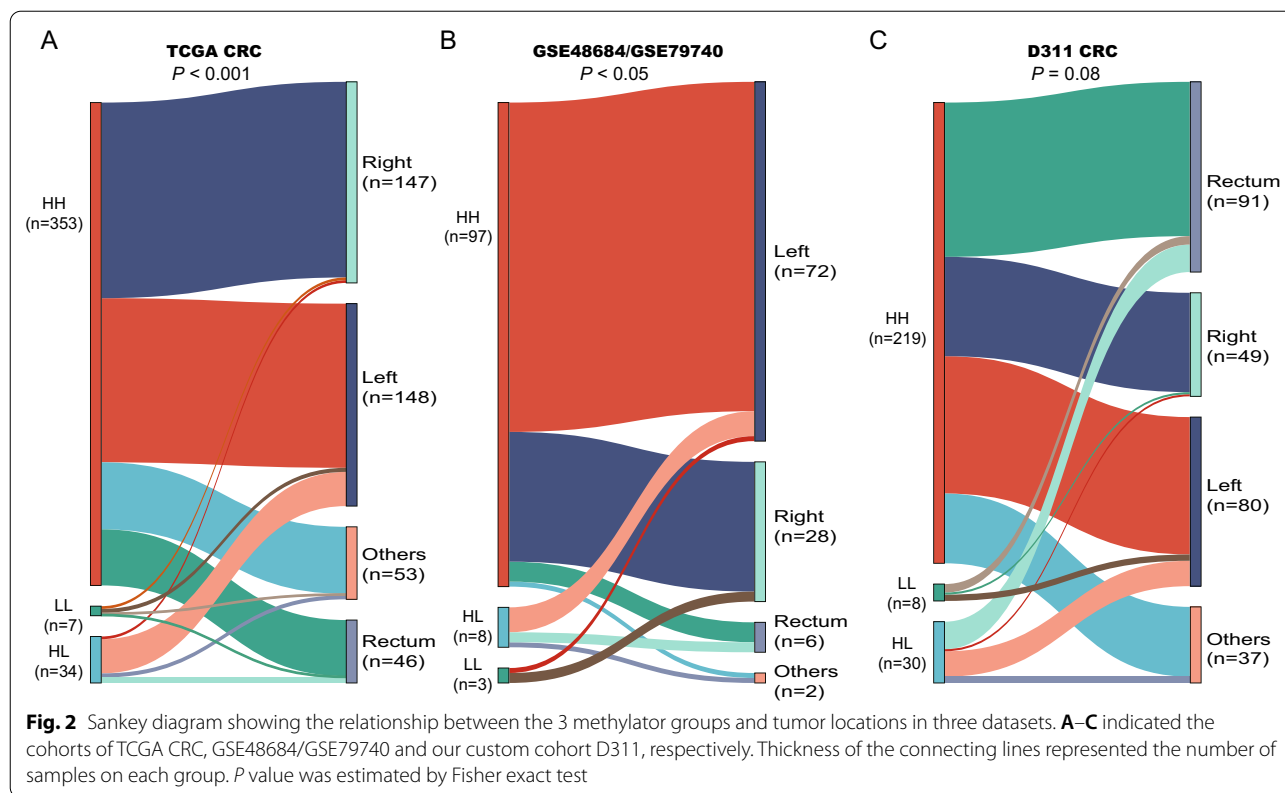
### The association of methylator groups with genomic variations

Microsatellite instability and hypermutation have been regarded as important molecular characteristics of CRCs. Compared to the other two groups, HH group CRCs presented the highest mutation load (Fig. 3A,  $P < 0.05$ ) in the TCGA CRC dataset. According to the MANTIS scores [42], which is used to evaluate the MSI status, we grouped the TCGA CRCs into MSI-H if the score is  $>0.4$  and MSS if the score is  $\leq 0.4$ . The  $\beta$  values of *SDC2* and *TFPI2* showed a high concordance with MANTIS scores in the MSI-H group (Fig. 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 [43], 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 in LL group (Additional file 1: Figure S2). 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 (Fig. 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 preferred in the right-sided colon (Fig. 3D, E,  $P < 0.001$ ), which possibly elucidated the potential mechanism that HL group CRCs were



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

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



mainly in left-sided colon and MSI-H CRCs were less likely in HL group. Additionally, 111 mutated genes were significantly enriched between the three groups (Additional file 1: Table S4), 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 (Fig. 4A, B).

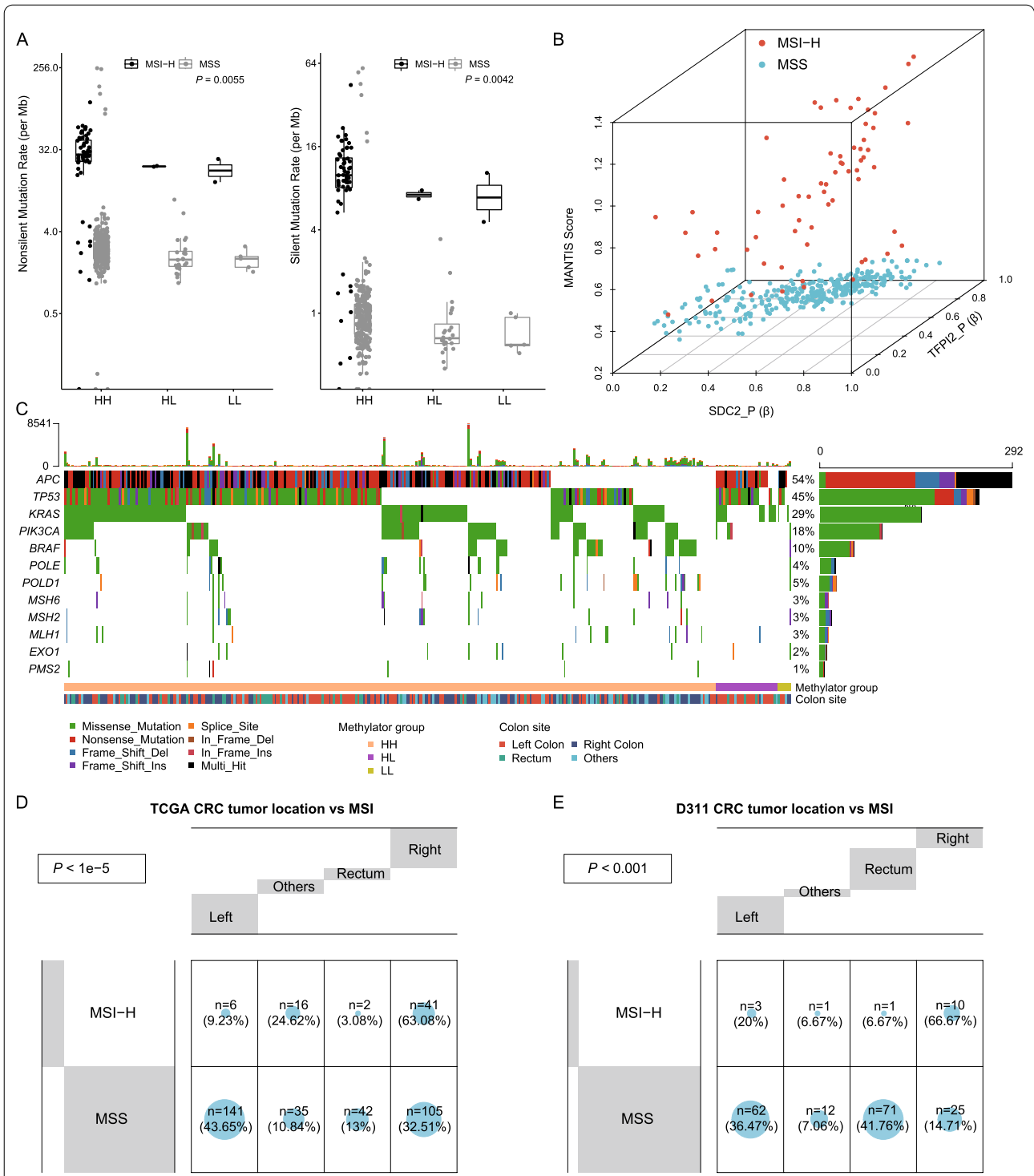
**The association of methylator groups with patient age**

Patient age is one of the risk factors for colon cancer, and we found a significantly older age in HH group patients, while it was the youngest for LL group patients (Fig. 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 (Fig. 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*, were used as a control to quantify the relative methylation levels of these two genes, also called ΔCt. The 2<sup>-ΔCt</sup> was then calculated for correlation analysis. Similar results were observed on the D311 dataset, though no strong

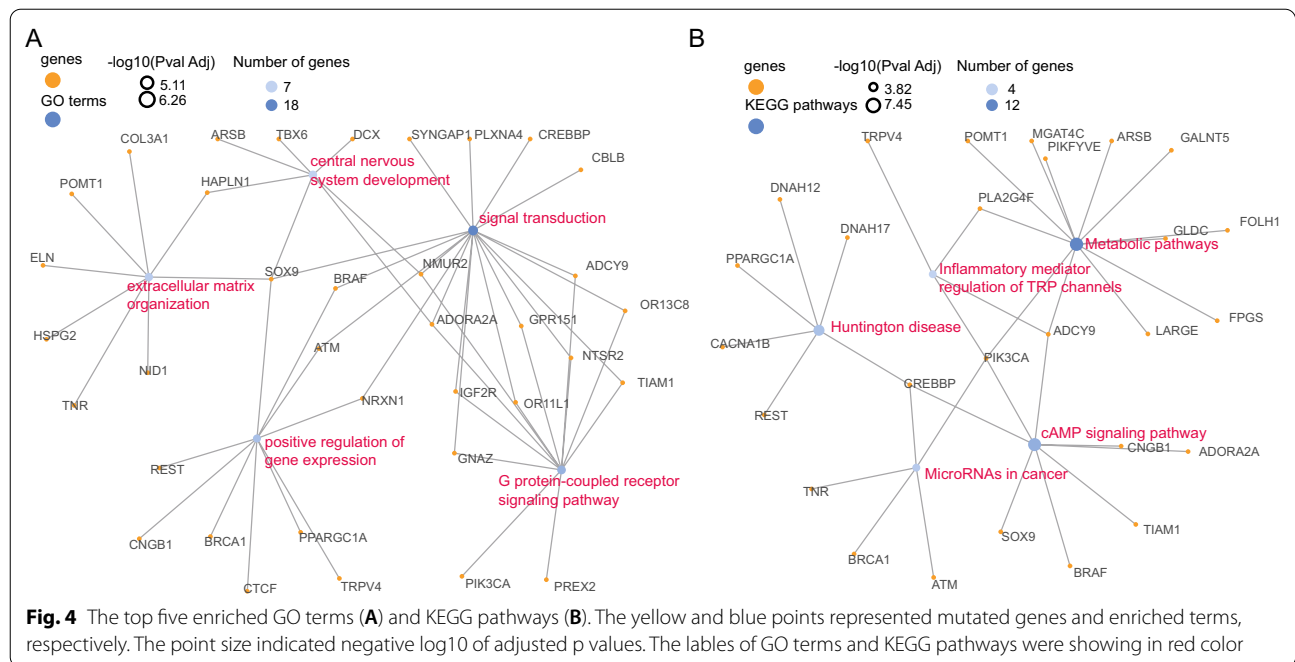
correlation was presented (Additional file 1: Figure S3), indicating that young patients might be more likely to be miss-detected.

**Identification of DEGs among the three methylator groups**

We performed differential expression analysis using the gene expression profiles of TCGA CRC dataset to identify group-specific DEGs. DEGs among the three groups were first identified, and then their average expression levels were calculated separately for each group. The group-specific DEGs were defined as the genes in one group that presented the highest average expression values. 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 (Fig. 6A). Functional enrichment analysis indicated that HH-specific DEGs were mainly related to the regulation of transcription and other processes (Fig. 6B), while LL-specific DEGs are enriched in the biological processes of extracellular matrix interaction (ECM) and cell migration (Fig. 6C). No GO terms and KEGG pathways were significantly enriched for HL-specific genes. 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.



**Fig. 3** Genomic characteristics of three methylator groups in TCGA cohort. **A** Non-silent and silent mutation rates of HH, HL and LL groups. **B** The correlation of MANTIS score with  $\beta$  values of *SDC2* and *TFP12* 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



### Discussion

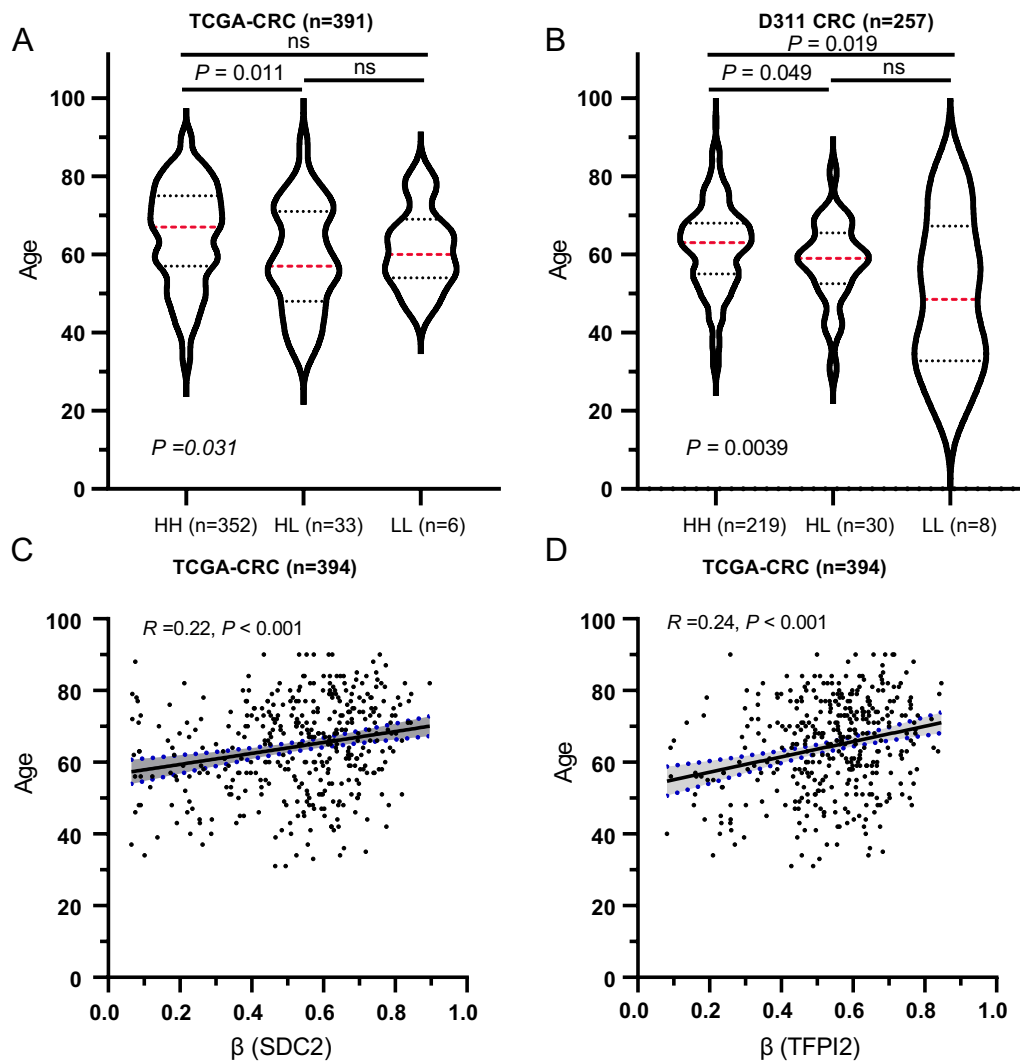
Quantifying aberrantly methylated genes was a feasible method for the early detection of CRCs. Several biomarkers have been demonstrated excellent performance in CRC early detection [24, 26, 27]. However, they often suffer the limitation with some positive samples miss detected, which directly affects 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 excellent performance of methylated *SDC2* and *TFPI2* alone or combined with other biomarkers for stool-based CRC detection [22, 44]. 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 a single target, nearly 10% of the samples had only one gene methylated, and here we defined them as HL group. Further analysis showed that more samples harboring high-methylated *TFPI2* and low-methylated *SDC2* were from HL group (about three 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 [45, 46], 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 [47]. Moreover, the incidence between proximal and distal CRCs also differs in age and sex. Epidemiologic studies have revealed a higher incidence rate of distal CRC on males than females, and the proximal patients were older than distal patients, although increased ages were associated with the incidence rates of all subsites [48–50]. These data reflect extensive distinctions in molecular pathogenesis between the two anatomical locations, which might significantly impact



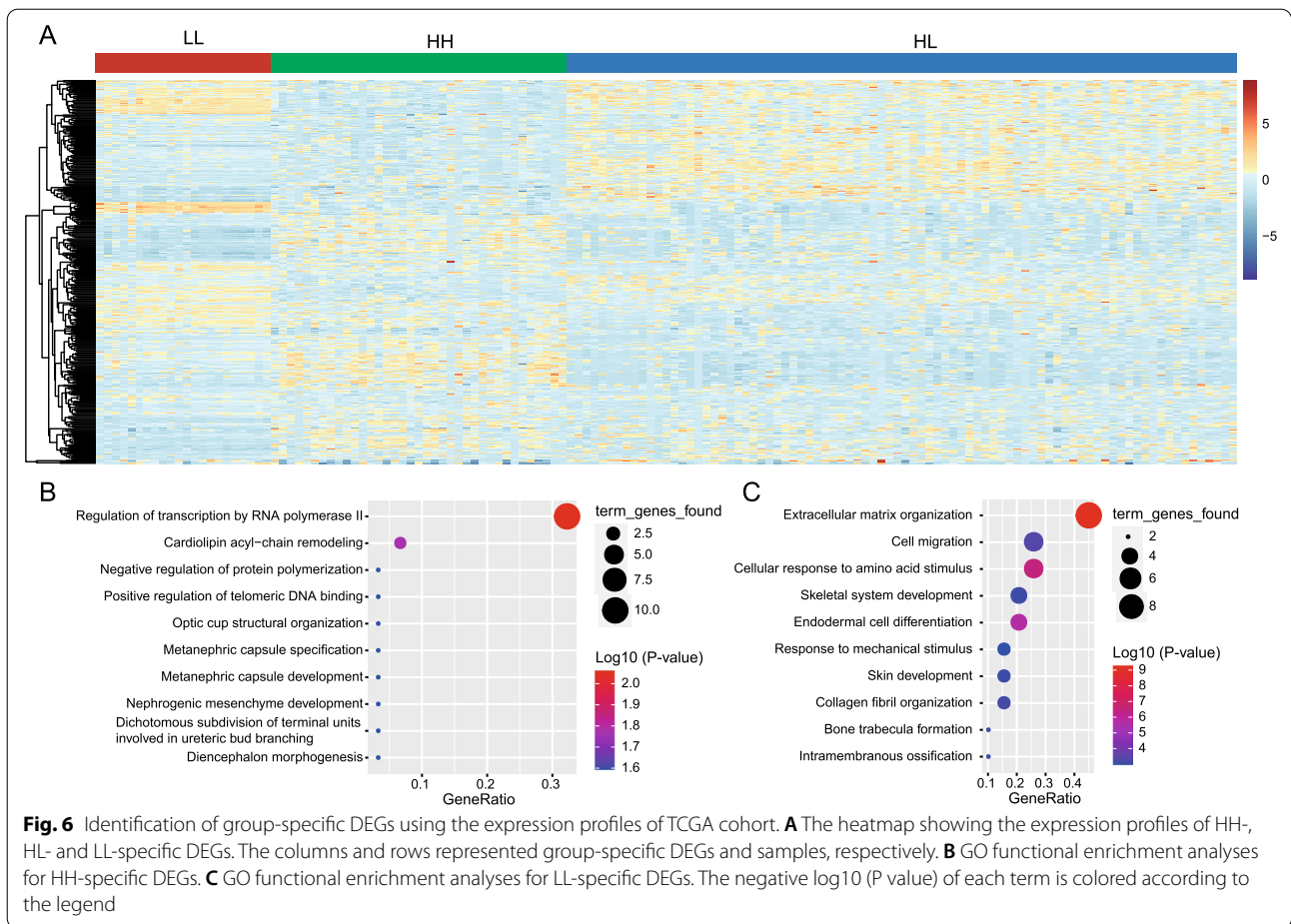


**Fig. 5** The association of patient age with methylation levels of *SDC2* and *TFPI2*. **A, B** Patient ages of HH, HL and LL groups in TCGA cohort (**A**) and D311 cohort (**B**). Patient age were not available for three samples in TCGA cohort. The dashed red line indicated median age. The upper and lower dashed lines represented 75% and 25% quantile. **C, D** The correlation of patient age with  $\beta$  values of *SDC2* and *TFPI2* in TCGA cohort. *P* value and correlation coefficient was calculated using *Pearson*' method

tumorigenesis. 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 [51, 52]. 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, and HL group CRCs were dominated by high-methylated *TFPI2* and low-methylated *SDC2*, which 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 *BRAF* 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 [53]. Therefore, the methylator phenotypes may be a stratification factor in clinical therapies. It has been reported that CIMP tumors showed significant associations with *BRAF* mutations, MSI-H [52]. These



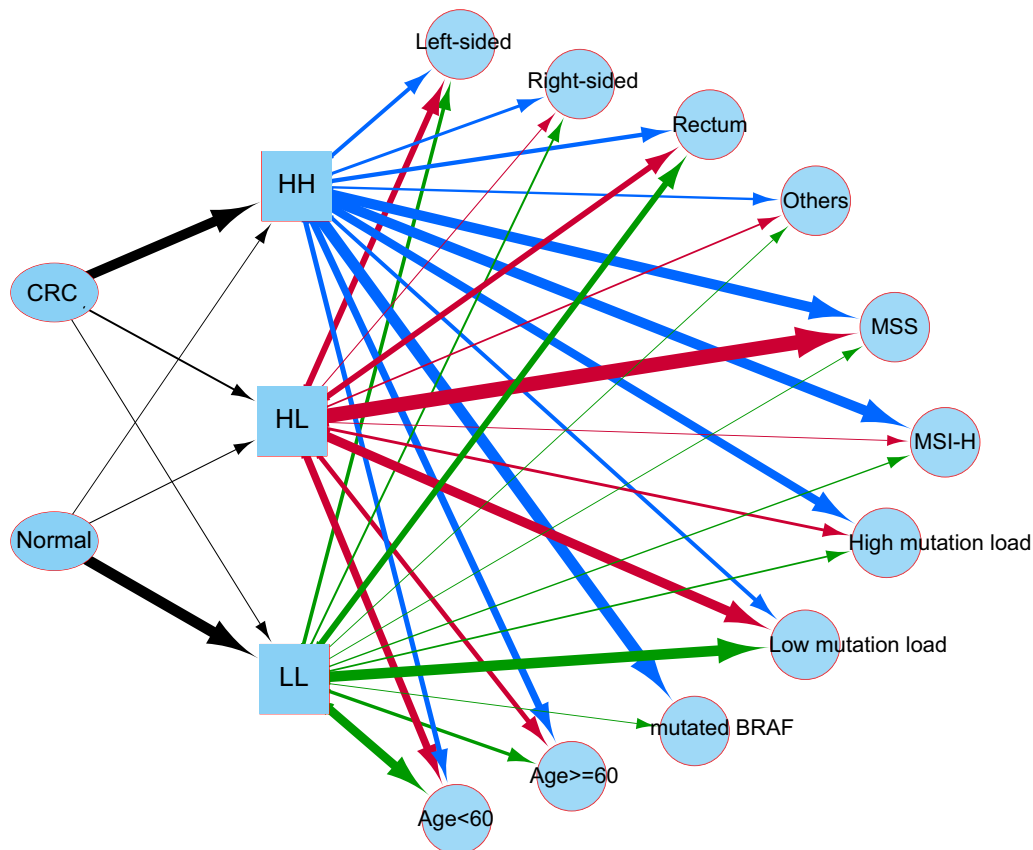
results, on the other hand, confirmed that molecular events, such as epigenetic variations, instability, aberrant DNA mutations, and MSI, are coupled with each other.

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 shown crucial roles of *SDC2* and *TFPI2* in the interaction of extracellular matrix with cell plasma [14, 54]. Besides, we also observed a significant enrichment of 111 mutant genes on GO term of extracellular matrix organization, which might, on 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 highly heterogeneous disease, 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, the midgut and the hindgut for the proximal and distal colon, respectively [47, 55]. 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 (Fig. 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 suggested that the performance of methylation-based markers might be affected by tumor locations, CIMP phenotypes, patient age, mutation load, and MSI. This study also highlighted the different methylation status of *SDC2* and *TFPI2* between left and right-sided CRCs, which could lead to a detection bias in clinical practice as the right-sided



**Fig. 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 according to the D311 cohort 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

tumors preferred to be detected by methylated *SDC2* and the left-sided by methylated *TFPI2*. Considering that left-sided CRCs are more prevalent in East Asian populations [56–58], the bias may enlarge the risk of CRC being missed when the two markers are applied for different ethnic groups. Therefore, examining physicians need to thoroughly evaluate the impact of multiple factors, including anatomical location, on a methylation-based test.

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12876-022-02175-3>.

**Additional file 1.** Supplementary Figures and Tables.

### Acknowledgements

We declare that submitted manuscript and other materials are not under consideration for publication elsewhere.

### Authors' contributions

Study design: HJY. Data collection and analysis: RXL, YTZ, KKW, TTL and QW. Manuscript writing: RXL, YTZ, KH and KKW. Manuscript and results revising: HJY and XPL. All authors read and approved the final manuscript.

### Funding

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

### 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.

### Declarations

#### Competing interests

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.

#### 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.

#### 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.

#### Author details

<sup>1</sup>Department of Pathology, The Fourth Affiliated Hospital of Henan University of Science and Technology, Anyang Tumor Hospital, Anyang, Henan, China. <sup>2</sup>Department of Transfusion, The First Affiliated Hospital of Zhengzhou University, Zhengzhou 450052, China. <sup>3</sup>Department of Gastrointestinal Surgery, Department of General Surgery, The First Affiliated Hospital of Anhui Medical University, Hefei, China. <sup>4</sup>Department of Gastroenterology, Zhongnan Hospital of Wuhan University, Wuhan, Hubei, China. <sup>5</sup>Wuhan Ammunition Life-Tech Company, Ltd., Wuhan, Hubei, China.

Received: 27 October 2021 Accepted: 17 February 2022

Published online: 28 February 2022

#### References

- Navarro M, Nicolas A, Ferrandez A, Lanas A. Colorectal cancer population screening programs worldwide in 2016: an update. *World J Gastroenterol*. 2017;23(20):3632–42.
- 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):E359–86.
- 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):235–43.
- Binefa G, Rodríguez-Moranta F, Teule A, Medina-Hayas M. Colorectal cancer: from prevention to personalized medicine. *World J Gastroenterol*. 2014;20(22):6786–808.
- Ang PW, Loh M, Liem N, Lim PL, Griew 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.
- 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):5389–96.
- 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.
- Boland CR, Goel A. Microsatellite instability in colorectal cancer. *Gastroenterology*. 2010;138(6):2073–87.
- 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):3874–86.
- 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):29730–6.
- 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.
- 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.
- Sierko E, Wojtukiewicz MZ, Kisiel W. The role of tissue factor pathway inhibitor-2 in cancer biology. *Semin Thromb Hemost*. 2007;33(7):653–9.
- Chand HS, Foster DC, Kisiel W. Structure, function and biology of tissue factor pathway inhibitor-2. *Thromb Haemost*. 2005;94(6):1122–30.
- 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.
- Lavergne M, Jourdan ML, Blechet C, Guyetant S, Pape AL, Heuze-Vourc'h N, Courty Y, Lerondel S, Sobilo J, Lochmann 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.
- Vaitkiene P, Skiriute D, Skauminas K, Tamasauskas A. Associations between TFPI-2 methylation and poor prognosis in glioblastomas. *Medicina (Kaunas)*. 2012;48(7):345–9.
- 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):779–83.
- 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.
- 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):1205–7.
- Glöckner SC, Dhir M, Yi JM, McGarvey KE, Van Neste L, Louwagie J, Chan TA, Kleeberger W, de Bruine 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):4691–9.
- 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.
- 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 Epigenet*. 2019;11(1):51.
- 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.
- 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):5264–71.
- 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):2423–32.
- 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):5619–28.
- Wang Z, Shang J, Zhang G, Kong L, Zhang F, Guo Y, Dou Y, Lin J. Evaluating the clinical performance of a dual-target stool DNA test for colorectal cancer detection. *J Mol Diagn*. 2022;24(2):131–43.
- Zhang L, Dong L, Lu C, Huang W, Yang C, Wang Q, Wang Q, Lei R, Sun R, Wan K et al: methylation of *SDC2/TFPI2* and its diagnostic value in colorectal tumorous lesions. *Front Mol Biosci* 2021, 8.
- Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, Sabedot TS, Malta TM, Pagnotta SM, Castiglioni I, et al. TCGAAbiolinks: an R/bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res*. 2016;44(8):e71.
- 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):418–29.
- 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):203–9.
- 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 Genom*. 2011;4:84.
- 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):1250–2.

35. 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):1343–9.
36. 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.
37. Dedeurwaerder S, Defrance M, Calonne E, Denis H, Sotiriou C, Fuks F. Evaluation of the Infinium methylation 450K technology. *Epigenomics-UK*. 2011;3(6):771–84.
38. 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):7141–52.
39. 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.
40. 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):W478–83.
41. 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.
42. 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):7452–63.
43. Toyota M, Ahuja N, Ohe-Toyota M, Herman JG, Baylin SB, Issa JJ. CpG island methylator phenotype in colorectal cancer. *Proc Natl Acad Sci*. 1999;96(15):8681.
44. 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 Epigenet*. 2017;9(1):126.
45. 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):1287–97.
46. 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):467–75.
47. Missiaglia E, Jacobs B, D'Ario G, Di Narzo AF, Sonesson 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.
48. Bufill JA. Colorectal cancer: evidence for distinct genetic categories based on proximal or distal tumor location. *Ann Intern Med*. 1990;113(10):779–88.
49. Lee GH, Malietzis G, Askari A, Bernardo D, Al-Hassi HO, Clark SK. Is right-sided colon cancer different to left-sided colorectal cancer? A systematic review. *Eur J Surg Oncol*. 2015;41(3):300–8.
50. Nelson RL, Dollear T, Freels S, Persky V. The relation of age, race, and gender to the subsite location of colorectal carcinoma. *Cancer Am Cancer Soc*. 1997;80(2):193–7.
51. 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.
52. 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):1188–201.
53. Sanz-Garcia E, Argiles G, Elez E, Tabernero J. BRAF mutant colorectal cancer: prognosis, treatment, and new perspectives. *Ann Oncol*. 2017;28(11):2648–57.
54. Xian X, Gopal S, Couchman JR. Syndecans as receptors and organizers of the extracellular matrix. *Cell Tissue Res*. 2010;339(1):31–46.
55. 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.
56. Bai Y, Gao J, Zou D, Li Z. Distribution trends of colorectal adenoma and cancer: a colonoscopy database analysis of 11 025 Chinese patients. *J Gastroen Hepatol*. 2010;25(10):1668–73.
57. Soon M, Kozarek RA, Ayub K, Soon A, Lin T, Lin OS. Screening colonoscopy in chinese and western patients: a comparative study. *Off J Am College Gastroenterol ACG*. 2005;100(12):2749–55.
58. Qing S, Rao K, Jiang H, Wexner SD. Racial differences in the anatomical distribution of colorectal cancer: a study of differences between American and Chinese patients. *World J Gastroenterol*. 2003;9(4):721–5.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

