Multiple gene-specific DNA methylation in blood leukocytes and colorectal cancer risk: a case-control study in China

Yupeng Liu1, Yibaina Wang1, Fulan Hu1, Hongru Sun1, Zuoming Zhang1, Xuan Wang1, Xiang Luo1, Lin Zhu1, Rong Huang1, Yan Li1, Guangxiao Li1, Xia Li1, Shangqun Lin1, Fan Wang1, Yanhong Liu2, Jiesheng Rong3, Huiping Yuan4 and Yashuang Zhao1

1Department of Epidemiology, Public Health College, Harbin Medical University, Harbin 150081, Heilongjiang Province, The People’s Republic of China
2Department of Clinical Laboratory, The Second Affiliated Hospital of Harbin Medical University, Harbin 150081, Heilongjiang Province, The People’s Republic of China
3Department of Orthopedics Surgery, The Second Affiliated Hospital of Harbin Medical University, Harbin 150081, Heilongjiang Province, The People’s Republic of China
4Key Laboratory of Ophthalmology, Department of Ophthalmology, The Second Affiliated Hospital of Harbin Medical University, Harbin 150081, Heilongjiang Province, The People’s Republic of China

Correspondence to: Yashuang Zhao, email: zhao_yashuang@263.net

Keywords: colorectal cancer risk, DNA methylation, leukocytes

Received: February 03, 2017 Accepted: April 07, 2017 Published: May 22, 2017

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

ABSTRACT

The relationship between gene-specific DNA methylation in peripheral blood leukocytes and colorectal cancer (CRC) susceptibility is unclear. In this case-control study, the methylation status of a panel of 10 CRC-related genes in 428 CRC cases and 428 cancer-free controls were detected with methylation-sensitive high-resolution melting analysis. We calculated a weighted methylation risk score (MRS) that comprehensively combined the methylation status of the panel of 10 genes and found that the MRS_10 was significantly associated with CRC risk. Compared with MRS-Low group, MRS-High group and MRS-Medium group exhibited a 6.51-fold (95% CI, 3.77-11.27) and 3.85-fold (95% CI, 2.72-5.45) increased risk of CRC, respectively. Moreover, the CRC risk increased with increasing MRS_10 ($P_{\text{trend}} < 0.0001$). Stratified analyses demonstrated that the significant association retained in both men and women, younger and older, and normal weight or underweight and overweight or obese subjects. The area under the receiver operating characteristic curves for the MRS_10 model was 69.04% (95% CI, 65.57-72.66%) and the combined EF and MRS_10 model yielded an AUC of 79.12% (95% CI, 76.22-82.15%). Together, the panel of 10 gene-specific DNA methylation in leukocytes was strongly associated with the risk of CRC and might be a useful marker of susceptibility for CRC.

INTRODUCTION

Colorectal cancer (CRC), with an estimated 1,360,602 newly diagnosed cases and 693,933 deaths in 2012, is the third-most common cancer in men and the second-most in women worldwide [1]. In China, CRC is the fifth-most common cancer in men and the fourth-most in women, with an estimated 376,300 newly diagnosed cases and 191,000 deaths in 2015 [2]. The cancerization of colon mucosal epithelial cells is a complex and multifactorial, gradual process, characterized by the accumulation of cancer-specific genetic and epigenetic alterations. Epigenetic DNA modifications, including aberrant DNA methylation, are recognized as major and causal epigenetic events that occur during CRC initiation [3]. Accumulating evidence suggests that the global hypomethylation of DNA might induce chromosomal instability and gene-specific hypermethylation can silence tumor suppressor genes, all of which might contribute to CRC formation.
Until now, studies have mainly focused on tumor-derived DNA methylation changes, and studied the relationship between DNA methylation status in tumor tissue and prognosis of various cancers including CRC. The aberrant methylated genes in tumor tissue are commonly involved in aspects of cell function such as cell cycle regulation (CDKN2A (also known as p16) and IGF2), DNA mismatch repair (MLH1 and MGMT), apoptosis (DAPK1), cell adhesion (CDH1) and signal transduction (APC and WIF1) in CRC [4–10]. It is clear that tumors do not develop as an isolated phenomenon in their target tissue, other organ systems including the immune system (such as peripheral blood leukocytes) are also involved in tumor initiation [11]. In addition, altered DNA methylation in peripheral blood leukocytes is associated with environmental exposures encountered throughout life [12, 13]. Meanwhile, several experiments in vitro showed that environmental factors may disrupt epigenetic balance and may cause methylation abnormalities, and these aberrations have been associated with cancer susceptibility [14, 15]. Therefore, DNA methylation alterations in leukocytes may reflect epigenetic modifications, environmental exposures, or interactions between these factors that increase disease susceptibility. However, only a few studies have assessed the association between DNA methylation in leukocytes and the risk of CRC or colorectal adenomas. Some studies focused on genomic methylation of leukocyte DNA in relation to the risk CRC [16–19] or colorectal adenomas [20, 21], and other studies have reported an association between gene-specific methylation in leukocytes and the risk of CRC [22–26] or colorectal adenomas [27–29]. In addition, recently, several studies have reported gene-specific methylation alterations in leukocytes from patients with various cancers, including bladder, breast, renal, and head and neck cancer [30–34]. The results of these studies did imply aberrant methylation of multiple genes in leukocytes might predispose toward susceptibility to CRC, like genetic variants of germline DNA. Therefore, we carried out this case-control study to investigate the associations between leukocyte-derived DNA methylation in a panel of 10 genes and the risk of CRC.

RESULTS

Main characteristics of the participants

Table 1 shows the basic characteristics of the cases and their matched controls. The CRC cases exhibited an overall lower body mass index than the controls. Of the 914 eligible DNA samples, methylation status of the 10 genes was successfully assessed in 856 samples (93.65%), which were included in the final analysis. The characteristics of demographic variables and questionnaire-derived variables before and after multiple imputations are listed in Supplementary Table 1.

DNA methylation and CRC risk

CRC risk was significantly associated with the methylation of DAPK1, IGF2, MINT31, NEUROG1 and WIF1 (Table 2). A marginally significant association was observed for MGMT. For the methylation of APC, CDH1, p16 and MLH1, there were no significant differences between the cases and the controls regardless of adjustment. For MRS_10, compared with the subjects in the MRS-Low group (57.36% of participants), the subjects in MRS-High (11.80% of participants) and MRS-Medium groups (30.84% of participants) exhibited a 6.51-fold (95% CI, 3.77 to 11.27, P<0.0001) and 3.85-fold (95% CI, 2.72 to 5.45, P<0.0001) increased risk for CRC, respectively. Moreover, the CRC risk increased with increasing MRS_10 (P_trend < 0.0001).

Subgroup analysis

For MRS_10, the MRS_High and the MRS-Medium groups conferred an increased risk of CRC among both men and women, although the effect was attenuated in women compared with men (Supplementary Table 2). Additionally, significant associations were observed between the methylation of DAPK1, IGF2, NEUROG1 and WIF1 and CRC risk in both men and women. However, the methylation of MGMT and MINT31 only displayed statistically significant associations with CRC in men.

The methylation of DAPK1, IGF2, NEUROG1 and WIF1 was associated with CRC risk in both the younger (<60 years) and older groups (≥60 years) (Supplementary Table 3), whereas the associations between the methylation of MGMT and MINT31 and CRC risk were significant only in the younger participants. Clearly significant associations were observed between the MRS-High and the MRS_Medium groups and CRC risk among both the younger and the older participants.

According to body-mass index (BMI), the MRS_High and the MRS-Medium groups conferred a similarly increased risk of CRC in both normal weight or underweight subjects (<24) and overweight or obese subjects. For individual gene-specific methylation, the methylation of DAPK1, IGF2, NEUROG1 and WIF1 was associated with CRC risk in both of the normal weight or underweight group (<24) and the overweight or obese group (≥24) (Supplementary Table 4), whereas the associations between the methylation of CDH1, MGMT and MINT31 and CRC risk were significant only in the overweight or obesity participants.

Interactions between EF and DNA methylation

The significant interaction between increased intake of fruit and the methylation of IGF2 displayed an antagonistic effect on the risk of CRC. Alternatively, the significant interaction between increased consumption
of fish stewed with brown sauce and the methylation of CDH1 displayed a synergistic effect on the risk of CRC. Furthermore, the interactions between increased consumption of coarse grains and the methylation of IGF2 and between increased consumption of pork and the methylation of NEUROG1 demonstrated a marginally significant antagonistic effect on the risk of CRC (Table 3).

**Pyrosequencing verification**

For DAPK1 and MLH1, the methylation status obtained via MS-HRM was compared with the mean methylation level based on quantitative pyrosequencing and the results indicated that the MS-HRM results were well confirmed by the pyrosequencing results (Supplementary Figure 1). The Spearman correlation coefficients between these two methylation assessment techniques were high (r=0.7147, 95% CI, 0.6145 to 0.8077, P<0.0001 for DAPK1; r=0.6089, 95% CI, 0.3972 to 0.7392, P<0.0001 for MLH1), and the AUC was 0.8699 (95% CI, 0.7919 to 0.9479, P<0.0001) for DAPK1 and 0.9113 (95% CI, 0.8445 to 0.9782, P<0.0001) for MLH1. Additional results about the Bland-Altman plots are provided in the Supplementary Materials.

**Performance of the MRS_10 model**

We developed two basic models (including the MRS_10 model and the EF-only model) and a combined model (the combined EF and MRS_10 model). The AUC for the MRS_10 model was 69.04% (95% CI, 65.57 to 72.66%, P<0.0001), which represented significantly higher discrimination accuracy than any individual gene methylation (Figure 1). The AUC for the EF-only model was 72.94% (95% CI, 65.76 to 79.21%, P<0.0001). Notably, the combined EF and MRS_10 model yielded an AUC of 79.12% (95% CI, 76.22 to 82.15%, P<0.0001).

---

**Table 1: Main characteristics of colorectal cancer cases and controls**

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Number of cases (%)</th>
<th>Number of controls (%)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number</td>
<td>428</td>
<td>428</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>59.37 (10.30)</td>
<td>59.36 (10.35)</td>
<td>0.99</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>266 (62.15)</td>
<td>266 (62.15)</td>
<td>1.00</td>
</tr>
<tr>
<td>Female</td>
<td>162 (37.85)</td>
<td>162 (37.85)</td>
<td></td>
</tr>
<tr>
<td>BMI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;18.50</td>
<td>28 (6.54)</td>
<td>26 (6.07)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>18.5-24.00</td>
<td>219 (51.17)</td>
<td>146 (34.11)</td>
<td></td>
</tr>
<tr>
<td>24.0-28.00</td>
<td>150 (35.05)</td>
<td>139 (32.48)</td>
<td></td>
</tr>
<tr>
<td>≥28.00</td>
<td>31 (7.24)</td>
<td>117 (27.34)</td>
<td></td>
</tr>
<tr>
<td>Tumour site</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colon</td>
<td>177 (41.35)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rectum</td>
<td>251 (58.65)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pathological morphology</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protruding type</td>
<td>237 (55.37)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ulcerative type</td>
<td>135 (31.54)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other types</td>
<td>56 (13.09)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Degree of differentiation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>43 (10.05)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td>350 (81.78)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>34 (7.94)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>1 (0.23)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Histological classification</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adenocarcinoma</td>
<td>395 (92.29)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other types</td>
<td>33 (7.71)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dukes stage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A-B</td>
<td>231 (53.97)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C-D</td>
<td>157 (36.68)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>40 (9.35)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

BMI: body mass index; SD: standard deviation.
Table 2: Associations between methylation at individual genes, MRS_10 and the risk of CRC

<table>
<thead>
<tr>
<th>DNA methylation status</th>
<th>Cases (%)</th>
<th>Controls (%)</th>
<th>Crude OR</th>
<th>95% CI</th>
<th>P</th>
<th>OR&lt;sub&gt;adjusted&lt;/sub&gt;</th>
<th>95% CI</th>
<th>P</th>
<th>OR&lt;sub&gt;adjusted&lt;/sub&gt;</th>
<th>95% CI</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>APC</td>
<td>Negative</td>
<td>412 (96.26)</td>
<td>419 (97.90)</td>
<td>1.00</td>
<td></td>
<td>1.00</td>
<td>1.00</td>
<td></td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Positive</td>
<td>16 (3.74)</td>
<td>9 (2.10)</td>
<td>1.81</td>
<td>0.79-4.14</td>
<td>0.16</td>
<td>1.77</td>
<td>0.76-4.11</td>
<td>0.19</td>
<td>1.82</td>
<td>0.74-4.49</td>
</tr>
<tr>
<td>CDH1</td>
<td>Negative</td>
<td>405 (94.63)</td>
<td>392 (91.59)</td>
<td>1.00</td>
<td></td>
<td>1.00</td>
<td>1.00</td>
<td></td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Positive</td>
<td>23 (5.37)</td>
<td>36 (8.41)</td>
<td>0.62</td>
<td>0.36-1.06</td>
<td>0.08</td>
<td>0.59</td>
<td>0.34-1.02</td>
<td>0.06</td>
<td>0.65</td>
<td>0.36-1.19</td>
</tr>
<tr>
<td>CDKN2A</td>
<td>Negative</td>
<td>419 (97.90)</td>
<td>424 (99.07)</td>
<td>1.00</td>
<td></td>
<td>1.00</td>
<td>1.00</td>
<td></td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Positive</td>
<td>9 (2.10)</td>
<td>4 (0.93)</td>
<td>2.28</td>
<td>0.70-7.45</td>
<td>0.17</td>
<td>2.26</td>
<td>0.68-7.51</td>
<td>0.18</td>
<td>1.99</td>
<td>0.56-7.04</td>
</tr>
<tr>
<td>DAPK1</td>
<td>Negative</td>
<td>322 (75.23)</td>
<td>385 (89.95)</td>
<td>1.00</td>
<td></td>
<td>1.00</td>
<td>1.00</td>
<td></td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Positive</td>
<td>106 (24.77)</td>
<td>43 (10.05)</td>
<td>2.95</td>
<td>2.01-4.33</td>
<td>&lt;0.0001</td>
<td>2.93</td>
<td>1.98-4.33</td>
<td>&lt;0.0001</td>
<td>2.95</td>
<td>1.94-4.49</td>
</tr>
<tr>
<td>IGF2</td>
<td>Negative</td>
<td>334 (78.04)</td>
<td>389 (90.89)</td>
<td>1.00</td>
<td></td>
<td>1.00</td>
<td>1.00</td>
<td></td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Positive</td>
<td>94 (21.96)</td>
<td>39 (8.41)</td>
<td>2.81</td>
<td>1.88-4.19</td>
<td>&lt;0.0001</td>
<td>2.63</td>
<td>1.75-3.96</td>
<td>&lt;0.0001</td>
<td>2.54</td>
<td>1.65-3.92</td>
</tr>
<tr>
<td>MGMT</td>
<td>Negative</td>
<td>365 (85.28)</td>
<td>389 (90.89)</td>
<td>1.00</td>
<td></td>
<td>1.00</td>
<td>1.00</td>
<td></td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Positive</td>
<td>63 (14.72)</td>
<td>39 (9.11)</td>
<td>1.72</td>
<td>1.13-2.63</td>
<td>0.01</td>
<td>1.48</td>
<td>0.96-2.27</td>
<td>0.08</td>
<td>1.82</td>
<td>1.00-2.60</td>
</tr>
<tr>
<td>MINT31</td>
<td>Negative</td>
<td>409 (95.56)</td>
<td>423 (98.83)</td>
<td>1.00</td>
<td></td>
<td>1.00</td>
<td>1.00</td>
<td></td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Positive</td>
<td>19 (4.44)</td>
<td>5 (1.17)</td>
<td>3.93</td>
<td>1.45-10.62</td>
<td>0.01</td>
<td>4.02</td>
<td>1.47-10.99</td>
<td>0.01</td>
<td>4.27</td>
<td>1.52-12.05</td>
</tr>
<tr>
<td>MLH1</td>
<td>Negative</td>
<td>412 (96.26)</td>
<td>420 (98.13)</td>
<td>1.00</td>
<td></td>
<td>1.00</td>
<td>1.00</td>
<td></td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Positive</td>
<td>16 (3.74)</td>
<td>8 (1.87)</td>
<td>2.04</td>
<td>0.86-4.82</td>
<td>0.11</td>
<td>1.94</td>
<td>0.81-4.65</td>
<td>0.14</td>
<td>1.72</td>
<td>0.68-4.34</td>
</tr>
<tr>
<td>NEUROG1</td>
<td>Negative</td>
<td>363 (84.81)</td>
<td>401 (93.69)</td>
<td>1.00</td>
<td></td>
<td>1.00</td>
<td>1.00</td>
<td></td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Positive</td>
<td>65 (15.19)</td>
<td>27 (6.31)</td>
<td>2.66</td>
<td>1.66-4.26</td>
<td>&lt;0.0001</td>
<td>2.61</td>
<td>1.62-4.21</td>
<td>&lt;0.0001</td>
<td>2.57</td>
<td>1.55-4.25</td>
</tr>
<tr>
<td>WIFI</td>
<td>Negative</td>
<td>354 (82.71)</td>
<td>392 (91.59)</td>
<td>1.00</td>
<td></td>
<td>1.00</td>
<td>1.00</td>
<td></td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Positive</td>
<td>74 (17.29)</td>
<td>36 (8.41)</td>
<td>2.28</td>
<td>1.49-3.48</td>
<td>&lt;0.0001</td>
<td>2.26</td>
<td>1.47-3.48</td>
<td>&lt;0.0001</td>
<td>2.44</td>
<td>1.53-3.87</td>
</tr>
<tr>
<td>MRS_10</td>
<td>Low</td>
<td>172 (40.18)</td>
<td>319 (74.53)</td>
<td>1.00</td>
<td></td>
<td>1.00</td>
<td>1.00</td>
<td></td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>177 (41.36)</td>
<td>87 (20.33)</td>
<td>3.77</td>
<td>2.75-5.18</td>
<td>&lt;0.0001</td>
<td>3.66</td>
<td>2.66-5.05</td>
<td>&lt;0.0001</td>
<td>3.85</td>
<td>2.72-5.45</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>79 (18.46)</td>
<td>22 (5.14)</td>
<td>6.67</td>
<td>4.01-11.06</td>
<td>&lt;0.0001</td>
<td>6.41</td>
<td>3.84-10.71</td>
<td>&lt;0.0001</td>
<td>6.51</td>
<td>3.77-11.27</td>
</tr>
<tr>
<td></td>
<td>Medium or High</td>
<td>256 (59.81)</td>
<td>109 (25.47)</td>
<td>4.36</td>
<td>3.26-5.83</td>
<td>&lt;0.0001</td>
<td>4.22</td>
<td>3.14-5.66</td>
<td>&lt;0.0001</td>
<td>4.39</td>
<td>3.19-6.05</td>
</tr>
</tbody>
</table>

CI: confidence interval; CRC: colorectal cancer; MRS: methylation risk score; OR: odds ratio.

* According to the Youden index, a 1% level of methylation was used as the cut-off value for APC, CDH1 and IGF2, and a 0% level of methylation was used as the cut-off value for CDKN2A, DAPK1, MGMT, MINT31, MLH1, NEUROG1 and WIFI. The positive indicates methylated status and the negative indicates unmethylated status.  

* The rates represent the percentages of all cases and controls, respectively.  

* ORs adjusted for age, gender and BMI.  

* ORs adjusted for age, gender, BMI, occupational physical activity, smoking, and consumption of coarse grains, fish stewed with brown sauce, fried food, leftovers and pork.  

* P values < 0.05 are in bold.
representing an increase of 6.18% (95% CI, 3.99 to 8.54%, \( P < 0.0001 \)) compared to the EF-only model. Based on the NRI, the IDI and the AUC difference, the improvement of adding MRS_10 to the EF-only model was statistically significant (Supplementary Table 5). The NRI and the IDI for addition of the MRS_10 to the EF-only model was 14.72% (95% CI, 8.92 to 20.52%, \( P < 0.0001 \)) and 9.30% (95% CI, 7.36 to 11.23%, \( P < 0.0001 \)), respectively.

### DISCUSSION

This study found that MRS_10 were strongly associated with the risk of CRC, indicating a positive relationship between leukocyte-derived DNA methylation of the panel of 10 genes and CRC risk. By using of MRS_10, compared with the MRS-Low group, subjects with MRS-High and MRS-Medium showed 6.51-fold and 3.85-fold increased risk of CRC, respectively. Stratified analyses demonstrated that these significant associations retained in both men and women, younger and older, and normal weight or underweight and overweight or obese subjects.

The findings demonstrated that the DNA methylation pattern in peripheral blood leukocytes is a detectable biomarker for CRC risk assessment. This observation is similar to those results for other cancers reported in previous studies [31–34] suggesting that DNA methylation alterations in peripheral blood are potential biomarkers for risk prediction.

The gene involved in the study illustrated a wide range of cancer related cellular events, for example, CDH1 encoded a member of the family of cell adhesion, and abrogation in its expression have been involved in unregulated growth and invasion of adjacent tissues in

### Table 3: Effects of interactions between environmental factors and gene methylation on the risk of colorectal cancer

<table>
<thead>
<tr>
<th>Gene methylation</th>
<th>Environmental factor</th>
<th>Gene methylation</th>
<th>Environmental factor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Consumption of fruits (times/week)</td>
<td></td>
<td>Consumption of coarse grains (g/week)</td>
</tr>
<tr>
<td></td>
<td>&lt;2</td>
<td>≥2</td>
<td>Interaction</td>
</tr>
<tr>
<td>IGF2</td>
<td>OR(_{eg}) (95% CI)</td>
<td>OR(_{i}) (95% CI)</td>
<td>( P )</td>
</tr>
<tr>
<td>Negative</td>
<td>1</td>
<td>0.91(0.74-1.11)</td>
<td>Negative</td>
</tr>
<tr>
<td>Positive</td>
<td>3.72(2.08-6.66)</td>
<td>1.64(1.00-2.71)</td>
<td>0.35(0.14-0.88)</td>
</tr>
<tr>
<td>CDH1</td>
<td>OR(_{eg}) (95% CI)</td>
<td>OR(_{i}) (95% CI)</td>
<td>( P )</td>
</tr>
<tr>
<td>Negative</td>
<td>1</td>
<td>1.50(1.15-1.95)</td>
<td>Negative</td>
</tr>
<tr>
<td>Positive</td>
<td>0.31(0.15-0.66)</td>
<td>2.00(0.81-4.96)</td>
<td>3.82(1.11-13.15)</td>
</tr>
</tbody>
</table>

CI: confidence interval; OR\(_{eg}\): odds ratio for the combined effect of the gene methylation pattern and the environmental factor; OR\(_{i}\): odds ratio for the interaction between the gene methylation pattern and the environmental factor.

* The OR\(_{i}\) was adjusted for age, gender and BMI.

\( P \) values \(< 0.05\) are in bold.
Figure 1: Receiver operating characteristic (ROC) curves and the corresponding area under the curves (AUC) analyses of prediction models of CRC risk. (A) The MRS_10 model versus any single gene methylation pattern. (B) Comparisons of the prediction accuracy by the EF-only model, the MRS_10 model and the combined EF and MRS_10 model. The straight dotted line (reference line), corresponding to an AUC of 0.5, indicates that the model is no more accurate than random classification in predicting CRC risk.

(Continued)
carcinogenesis [35]. Likewise, as a tumor suppressor gene, \textit{CDKN2A} played an important control role in cell cycle regulating during the G1 phase [36]. \textit{DAPK1} was a positive mediator of apoptosis which executed juncture of cell death signaling, and its loss or inactivation has been linked to human tumor [37]. \textit{WIF1} was a secreted inhibitory factor of Wnt pathway, which played a pivotal role in blockade of Wnt signaling and induced apoptosis in colorectal cancer cells [10]. A growing body of evidence suggested that aberrant hypermethylation of the promoter can lead to reduced expression of these genes including \textit{APC} [9], \textit{CDH1} [38], \textit{CDKN2A} [39], \textit{DAPK1} [40], \textit{IGF2} [41], \textit{MGMT} [42], \textit{MLH1} [43], \textit{NEUROG1} [44] and \textit{WIF1} [45]. In contrast to the methylation status of a single gene, the MRS, which summarized the data for the methylation of multiple genes, might demonstrate systematically altered gene methylation profiles in subjects and might more comprehensively represent the susceptibility of an individual to CRC. In addition, sensitivity analyses by omitting each individual locus showed a robust association between the MRS and the risk of CRC (Supplementary Table 7).

The exact mechanism underlying the alterations in the methylation of peripheral blood-derived DNA among individuals who are susceptible to CRC remains unclear. Alterations in leukocyte-derived DNA methylation may reflect a response of the hematopoietic system to tumorlogenesis and may be partially explained by systemic differences in the methylation signatures of leukocyte subpopulations during tumorigenesis [46]. The leukocyte-derived DNA methylation profiles represented the overall methylation status; nevertheless, isolation of specific cell subpopulations is difficult in epidemiologic studies [47]. Additionally, in the present case-control study, it is not possible to determine the etiologically relevant time windows of DNA methylation or confirm the temporal sequence of DNA methylation and CRC occurrence. Although we cannot definitely determine whether these methylation alterations in peripheral blood-derived DNA represent an early response of the hematologic system to the presence of tumor cells or appear before tumor development [31], the observed significant association between gene-specific DNA methylation in leukocytes and CRC risk is likely to be useful in identifying the population at high risk of CRC.

Importantly, circulating tumor cell DNA is unlikely to interfere with the results of leukocyte-derived DNA methylation because the number of circulating tumor cells is negligible and the effects of their aberrant methylation can be excluded [32, 48]. Another potential source of support of our findings is that accumulating studies have shown no significant correlation between tissue and blood DNA methylation [49, 50]. We also compared the DNA methylation levels of six genes in peripheral-blood

---

\textbf{Figure 1 (Continued):}(C) The combined EF and MRS_10 model versus the EF-only model. ROC curves and AUC analyses were computed for a reference prediction model (blue) and for an extended model (red) including additional risk predictors. Indicated are the increases in AUC (delta-AUC) obtained by adding the additional predictors.
leukocytes and colorectal tumor tissues from 217 CRC patients; no statistically significant correlations were observed (Supplementary Table 8).

The potential biological mechanism for EF interfering with DNA methylation processes in carcinogenesis is complex and variable. Several researchers have suggested the plausible mechanisms, including that DNA methylation requires methyl group donors such as S-adenosylmethionine (SAM), while environmental factors may affect SAM synthesis and alter DNA methyltransferase (DNMT) activity [51]. In this study, we found that smoking was associated with DNA hypomethylation, which was consistent with several previous studies [52, 53]. Smoking can induce DNMT1 overexpression and subsequently result in hypermethylation of the promoter of tumor suppressor genes, which could lead to tumorigenesis [54]. The molecular mechanism underlying the link between consumption of pork or stewed fish with brown sauce and DNA methylation remains unclear. Saturated fatty acids and heterocyclic amines, which form during the high-temperature cooking of pork and fish, may partly account for the aberrant hypermethylation [55, 56]. Alteration of DNA methylation is a gradual and reversible process, thus, there is a critical window of opportunity through which we might inhibit or reverse the process and counteract cancer by making changes in diet and lifestyle [57].

We initially used MS-HRM to assess DNA methylation status. MS-HRM technology has been shown to reliably and accurately evaluate low-level methylation [58]. By comparing the estimates of methylation of DAPK1 and MLH1 based on MS-HRM and pyrosequencing, we found that the MS-HRM results were highly consistent with the pyrosequencing results. This consistency was also supported by a previous study that analyzed APC and CDKN2A methylation and found a high correlation between the results from these two techniques [59]. Only a small subset of genes and participants were analyzed via pyrosequencing because of its high cost. From a cost perspective, the assessment of peripheral blood-derived DNA methylation via MS-HRM might be an efficient strategy for the early detection of individuals who are at high risk for CRC.

In this study, we developed the weighted MRS_10 that combined these 10 genes in a leukocyte-derived DNA methylation marker panel and found that the AUCs for the MRS_10 model and the combined EF and MRS_10 model reached 0.6904 and 0.7912, respectively. Compared to the EF-only model with a similar AUC to those reported in previous studies (AUCs ranging from 0.61 to 0.76) [60–69], adding the MRS_10 significantly improved the discriminatory performance, as demonstrated by a 0.0618 units of AUC gain. However, the combined EF and MRS model is needed to be further validated in future prospective studies. To evaluate whether adding the MRS_10 to the EF-only model improved risk prediction performance, we assessed the accuracy of the combined model in classifying individuals as cases or controls based on the NRI and the IDI, which are useful statistics that have gained increasing acceptance for the evaluation of new biomarkers and risk models [70, 71]. Adding MRS_10 to the EF-only model resulted in the reclassification of 14.72% of the subjects into more accurate risk categories. This advancement might improve the selection of those who require more frequent screening and shorter follow-up intervals. However, NRI is sensitive to arbitrary cut-off values [72]. Therefore, we reset the cut-off points to 0.3, 0.4, 0.6 or 0.7 and found similar NRIs (Supplementary Table 9), which implied that the discriminatory improvement was robust across the cut-off values used in our study.

Our study has several additional limitations. First, the current panel of gene methylation sites incorporated into the calculation of MRS_10 might not be ideal because we have identified the genes from published studies rather than next-generation sequencing, which would likely reveal additional methylation biomarkers. Therefore, the present MRS_10 models in this study must be updated. Another potential limitation of this study is the restriction of EFs included in the combined EF and MRS_10 model. The age and gender matching design inherent to our present study removes two CRC-associated variables. Additionally, the discriminatory performance of the combined model might be improved by including additional factors such as colonoscopy or sigmoidoscopy findings.

In summary, the MRS_10 combining a panel of 10 gene-specific DNA methylation in leukocytes seemed to be a promising and robust risk prediction tool and might be a useful marker of susceptibility for CRC. The MRS_10 may be useful for the identification of individuals who are at high risk of developing CRC. However, our results should be further validated in future studies.

MATERIALS AND METHODS

Study population

We included primary sporadic CRC cases diagnosed at the Third Affiliated Hospital (from June 2004 to May 2005 and May 2007 to January 2008) and the Second Affiliated Hospital of Harbin Medical University (from October 2010 to December 2011) in Harbin, China. Cancer-free controls were selected contemporaneously from the Second Affiliated Hospital of Harbin Medical University by individual matching each case according to gender and age (±2 years). All participants were Chinese. All CRC cases were newly diagnosed, histologically confirmed, and alive at the time of initial contact; and the exclusion criteria included subjects with metastatic colorectal carcinoma, adenomatous polyposis coli, or a family history of CRC in first-degree relatives according
to the Amsterdam criteria [73]. The participants were interviewed face-to-face to complete a structured standard questionnaire, which was partially adopted from the report by Shu et al [74]. The questionnaire queried information on demographic characteristics and potential risk factors for CRC, including family history, smoking, alcohol drinking, occupational physical activity, and diet consumption. Dietary consumption over the past year was assessed using a validated food frequency questionnaire (FFQ) [74]. The FFQ included 9 major food groups, which represent most of the common foods in Northeast China. The food items included coarse grains, dairy products, fish stewed with brown sauce, fried food, fresh fruits, green vegetables, leftover, pork, and soybean products.

Through professional training, the investigators made efforts to reduce recall and investigation bias. All study subjects gave informed consent, and appropriate ethical approval for sample collection was obtained from the Ethics Committee of Harbin Medical University prior to the study. Peripheral blood (5 milliliters) was donated before chemotherapy or adjuvant radiotherapy and was stored in a freezer. The sample size estimation was presented in the Supplementary Materials. In total, 553 eligible patients were recruited; 32 did not complete the questionnaire, and another 19 did not provide sufficient blood samples. Therefore, 502 extracted DNA samples were available for the CRC cases, with a response rate of 90.78%. Similarly, 1,210 subjects were eligible as controls, and 1,083 provided complete questionnaire data and blood samples with a response rate of 89.50%. Because eligible matched controls could not be found for 45 cases, 457 controls were ultimately enrolled.

Genomic DNA extraction and bisulfite modification

DNA was extracted from buffy coats using the QIAamp DNA Blood Mini Kit (Qiagen, Hilden, Germany) and was then bisulfite-modified using the EpiTect Plus DNA Bisulfite Kit (Qiagen) according to the manufacturer’s protocols. Detailed methods are provided in the Supplementary Materials.

Methylation-sensitive high-resolution melting (MS-HRM) assays

Fourteen CRC-related genes, including tumor suppressor genes such as APC, CDH1, CDKN2A (also known as p16), CDKN2B (also known as p15), DAPK1, GSTP1, MGMT, MLH1, PTEN and WIF1 and CpG island methylator phenotype-related markers such as IGF2, APBA1 (also known as MINT1), MINT31 and NEUROG1, were selected according to review of literatures. Eight primer pairs were selected from previously published studies [75–81]; the remaining primers were designed using Methprimer software [82]. The primer sequences used are listed in Supplementary Table 10. The PCR mixture consisted of a total volume of 10 μl containing 2× LightCycler 480 High Resolution Melting Master Mix (Roche Applied Science, Mannheim, Germany), 3 mM MgCl2, 0.2-0.4 μM of each primer and approximately 10 ng of bisulfite-modified template DNA. PCR amplification and MS-HRM analyses were performed using the LightCycler 480 platform (Roche), and the resulting data were analyzed using software module of Gene Scanning (Roche). A set of methylation standards (100, 25, 10, 5, 2, 1, and 0% methylated DNA) were prepared by mixing commercially available methylated and unmethylated DNA (Zymo Research); these standards were used to semi-quantitatively measure the DNA methylation level in the samples. Normalized melting curves of MS-HRM assays for each gene were shown in Supplementary Figure 2. In addition, a blank control (non-template control) sample was included in each batch, and all reactions were performed in duplicate. A third trial was conducted for the samples that presented inconsistent results between the two trials.

Two investigators (Y.L. and Y.W.) blinded to outcome and other predictive variables assessed the MS-HRM results, and discrepancies were resolved by discussion and consensus with another investigator (Y.Z.). The DNA methylation pattern was first tested in 167 cases and their matched controls. Four genes displaying no abnormal methylation (PTEN, GSTP1, APBA1 and CDKN2B) were tested no further and were excluded from the analysis. Finally, the other 10 genes were assessed in all 457 cases and their matched controls.

Pyrosequencing verification

To verify the results of MS-HRM, we performed pyrosequencing in a subset of the samples (further details are provided in the Supplementary Materials and the primer sets are listed in Supplementary Table 11).

Methylation risk score (MRS) computation

First, we coded each gene methylation status as 0 or 1 for non-methylation or methylation, respectively, according to the optimal cut-off value as determined by the Youden index (J = max {sensitivity + specificity – 1}) using the receiver operating characteristic (ROC) curve analysis. Then, we computed the weighted MRS for each individual by multiplying the methylation status by the β-coefficients for each gene included in the model and then dividing by the number of genes included in the model. Simply, the weighted MRS was computed using the following equation:

\[
MRS = \frac{\sum i \beta_i x_i}{k}
\]

Where \( \beta_i \) is the β-coefficient for gene \( i \), \( x_i \) is the methylation status of the same gene \( i \) (0 or 1), and \( k \) is the
total number of genes included in the model. According to the formulation listed above, we generated MRS_10 comprehensively considering the methylation status of all the 10 genes. Using the MRS_10 model, the subjects were classified as follows: MRS-Low, predicted probability ≤0.5; MRS-Medium, 0.5< predicted probability ≤0.7; or MRS-High, predicted probability >0.7.

Missing data analysis and imputation

Definitions of questionnaire-derived variables are provided in the Supplementary Materials. All questionnaire-derived variables were analyzed via missing value analysis and were imputed via multiple imputation (further details are provided in the Supplementary Materials).

Model development and performance

We selected CRC-associated risk factors by using of multivariable logistic regression model with backward conditional selection method ($P$ values of 0.05 and 0.10 were specified as the thresholds for entry and removal of variables, respectively), and we further included gender and age in the final model. Interactions between environmental factors (EF) and gene methylation were also explored, but we did not included interaction terms in the models because interaction terms rarely add to the predictive ability of the model [83, 84]. A multivariable logistic regression model was used to fit the CRC risk prediction models: Model 1 was the MRS_10 model; Model 2 was an EF-only model, which exclusively contained questionnaire-derived risk factors; and Model 3 was a combined model incorporating EF and MRS_10. For each logistic regression model, the regression coefficient estimates were corrected according to the pooled imputed datasets. The estimates of regression coefficients for each model were listed in Supplementary Table 12.

To assess the discriminatory accuracy of each model, we generated ROC curves and calculated the areas under ROC curves (AUCs). The 95% CI for the AUC were estimated using the bootstrapping method (1000 replicates). We used the method of DeLong [85] to assess the differences in the AUC between the models. We further calculated the categorical net reclassification improvement (NRI), the integrated discrimination improvement (IDI) to evaluate the improvement in the discriminatory accuracy of the model (considering 0.5 as the cut-off point).

Statistical analysis

The differences in characteristics between the cases and the controls were tested using Student’s $t$-test for continuous variables or the $\chi^2$ test for categorical variables. We used both univariable and multivariable logistic regression models to estimate odds ratios (ORs) and 95% confidence interval (CI) for the associations between gene-specific DNA methylation or EF and CRC risk. The effects of interactions between EF and the methylation status at individual genes or MRS_10 on the risk of CRC were evaluated on a multiplicative scale with a product-term coefficient using multivariable logistic regression models. Additionally, to assess the associations between EF and DNA methylation among all the subjects, we used multivariable logistic regression models that included age, gender, BMI and case-control status that allowed us to adjust for the potential influence from case-control status. We also assessed the associations by case and control separately. All statistical tests were two-sided except for the test of the increase in AUC (one-sided) and statistical significance was defined as $P$ values of less than 0.05. Missing value analysis, multiple imputations and multivariable logistic regression analysis were performed using SPSS Statistics version 19.0 (IBM, Inc., USA). The 95% CIs for the AUC and the NRI and the IDI were estimated using the pROC package and the PredictABEL package in R software version 2.15.3, respectively. All other statistical analyses were performed using SAS software version 9.1 (SAS Institute Inc., USA).

Author contributions


ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (grant numbers 81473055, 30972539).

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

REFERENCES


