



Novel epigenetic markers for gastric cancer risk stratification in individuals after *Helicobacter pylori* eradication

Masahiro Maeda^{1,2} · Satoshi Yamashita¹ · Taichi Shimazu³ · Naoko Iida¹ · Hideyuki Takeshima¹ · Takeshi Nakajima⁴ · Ichiro Oda⁴ · Sohachi Nanjo⁵ · Chika Kusano⁶ · Akiko Mori¹ · Hiroshi Moro¹ · Harumi Yamada^{1,2} · Shoichiro Tsugane³ · Toshiro Sugiyama⁵ · Yoshiharu Sakai² · Toshikazu Ushijima¹

Received: 19 October 2017 / Accepted: 27 January 2018 / Published online: 9 February 2018
© The International Gastric Cancer Association and The Japanese Gastric Cancer Association 2018

Abstract

Background The risk stratification of healthy individuals after *Helicobacter pylori* eradication is an urgent issue. The assessment of aberrant DNA methylation accumulated in gastric tissues with normal appearance, which can reflect overall epigenomic damage, is a promising strategy. We aimed to establish novel epigenetic cancer risk markers for *H. pylori*-eradicated individuals.

Methods Gastric mucosa was collected from eight healthy volunteers without *H. pylori* infection (G1), 75 healthy individuals with gastric atrophy (G2), and 94 gastric cancer patients (G3) after *H. pylori* eradication. Genome-wide analysis was conducted using Infinium 450 K and differentially methylated probes were screened using large difference and iEVORA-based methods. Bisulfite pyrosequencing was used for validation.

Results Screening, using 8 G1, 12 G2, and 12 G3 samples, isolated 57 candidates unmethylated in G1 and differentially methylated in G3 compared with G2. Validation for nine candidate markers (*FLT3*, *LINC00643*, *RPRM*, *JAM2*, *ELMO1*, *BHLHE22*, *RIMS1*, *GUSBP5*, and *ZNF3*) in 63 G2 and 82 G3 samples showed that all of them had significantly higher methylation levels in G3 than in G2 ($P < 0.0001$). Their methylation levels were highly correlated, which indicated that they reflect overall epigenomic damage. The candidates had sufficient performance (AUC: 0.70–0.80) and high odds ratios (5.43–23.41), some of which were superior to a previous marker, *miR-124a-3*. The methylation levels of our novel markers were not associated with gastric atrophy, gender, or age.

Conclusions Novel epigenetic markers for gastric cancer risk optimized for *H. pylori*-eradicated individuals were established.

Keywords Epigenetics · Cancer risk marker · DNA methylation · Gastric cancer · *Helicobacter pylori*

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10120-018-0803-4>) contains supplementary material, which is available to authorized users.

✉ Masahiro Maeda
tushijim@ncc.go.jp

¹ Division of Epigenomics, National Cancer Center Research Institute, Tokyo, Japan

² Department of Gastrointestinal Surgery, Kyoto University Graduate School of Medicine, Kyoto, Japan

³ Epidemiology and Prevention Group, Center for Public Health Sciences, National Cancer Center, Tokyo, Japan

Introduction

Risk stratification is critically important for the early detection of gastric cancer, which can lead to improved curability and reduced mortality. In Japan, national health insurance approved the eradication therapy of *Helicobacter pylori* (*H. pylori*) for the indication of chronic gastritis to prevent gastric

⁴ Endoscopy Division, National Cancer Center Hospital, Tokyo, Japan

⁵ Third Department of Internal Medicine, University of Toyama, Toyama, Japan

⁶ Division of Gastroenterology and Hepatology, Department of Medicine, Nihon University School of Medicine, Tokyo, Japan

cancers in February 2013, based on clinical evidence [1]. This has led to an explosive increase in healthy individuals after *H. pylori* eradication [2]. However, *H. pylori* eradication does not completely inhibit gastric cancer development [1]. Therefore, periodic surveillance for all such people is currently recommended and has emerged as a burden on both the individual and medical service providers. Risk stratification can optimize the surveillance interval for individuals, which helps to manage the social burden. Therefore, the establishment of a risk stratification system for *H. pylori*-eradicated healthy individuals is an urgent issue.

To this end, epigenetic risk markers have been highlighted as a promising biomarker for cancer risk stratification. Epigenetic alterations, namely aberrant DNA methylation, are induced in gastric mucosa with normal appearance by *H. pylori* infection-triggered chronic inflammation [3–5]. After *H. pylori* eradication, methylation levels decrease, depending upon individual marker genes, and the decreased levels persist for a long time [6, 7]. Persistent methylation levels after eradication are considered to reflect methylation in stem cells [8] and are closely correlated with gastric cancer risk [9, 10]. Methylation markers are considered to be a reflection of the overall epigenomic damage in gastric mucosa that leads to gastric cancer [11]. Using previously isolated methylation markers, such as *miR-124a-3* [12, 13], we conducted a multicenter prospective cohort study for the prediction of the risk of metachronous gastric cancers and found the advocated epigenetic cancer risk diagnosis as a promising strategy [14, 15].

However, our previous markers were isolated from relatively small sample sets and by methylated DNA immunoprecipitation (MeDIP)-CpG island microarray, which is a low-resolution technique [13]. The recent development of a bisulfite-based BeadArray has enabled high-throughput epigenetic analysis with a much more comprehensive coverage of the genome. In particular, the BeadArray reaches a higher resolution using more than 450 thousand probes; additionally, the BeadArray can accurately measure DNA methylation levels through the accurate detection of slight differences. Therefore, using this new technology, it is likely that we can establish novel methylation markers with a higher accuracy than that of the previous markers.

In this study, we aimed to establish novel methylation markers for the risk stratification of gastric cancer optimized for *H. pylori*-eradicated individuals using the BeadArray technology.

Materials and methods

Tissue sample collection

A total of 232 normal or non-cancerous gastric mucosa samples collected from 169 subjects in two previous

studies [14, 16]. In addition, eight normal gastric mucosa samples were collected from healthy volunteers without a history of *H. pylori* infection and three peripheral blood samples from healthy volunteers for screening. Further, four gastric mucosa samples from healthy volunteers with current *H. pylori* infection and four without a history of *H. pylori* infection were collected for expression microarray analysis. All gastric samples were endoscopically biopsied from the fixed antral region (2 cm from the pyloric ring on the lesser curvature), and stored in RNAlater (Thermo Fisher Scientific, MA, USA) at -80°C . For *H. pylori*-eradicated individuals, the samples were collected 6 months or more after eradication. Genomic DNA was extracted using the phenol/chloroform method.

All participants were classified into three groups (Groups 1, 2, and 3) according to their cancer risk (low, intermediate, and high, respectively) defined by the history of *H. pylori* infection and gastric cancer status (Supplementary Table 1). Group 1 (G1) consisted of four young (mean age \pm SD, 28.8 ± 4.2 years) and four older (mean age \pm SD, 71.0 ± 3.1 years) healthy volunteers without a history of *H. pylori* infection, and was considered a low-risk group. Group 2 (G2) consisted of 75 healthy individuals who underwent cancer screening at the Research Center for Cancer Prevention and Screening at the National Cancer Center, Japan, between April 2009 and September 2009 and experienced the successful eradication of *H. pylori* [16]. All the G2 participants were accompanied by gastric mucosal atrophy, and G2 was considered an intermediate-risk group. G2 samples were consecutively collected both before and after *H. pylori* eradication, and all the samples had sufficient quality and quantity for use in this study. Group 3 (G3) consisted of 94 gastric cancer patients who underwent the first endoscopic submucosal dissection at the National Cancer Center Hospital [14]. Such gastric cancer patients are known to have a much higher incidence of subsequent (metachronous) gastric cancer even after *H. pylori* eradication [17], reaching 3.0% per year, while *H. pylori*-eradicated healthy individuals (G2) have a low incidence of 0.23% per year [18]. All G3 patients had a history of *H. pylori* infection accompanied by gastric mucosal atrophy.

Gastric mucosal atrophy was determined before successful *H. pylori* eradication in accordance with the Kimura–Takemoto classification [19], which was correlated with the degree of histological atrophy. The extent of gastric mucosal atrophy was classified into three grades: mild (closed types I and II), moderate (closed type III and open type I), and severe (open types II and III).

All previous studies and the current study were approved by the relevant Institutional Review Boards, and written informed consent was obtained from all participants.

Preparation of a screening set and a validation set

To prepare a screening set, 12 samples were randomly selected from G2 and G3, respectively (Supplementary Table 2). In the screening set, the age and extent of gastric atrophy were not significantly different between G2 and G3 (Supplementary Table 1). In addition, G1 samples (Supplementary Table 1) were used for the selection of unmethylated probes in normal gastric mucosa. The remaining participants were used as the validation set (Supplementary Table 3). In the validation set, age distribution was comparable between G2 and G3, but G3 included significantly more males and individuals with a more severe extent of gastric atrophy than G2 (Supplementary Table 1).

Genome-wide DNA methylation analysis

A comprehensive genome-wide screening of differentially methylated CpG sites was conducted using an Infinium HumanMethylation450 BeadChip array (Illumina, CA, USA) as described previously [20]. Among the total 485,512 probes, the probes on the sex chromosomes were excluded and the remaining 473,864 probes on autosomes were analyzed. The methylation level of each probe was represented by a β value, which was in the range from 0 (completely unmethylated) to 1 (completely methylated). To adjust for probe design biases, the intra-array normalization was conducted using a peak-based correction method, beta-mixture quantile dilation (BMIQ) [21].

iEVORA-based screening

The Epigenetic Variable Outliers for Risk prediction Algorithm (iEVORA) is a novel statistical algorithm based on the hypothesis that differentially variable (DV) and methylated (DM) CpGs (DVMCs) are more likely to indicate cancer risk [22, 23]. In the iEVORA algorithm, significant DV CpGs are selected by Bartlett's test FDRs and the significant DVMCs were ranked by DM t statistic. The iEVORA algorithm was computed by the R script 'iEVORA.R' and probes with a Bartlett's test FDR of less than 0.001 and an unadjusted P value of less than 0.05, based on the t test, were selected.

Quantitative methylation analysis

DNA (1 μ g) was modified by sodium bisulfite using an innuCONVERT Bisulfite Basic kit (Analytik Jena AG, Germany) and EZ DNA Methylation Kit (Zymo Research, CA, USA). The modified DNA was suspended in 40 μ L elution buffer and a 1 μ L aliquot was used for quantitative methylation analysis. Quantitative methylation-specific polymerase chain reaction (qMSP) was performed using an iCycler Thermal Cycler (Bio-Rad Laboratories, CA, USA),

as previously described [12, 13]. The primer sets specific to methylated and unmethylated sequences and PCR conditions are listed in Supplementary Table 4. Standard DNAs for methylated and unmethylated sequences were prepared by cloning the PCR products from fully methylated and unmethylated control DNAs, respectively, into pGEM-T Easy vector (Promega, WI, USA). The number of molecules in a sample was determined through the comparison of its amplification with those of standard DNA with defined numbers of molecules (10^1 – 10^6 molecules). The methylation levels were defined as the fraction of methylated molecules in the total number of DNA molecules (the number of methylated and unmethylated molecules).

Bisulfite pyrosequencing was performed on bisulfite-converted DNA using the PSQ 96 Pyrosequencing System (Qiagen, CA, USA) as previously described [24]. Data analysis was conducted using PyroMark Q96 ID software (Qiagen, version 2.5.8). The primer sequences and PCR conditions are listed in Supplementary Table 4.

Gene expression analysis by microarray

Expression analysis was conducted with a GeneChip Human Genome U133 Plus 2.0 expression microarray (Affymetrix, CA, USA), using a pool of 4 samples from healthy individuals with current *H. pylori* infection, and another pool of 4 samples from individuals without a history of *H. pylori* infection. Data processing was conducted using GeneChip operating software (ver. 1.4). The signal intensity of each probe was normalized so that the average signal intensity of all the probes on a microarray would be 500. Mean signal intensity of all probes for a gene was used as the gene expression level. Genes were classified into those with high (> 1000), moderate (250–1000), and low (< 250) transcription, according to their signal intensities as previously described [25].

Statistical analysis

The mean methylation levels were compared by Welch's t test and by a paired t test to compare G2 samples before and after eradication. The clinical factors between G2 and G3 were compared by Welch's t test or the Chi-squared test. These analyses were computed using GraphPad Prism software (GraphPad Software, CA, USA). Pearson's correlation coefficients and the P values were calculated by Excel software. The receiver-operating characteristic (ROC) curves and AUCs (area under the curve) were computed using the R package, ROCR. The P values were obtained by a two-sided test and considered as significant if less than 0.05.

Table 1 Candidates for epigenetic markers

No.	IllumID	Human genome (Hg19)		UCSC_RefGene_Name	UCSC_RefGene_Group	Relation_to_UCSC_CpG_Island				
		CHR	MAPINFO							
(A) The large difference analysis										
Top 10 (TT;P value)										
1	cg05598562	13	28674876	<i>FLT3</i>	TSS200	Island				
2	cg03401096	11	123301171			Island				
3	cg15897970	14	62583962	<i>LINC00643</i>	TSS200	Island				
4	cg15400238	2	154335640	<i>RPRM</i>	TSS1500	S_Shore				
5	cg13769223	21	27011963	<i>JAM2</i>	1stExon	Island				
6	cg07122245	7	37488428	<i>ELMO1</i>	5'UTR	Island				
7	cg01257828	3	192126996	<i>FGF12</i>	TSS200	Island				
8	cg04942472	16	58497239	<i>NDRG4</i>	TSS1500	Island				
9	cg13282152	8	65492683	<i>BHLHE22</i>	TSS200	N_Shore				
10	cg01941671	4	81952099	<i>BMP3</i>	TSS200	Island				
Top 10 (Delta β value)										
1	cg03401096	11	123301171			Island				
2	cg11092616	6	72596493	<i>RIMS1</i>	TSS200	S_Shore				
3	cg07122245	7	37488428	<i>ELMO1</i>	5'UTR	Island				
4	cg05598562	13	28674876	<i>FLT3</i>	TSS200	Island				
5	cg04942472	16	58497239	<i>NDRG4</i>	TSS1500	Island				
6	cg10748160	4	6202384	<i>JAKMIP1</i>	TSS200	Island				
7	cg14186641	6	88876741	<i>CNR1</i>	TSS1500	Island				
8	cg13769223	21	27011963	<i>JAM2</i>	1stExon	Island				
9	cg13282152	8	65492683	<i>BHLHE22</i>	TSS200	N_Shore				
10	cg15400238	2	154335640	<i>RPRM</i>	TSS1500	S_Shore				
No.	IllumID	Methylation levels (mean beta value)				High vs Intermediate	High-quality pyrosequencing primer			
		Blood	Low-risk (G1)	Intermediate-risk (G2)	High-risk (G3)			Delta [High-Intermediate]	P(TT)	q(TT)
(A) The large difference analysis										
Top 10 (TT;P value)										
1	cg05598562	0.009	0.014	0.099	0.426	0.106	0.327	3.97E-07	3.89E-03	Designed
2	cg03401096	0.016	0.036	0.137	0.508	0.137	0.371	4.93E-06	6.97E-03	Not designed
3	cg15897970	0.059	0.052	0.166	0.404	0.126	0.238	6.62E-05	1.27E-02	Designed
4	cg15400238	0.157	0.059	0.139	0.425	0.166	0.286	1.18E-04	1.37E-02	Designed
5	cg13769223	0.025	0.051	0.149	0.439	0.160	0.290	1.28E-04	1.38E-02	Designed

Table 1 (continued)

No.	IllumID	Methylation levels (mean beta value)				High vs Intermediate			High-quality pyrosequencing primer			
		Blood	Low-risk (G1)		Intermediate-risk (G2)	SD	High-risk (G3)	SD		Delta [High-Intermediate]	P(TT)	q(TT)
			Young	Old								
6	cg07122245	0.103	0.038	0.035	0.139	0.145	0.469	0.185	0.330	1.41E-04	1.41E-02	Designed
7	cg01257828	0.006	0.055	0.033	0.175	0.091	0.402	0.133	0.227	1.49E-04	1.41E-02	Not designed
8	cg04942472	0.009	0.059	0.063	0.176	0.132	0.471	0.177	0.295	2.48E-04	1.51E-02	Not Designed
9	cg13282152	0.016	0.033	0.022	0.114	0.074	0.403	0.190	0.288	3.37E-04	1.57E-02	Designed
10	cg01941671	0.124	0.036	0.019	0.127	0.102	0.350	0.141	0.222	3.92E-04	1.59E-02	Not designed
Top 10 (Delta β value)												
1	cg03401096	0.016	0.036	0.016	0.137	0.152	0.508	0.137	0.371	4.93E-06	6.97E-03	Not designed
2	cg11092616	0.005	0.012	0.006	0.149	0.153	0.495	0.224	0.347	4.22E-04	1.59E-02	Designed
3	cg07122245	0.103	0.038	0.035	0.139	0.145	0.469	0.185	0.330	1.41E-04	1.41E-02	Designed
4	cg05598562	0.009	0.014	0.013	0.099	0.110	0.426	0.106	0.327	3.97E-07	3.89E-03	Designed
5	cg04942472	0.009	0.059	0.063	0.176	0.132	0.471	0.177	0.295	2.48E-04	1.51E-02	Not Designed
6	cg10748160	0.038	0.026	0.018	0.104	0.126	0.399	0.191	0.295	4.07E-04	1.59E-02	Not designed
7	cg14186641	0.005	0.030	0.020	0.207	0.177	0.501	0.153	0.294	4.14E-04	1.59E-02	Not designed
8	cg13769223	0.025	0.051	0.020	0.149	0.129	0.439	0.160	0.290	1.28E-04	1.38E-02	Designed
9	cg13282152	0.016	0.033	0.022	0.114	0.074	0.403	0.190	0.288	3.37E-04	1.57E-02	Designed
10	cg15400238	0.157	0.059	0.041	0.139	0.101	0.425	0.166	0.286	1.18E-04	1.37E-02	Designed
No.	IllumID	Human genome (Hg19)				UCSC_RefGene_Name			UCSC_RefGene_Group			Relation_to_UCSC_CpG_Island
		CHR		MAPINFO								
(B) The iEVORA-based analysis												
The large difference												
1	cg00432059	19	20012489	ZNF93	Body	S_Shore						
2	cg02636041	10	43698008	RASGEF1A	Body	Island						
Three sequential probes												
1	cg07655479	4	144480551	GUSBP5	TSS200	Island						
2	cg06123783	4	144480558	GUSBP5	TSS200	Island						
3	cg13487949	4	144480561	GUSBP5	TSS200	Island						
4	cg10166205	19	50316699	FUZ	TSS1500	S_Shore						
5	cg02709432	8	120651236	ENPP2	TSS200	S_Shore						
6	cg15653282	18	11689218	GNAL	5'UTR	Island						

Table 1 (continued)

No.	ImmID	Methylation levels (mean beta value)				High vs Intermediate				High-quality pyrosequencing primer		
		Blood		Intermediate-risk (G2)		Delta [High-Intermediate]	High-risk (G3)	SD	q(TT)			
		Low risk (G1)	Old	Young	Old						P(TT)	
(B) The iEVORA-based analysis												
The large difference												
1	cg00432059	0.069	0.008	0.019	0.034	0.025	0.271	0.204	0.236	5.03E-08	1.31E-05	Designed
2	cg02636041	0.024	0.008	0.011	0.019	0.021	0.223	0.165	0.204	5.33E-08	1.37E-05	Not designed
Three sequential probes												
1	cg07655479	0.145	0.009	0.056	0.047	0.020	0.203	0.125	0.156	8.79E-07	1.32E-04	Designed
2	cg06123783	0.044	0.011	0.029	0.032	0.014	0.166	0.113	0.134	4.20E-08	1.12E-05	Not designed
3	cg13487949	0.189	0.004	0.074	0.042	0.017	0.171	0.095	0.129	3.16E-06	3.69E-04	Not designed
4	cg10166205	0.023	0.003	0.005	0.005	0.005	0.109	0.134	0.104	6.27E-14	3.37E-10	Not designed
5	cg02709432	0.032	0.048	0.042	0.060	0.017	0.149	0.093	0.089	3.60E-06	4.04E-04	Not designed
6	cg15653282	0.007	0.002	0.004	0.009	0.007	0.074	0.092	0.065	1.49E-10	1.26E-07	Not designed

CHR chromosome number, NA not applicable, TT t test, P P value, q FDR

Results

Isolation of candidate novel markers by two algorithms

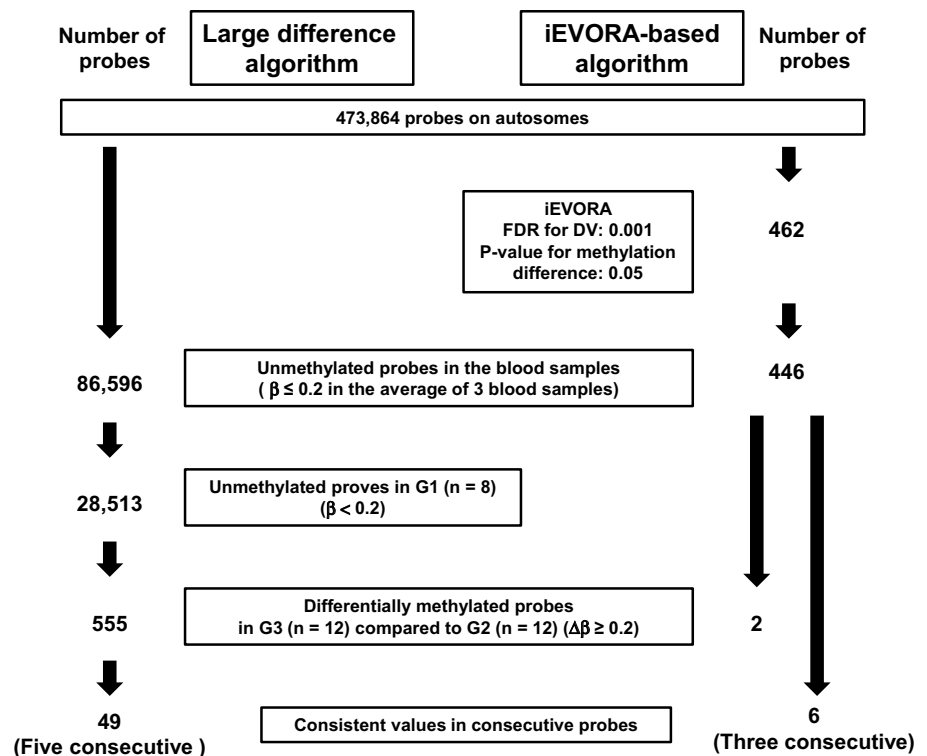
To isolate the methylation cancer risk markers, two screening algorithms (the large difference and iEVORA-based methods) were adopted (Fig. 1). In the large difference method, to minimize the influence of contamination of inflammatory cells, probes with high methylation in three peripheral blood cell samples were first removed from the 473,864 probes on autosomes to obtain 86,596 probes. Then, 28,513 probes unmethylated in G1 (healthy volunteers never infected with *H. pylori*) were selected and 555 probes with a large difference ($\Delta\beta \geq 0.2$) between G2 ($n = 12$; *H. pylori*-eradicated healthy individuals with gastric mucosal atrophy) and G3 ($n = 12$; gastric cancer patients after endoscopic treatment and *H. pylori*-eradication) were isolated. Finally, to avoid the isolation of an outlier signal in a region, 49 regions were selected in which five consecutive probes showed consistent values and the value for the central probe was used. From these 49 candidates, ten candidates selected based on their *P* values and another ten candidates (of which six overlapped) selected based on mean differences of the beta values were used for validation (Table 1).

In the iEVORA-based method, 462 probes with high variances in G3 ($n = 12$) were first selected. Then, similar to the large difference method, 446 probes unmethylated in the peripheral blood cells were selected. Among the 446 probes, two candidates were isolated based on the large mean differences between G2 ($n = 12$) and G3 ($\Delta\beta \geq 0.2$). Additionally, six regions in which three consecutive probes showed consistent values were selected and the value for the central probe was used. Collectively, eight candidates were isolated by the iEVORA-based method and used for validation (Table 1).

All candidate markers analyzed were validated by a different set

Among the 14 (large difference method) and eight (iEVORA-based method) candidates, we successfully designed pyrosequencing primers for nine regions, *FLT3*, *LINC00643*, *RPRM*, *JAM2*, *ELMO1*, *BHLHE22*, *RIMS1*, *GUSBP5*, and *ZNF93* (Supplementary Table 4). First, to analyze the effect of *H. pylori* eradication on their methylation levels, we analyzed 63 G2 samples taken before and after *H. pylori* eradication. Consistent with previous reports [6, 7], all candidate markers showed comparable or decreased methylation levels after eradication compared to those before eradication (Supplementary Fig. 1).

Fig. 1 The isolation of candidate novel epigenetic (DNA methylation) risk markers by two algorithms from 473,864 probes on autosomes. Left: the large difference algorithm. The probes unmethylated (β value < 0.2) in blood samples and G1 were first selected. Then, 555 probes with a large difference in methylation levels ($\Delta\beta \geq 0.2$) between G3 and G2 were selected. Finally, 49 regions in which five consecutive probes showed consistent values were obtained. Right: iEVORA-based algorithm. First, 462 probes were isolated by iEVORA. Subsequently, 446 unmethylated probes in the blood samples were selected. After consideration of both large difference and consecutive probes with consistent values, two and six probes were isolated, respectively



Next, the validation set of 63 G2 and 82 G3 samples was analyzed by the nine candidate markers and a previous marker (*miR-124a-3*) [12]. All the candidate markers and *miR-124a-3* had much higher methylation levels in G3 than in G2 ($P < 0.0001$; Fig. 2, Supplementary Table 5). Therefore, we were able to validate all the nine novel methylation markers for gastric cancer risk.

High correlations among novel methylation markers

To explore whether the novel methylation markers reflected the overall epigenomic damage that resulted from past exposure to environmental factors and host responses to them [26, 27], correlations among the markers were analyzed. As expected, high correlation coefficients ($R = 0.74$ – 0.97 ; Table 2) were observed between any two of the novel markers and *miR-124a-3*. In particular, the novel markers identified by the large difference method showed very strong correlations ($R = 0.82$ – 0.97) (Table 2, Supplementary Fig. 2). The known marker, *miR-124a-3*, was also highly correlated ($R = 0.66$ – 0.93) with the novel markers. These results suggested that our novel markers represented one shared entity, namely epigenomic damage as a result of *H. pylori* infection.

High-performance novel methylation markers

The clinical performance of the novel methylation cancer risk markers was assessed by ROC curve analysis (Fig. 3,

Supplementary Table 6). All the novel markers showed high AUCs (0.70–0.80), odds ratios (5.43–23.41), high-to-moderate sensitivity (0.59–0.94), and high-to-moderate specificity (0.59–0.83) (Supplementary Table 6). Compared with *miR-124a-3* (AUC; 0.74, odds ratio; 8.01), some novel markers had higher AUCs and better sensitivity or specificity. Thus, we were able to establish novel methylation markers for bisulfate pyrosequencing for the estimation of gastric cancer risk that was potentially superior to *miR-124a-3*.

Passenger methylation of novel marker genes in gastric carcinogenesis

To explore the possible roles of our novel marker genes in gastric carcinogenesis, we analyzed their expression levels in normal gastric mucosa using an expression microarray (Supplementary Table 7). Most of the marker genes showed very low expression levels irrespective of *H. pylori* infection statuses, which agreed with the data in the Genotype-Tissue Expression (GTEx, URL: <https://www.gtexportal.org/home/>) database (data not shown). This finding indicated that these genes are unlikely to play any biological roles in the stomach, and that their aberrant methylation is unlikely to be involved in gastric carcinogenesis, being passenger methylation. Nevertheless, *JAM2* and *ELMO1*, whose target probes were located in CpG sites near their transcription start sites (TSSs), showed moderate expression levels.

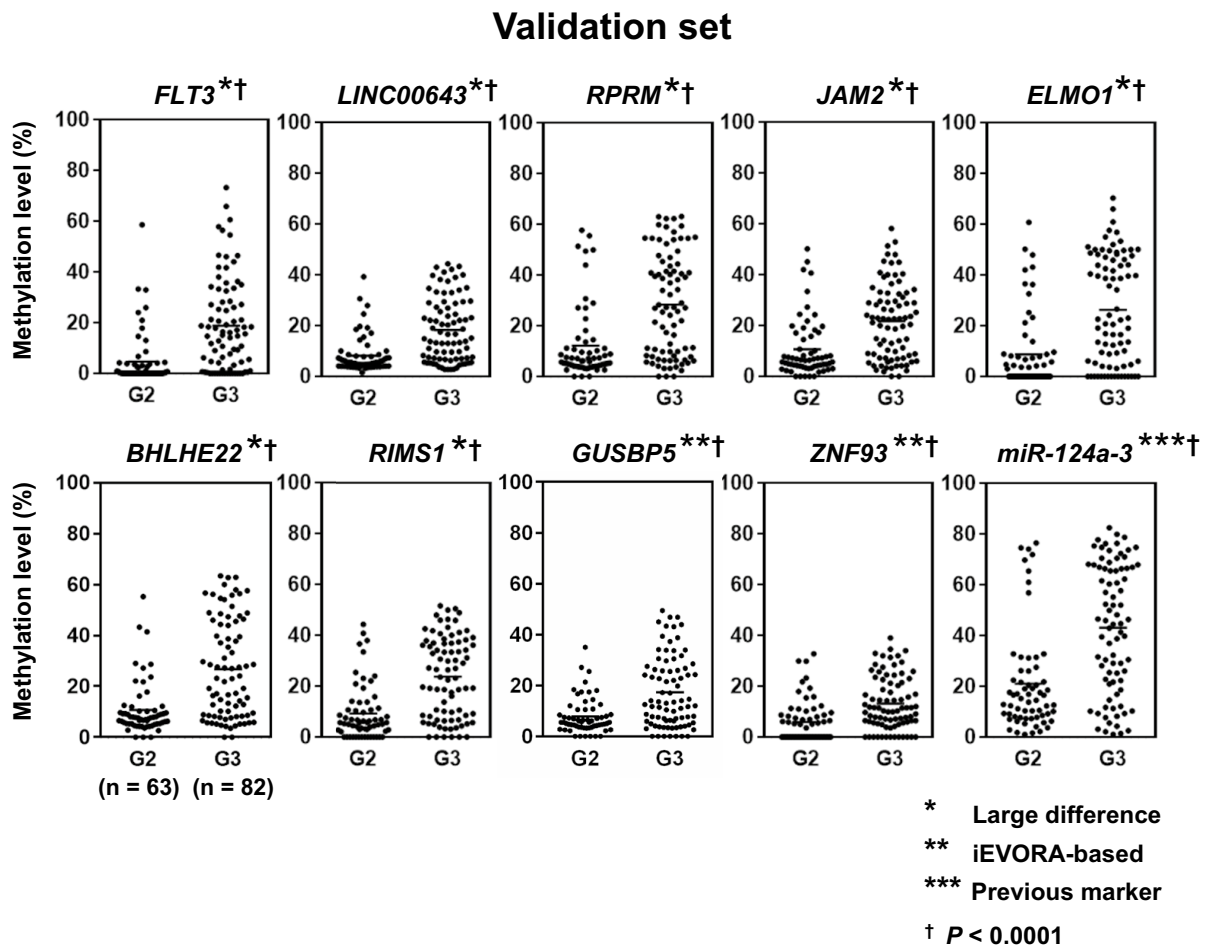


Fig. 2 The methylation levels of the nine candidate markers and a previous marker, *miR-124a-3*, in the validation set. The horizontal line represents the mean methylation level in each group. The meth-

ylation levels of all nine candidate markers, as well as *miR-124a-3*, were significantly higher in the high-risk group (G3, $n = 82$) than in the intermediate-risk group (G2, $n = 63$). $*P < 0.0001$

Relationships between methylation levels and the extent of gastric atrophy/gender/aging

Gastric atrophy and gender are also strong risk factors for gastric cancer [28, 29], and the efficacy of the novel methylation risk markers irrespective of these confounding factors was examined by subgroup analysis. First, the methylation levels of the markers did not show differences in the extent of gastric atrophy within G2 or G3 (Supplementary Fig. 3a and 3b). Among the individuals with moderate gastric atrophy, significantly higher methylation levels were observed in G3 ($n = 42$) than in G2 ($n = 26$). Significantly higher methylation levels were observed in G3 than in G2 for all novel markers and *miR-124a-3* (Supplementary Fig. 4, Supplementary Table 8a).

The subgroup analysis for men confirmed significantly higher methylation levels in G3 ($n = 73$) than in G2 ($n = 41$) for all markers. For women, the subgroup analysis also revealed higher methylation levels in G3 ($n = 9$) than in G2

($n = 22$) for all markers, although some of the differences were not significant owing to the limited sample size (Supplementary Table 8b).

In addition, age-related methylation is known to contribute to predisposition to carcinogenesis by inactivating specific genes [30]. We compared methylation levels between four young and four older healthy individuals never infected with *H. pylori* (G1). No candidate novel methylation markers showed significant age-related differences (Table 1). This suggests that the novel methylation markers are unlikely to be affected by aging. On the other hand, regarding *H. pylori*-eradicated individuals (G2 and G3), the methylation levels of some markers showed weak positive correlations with age (Supplementary Fig. 5).

Taken together, these results suggested that our novel markers have the potential to be applied to individuals after *H. pylori* eradication, irrespective of the extent of gastric atrophy, gender, and age.

Table 2 Correlation coefficient between methylation markers

	Large difference							iEVORA-based		Previous marker	
	Marker	<i>FLT3</i>	<i>LINC00643</i>	<i>RPRM</i>	<i>JAM2</i>	<i>ELMO1</i>	<i>BHLHE22</i>	<i>RIMS1</i>	<i>GUSBP5</i>	<i>ZNF93</i>	<i>miR-124a-3</i>
Large difference	<i>FLT3</i>	1									
	<i>LINC00643</i>	0.84	1								
	<i>RPRM</i>	0.86	0.90	1							
	<i>JAM2</i>	0.82	0.91	0.88	1						
	<i>ELMO1</i>	0.85	0.95	0.94	0.93	1					
	<i>BHLHE22</i>	0.84	0.88	0.88	0.87	0.90	1				
	<i>RIMS1</i>	0.83	0.95	0.93	0.93	0.97	0.89	1			
iEVORA-based	<i>GUSBP5</i>	0.80	0.80	0.80	0.76	0.79	0.77	0.80	1		
	<i>ZNF93</i>	0.74	0.83	0.76	0.78	0.77	0.78	0.80	0.88	1	
Previous marker	<i>miR-124a-3</i>	0.79	0.85	0.92	0.85	0.93	0.84	0.91	0.71	0.66	1

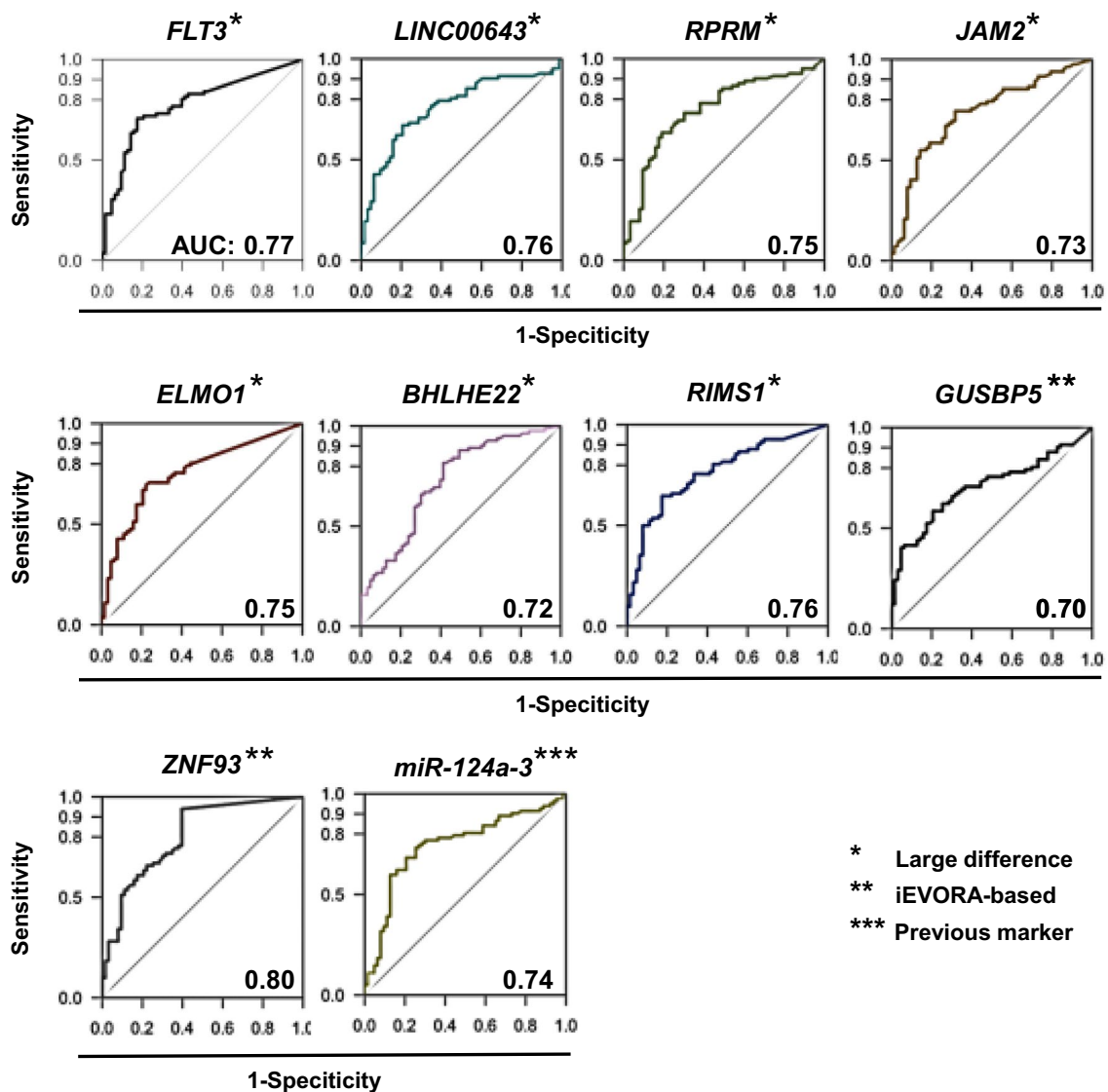


Fig. 3 The ROCs of nine novel methylation markers and *miR-124a-3*. The AUC values are shown in each box. The AUCs of some of the novel methylation markers were higher than that of *miR-124a-3*. AUC: area under the curve

Discussion

In the current study, we established nine novel epigenetic markers able to identify cancer patients even among *H. pylori*-eradicated individuals using a genome-wide screening based on two different algorithms. Our novel markers showed sufficiently high odds ratios, even among individuals with gastric atrophy. Some markers showed superior performance compared with our previous marker, *miR-124a-3*. We believe that BeadArray technology-based comprehensive methylation analysis and validation in a large number of samples provided us with promising methylation markers. Importantly, our epigenetic markers are not likely to be affected by possible contamination of blood cells in the biopsy sample because we eliminated such probes during our screening.

The novel methylation markers showed comparable or decreased methylation levels after eradication, compared to those before eradication, in agreement with previous reports [6, 7]. Our novel markers were selected as those showing very low methylation levels in blood cells. Therefore, the decrease in methylation levels after eradication is unlikely to have resulted from decreased infiltration of inflammatory blood cells after eradication, but likely resulted from the supply of new progenitor cells from unmethylated stem cells. Additionally, before eradication, a large fraction of inflammatory blood cells can reduce the fraction of epithelial cells. After eradication, an increased fraction of epithelial cells may result in a trade-off with decreased methylation levels in epithelial cells for some markers, leading to methylation levels comparable to those before eradication.

In general, genes with low expression are known to be susceptible to aberrant DNA methylation [25, 31, 32]. Indeed, seven marker genes had very low expression levels in the gastric mucosa, indicating that aberrant DNA methylation of these marker genes was passenger methylation. On the other hand, *JAM2* and *ELMO1* had moderate expression levels, and their target probes were located in CpG islands near the TSSs. Therefore, there remains a possibility that these two genes were methylation-silenced, being drivers. Methylation silencing of *JAM2*, involved in cell adhesion, and *ELMO1*, involved in cell motility, is reported in colorectal cancer [33–35]. However, more importantly, considering that the high correlations of methylation levels between the marker genes, the high susceptibility of the marker genes to aberrant methylation due to *H. pylori* infection, irrespective of their functions in gastric carcinogenesis, was considered the explanation for the good performance of these novel marker genes in risk prediction.

We also analyzed the effect of age-related methylation. However, no difference in methylation levels was observed between young and older healthy individuals never infected

with *H. pylori* (G1), and the effect of aging was considered to be minimal. At the same time, in *H. pylori*-eradicated individuals (G2 and G3), some markers were slightly associated with age. Importantly, methylation levels are known to increase over time to reflect the infection period [5]; therefore, the slight increase correlated with age may be related to the infection period.

The potential limitations of our study include the sample bias collected in two studies [14, 16]. Additionally, *H. pylori*-eradicated individuals who were collected as G2 may contain a small number of cancer patients (G3). This might have resulted in a lower sensitivity and specificity in our markers than expected. In addition, the differences in prevalence of the extent of gastric atrophy and gender were observed between G3 and G2. However, the subgroup analysis suggested a high potential applicability of our methylation markers.

One of the prominent applications of the novel markers is a large-scale multicenter prospective cohort study for the risk stratification of primary gastric cancer in healthy individuals with extensive gastric atrophy after *H. pylori* eradication (UMIN00016894). This nationwide, ongoing prospective study is expected to optimize the surveillance system based on stratified individual risk, which will contribute to precision medicine.

In conclusion, we established nine novel methylation markers that identify *H. pylori*-eradicated individuals with high risk of gastric cancer.

Acknowledgements The authors are grateful to Drs. Y. Kakugawa, Y. Otake and T. Gotoda for sample and data acquisition, and Dr. N. Hattori for her advice, and Drs. K. Ichimura, Y. Matsushita, and M. Kitahara of Division of Brain Tumor Translational Research in National Cancer Center Research Institute for their technical assistance with the experiments.

Funding Practical Research for Innovative Cancer Control from Japan Agency for Medical Research and Development, AMED (15ck0106023h0002, 16ck0106023h0003, 17ck0106267h0001).

Compliance with ethical standards

Conflict of interest MM and TU made a joint patent application with Sysmex Corporation for identified epigenetic markers.

Ethical approval All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1964 and later versions. Informed consent or substitute for it was obtained from all patients for being included in the study.

References

1. Lee YC, Chiang TH, Chou CK, Tu YK, Liao WC, Wu MS, et al. Association between *Helicobacter pylori* eradication and gastric

- cancer incidence: a systematic review and meta-analysis. *Gastroenterology*. 2016;150(1113–1124):e5.
2. Tsuda M, Asaka M, Kato M, Matsushima R, Fujimori K, Akino K, et al. Effect on *Helicobacter pylori* eradication therapy against gastric cancer in Japan. *Helicobacter*. 2017;22. <https://doi.org/10.1111/hel.12415>.
 3. Niwa T, Tsukamoto T, Toyoda T, Mori A, Tanaka H, Maekita T, et al. Inflammatory processes triggered by *Helicobacter pylori* infection cause aberrant DNA methylation in gastric epithelial cells. *Cancer Res*. 2010;70:1430–40.
 4. Niwa T, Toyoda T, Tsukamoto T, Mori A, Tatematsu M, Ushijima T. Prevention of *Helicobacter pylori*-induced gastric cancers in gerbils by a DNA demethylating agent. *Cancer Prev Res (Phila)*. 2013;6:263–70.
 5. Takeshima H, Niwa T, Toyoda T, Wakabayashi M, Yamashita S, Ushijima T. Degree of methylation burden is determined by the exposure period to carcinogenic factors. *Cancer Sci*. 2017;108:316–21.
 6. Nakajima T, Enomoto S, Yamashita S, Ando T, Nakanishi Y, Nakazawa K, et al. Persistence of a component of DNA methylation in gastric mucosae after *Helicobacter pylori* eradication. *J Gastroenterol*. 2010;45:37–44.
 7. Shin CM, Kim N, Lee HS, Park JH, Ahn S, Kang GH, et al. Changes in aberrant DNA methylation after *Helicobacter pylori* eradication: a long-term follow-up study. *Int J Cancer*. 2013;133:2034–42.
 8. Maeda M, Moro H, Ushijima T. Mechanisms for the induction of gastric cancer by *Helicobacter pylori* infection: aberrant DNA methylation pathway. *Gastric Cancer*. 2017;20:8–15.
 9. Nakajima T, Maekita T, Oda I, Gotoda T, Yamamoto S, Umemura S, et al. Higher methylation levels in gastric mucosae significantly correlate with higher risk of gastric cancers. *Cancer Epidemiol Biomark Prev*. 2006;15:2317–21.
 10. Maekita T, Nakazawa K, Mihara M, Nakajima T, Yanaoka K, Iguchi M, et al. High levels of aberrant DNA methylation in *Helicobacter pylori*-infected gastric mucosae and its possible association with gastric cancer risk. *Clin Cancer Res*. 2006;12:989–95.
 11. Ushijima T. Epigenetic field for cancerization. *J Biochem Mol Biol*. 2007;40:142–50.
 12. Ando T, Yoshida T, Enomoto S, Asada K, Tatematsu M, Ichinose M, et al. DNA methylation of microRNA genes in gastric mucosae of gastric cancer patients: its possible involvement in the formation of epigenetic field defect. *Int J Cancer*. 2009;124:2367–74.
 13. Nanjo S, Asada K, Yamashita S, Nakajima T, Nakazawa K, Maekita T, et al. Identification of gastric cancer risk markers that are informative in individuals with past *H. pylori* infection. *Gastric Cancer*. 2012;15:382–8.
 14. Asada K, Nakajima T, Shimazu T, Yamamichi N, Maekita T, Yokoi C, et al. Demonstration of the usefulness of epigenetic cancer risk prediction by a multicentre prospective cohort study. *Gut*. 2015;64:388–96.
 15. Maeda M, Nakajima T, Oda I, Shimazu T, Yamamichi N, Maekita T, et al. High impact of methylation accumulation on metachronous gastric cancer: 5-year follow-up of a multicentre prospective cohort study. *Gut*. 2017;66:1721–3.
 16. Shimazu T, Asada K, Charvat H, Kusano C, Otake Y, Kakugawa Y, et al. Association of gastric cancer risk factors with DNA methylation levels in gastric mucosa of healthy Japanese: a cross-sectional study. *Carcinogenesis*. 2015;36:1291–8.
 17. Mori G, Nakajima T, Asada K, Shimazu T, Yamamichi N, Maekita T, et al. Incidence of and risk factors for metachronous gastric cancer after endoscopic resection and successful *Helicobacter pylori* eradication: results of a large-scale, multicenter cohort study in Japan. *Gastric Cancer*. 2016;19:911–8.
 18. Take S, Mizuno M, Ishiki K, Nagahara Y, Yoshida T, Yokota K, et al. Baseline gastric mucosal atrophy is a risk factor associated with the development of gastric cancer after *Helicobacter pylori* eradication therapy in patients with peptic ulcer diseases. *J Gastroenterol*. 2007;42(Suppl 17):21–7.
 19. Kimura K, Takemoto T. An endoscopic recognition of the atrophic border and its significance in chronic gastritis. *Endoscopy*. 1969;1:87–97.
 20. Shigematsu Y, Niwa T, Yamashita S, Taniguchi H, Kushima R, Katai H, et al. Identification of a DNA methylation marker that detects the presence of lymph node metastases of gastric cancers. *Oncol Lett*. 2012;4:268–74.
 21. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, et al. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics*. 2013;29:189–96.
 22. Teschendorff AE, Gao Y, Jones A, Ruebner M, Beckmann MW, Wachter DL, et al. DNA methylation outliers in normal breast tissue identify field defects that are enriched in cancer. *Nat Commun*. 2016;7:10478.
 23. Teschendorff AE, Jones A, Fiegl H, Sargent A, Zhuang JJ, Kitchener HC, et al. Epigenetic variability in cells of normal cytology is associated with the risk of future morphological transformation. *Genome Med*. 2012;4:24.
 24. Matsuda Y, Yamashita S, Lee YC, Niwa T, Yoshida T, Gyobu K, et al. Hypomethylation of Alu repetitive elements in esophageal mucosa, and its potential contribution to the epigenetic field for cancerization. *Cancer Causes Control*. 2012;23:865–73.
 25. Takeshima H, Yamashita S, Shimazu T, Niwa T, Ushijima T. The presence of RNA polymerase II, active or stalled, predicts epigenetic fate of promoter CpG islands. *Genome Res*. 2009;19:1974–82.
 26. Ushijima T, Hattori N. Molecular pathways: involvement of *Helicobacter pylori*-triggered inflammation in the formation of an epigenetic field defect, and its usefulness as cancer risk and exposure markers. *Clin Cancer Res*. 2012;18:923–9.
 27. Hattori N, Ushijima T. Epigenetic impact of infection on carcinogenesis: mechanisms and applications. *Genome Med*. 2016;8:10.
 28. Ohata H, Kitauchi S, Yoshimura N, Mugitani K, Iwane M, Nakamura H, et al. Progression of chronic atrophic gastritis associated with *Helicobacter pylori* infection increases risk of gastric cancer. *Int J Cancer*. 2004;109:138–43.
 29. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer*. 2015;136:E359–86.
 30. Issa JP, Ottaviano YL, Celano P, Hamilton SR, Davidson NE, Baylin SB. Methylation of the oestrogen receptor CpG island links ageing and neoplasia in human colon. *Nat Genet*. 1994;7:536–40.
 31. Keshet I, Schlesinger Y, Farkash S, Rand E, Hecht M, Segal E, et al. Evidence for an instructive mechanism of de novo methylation in cancer cells. *Nat Genet*. 2006;38:149–53.
 32. Takeshima H, Ushijima T. Methylation destiny: Moira takes account of histones and RNA polymerase II. *Epigenetics*. 2010;5:89–95.
 33. Yagi K, Akagi K, Hayashi H, Nagae G, Tsuji S, Isagawa T, et al. Three DNA methylation epigenotypes in human colorectal cancer. *Clin Cancer Res*. 2010;16:21–33.
 34. Kok-Sin T, Mokhtar NM, Ali Hassan NZ, Sagap I, Mohamed Rose I, Harun R, et al. Identification of diagnostic markers in colorectal cancer via integrative epigenomics and genomics data. *Oncol Rep*. 2015;34:22–32.
 35. Oster B, Thorsen K, Lamy P, Wojdacz TK, Hansen LL, Birkenkamp-Demtroder K, et al. Identification and validation of highly frequent CpG island hypermethylation in colorectal adenomas and carcinomas. *Int J Cancer*. 2011;129:2855–66.