


Genome-wide profiling of normal gastric mucosa identifies *Helicobacter pylori*- and cancer-associated DNA methylome changes

Hae Dong Woo^{1,2}, Nora Fernandez-Jimenez¹, Akram Ghantous¹, Davide Degli Esposti¹, Cyrille Cuenin¹, Vincent Cahais¹, Il Ju Choi³, Young-Il Kim³, Jeongseon Kim² and Zdenko Herceg¹ 

¹Epigenetics Group, International Agency for Research on Cancer (IARC), 150 Cours Albert Thomas, Lyon, 69372, France

²Department of Cancer Biomedical Science, Graduate School of Cancer Science and Policy, National Cancer Center, Goyang, 10408, Republic of Korea

³Center for Gastric Cancer, National Cancer Center Hospital, National Cancer Center, Goyang, 10408, Republic of Korea

The large geographic variations in the incidence of gastric cancer (GC) are likely due to differential environmental exposures, in particular to *Helicobacter pylori* (*H. pylori*) infection. We aimed to investigate the impact of *H. pylori* on the epigenome in normal gastric mucosa and methylation changes associated with cancer risk independent of *H. pylori*. A discovery set of normal gastric mucosa from GC cases ($n = 42$) and controls ($n = 42$), nested in a large case-control study and stratified by *H. pylori* status, were subjected to genome-wide methylation profiling. Single-nucleotide polymorphism arrays from peripheral blood leukocytes were used to conduct methylation quantitative trait loci (mQTL) analysis. A validation set of gastric mucosa samples ($n = 180$) was used in the replication phase. We found 1,924 differentially methylated positions (DMPs) and 438 differentially methylated regions (DMRs) associated with *H. pylori* infection, most of which were hypermethylated. Significant methylation alterations identified in the initial set were successfully replicated. Furthermore, the *H. pylori*-associated DMP/Rs showed marked stability ('epigenetic memory') after *H. pylori* clearance. Interestingly, we found 152 DMRs associated with cancer risk independent of the *H. pylori* status in normal gastric mucosa. The methylation score derived from three biomarkers was a strong predictor of GC. Finally, the mQTL analysis indicated that the *H. pylori*- and cancer-specific methylation signatures were minimally affected by genetic variation. The comprehensively characterized methylome changes associated with *H. pylori* infection and GC risk in our study might serve as potential biomarkers for early cancer progression in tumour-free gastric mucosa.

Introduction

Gastric cancer (GC) is the second leading cause of cancer-related death worldwide.¹ The 5-year relative survival rate among people diagnosed with GC in Japan and the Republic of

Korea is approximately 73%,^{2,3} which is relatively high due to extensive nationwide GC screening programmes. However, despite progress in therapies and preventive strategies, the 5-year survival rates of GC patients in other countries are low

Key words: gastric cancer, *H. pylori*, biomarkers of cancer risk, methylation profiling, epigenetic memory

Abbreviations used: CpG: 5'-C-phosphate-G-3'; DMP: differentially methylated position; DMR: differentially methylated region; FDR: false discovery rate; GC: gastric cancer; GWAS: genome-wide association study; HM450 array: Infinium Human Methylation 450K BeadChip Kit (Illumina, San Diego, CA); IARC: International Agency for Research on Cancer; mQTL: methylation quantitative trait locus; ROC: receiver operating characteristic; SNP: single-nucleotide polymorphism; SVA: surrogate variable analysis

Additional Supporting Information may be found in the online version of this article.

Conflict of interest: The authors declare that they have no conflict of interest related to this article.

Grant sponsor: Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education; **Grant number:** 2015R1A6A3A03020453; **Grant sponsor:** National Cancer Center of Korea; **Grant number:** 1410260; **Grant sponsor:** IARC Epigenetics Group

DOI: 10.1002/ijc.31381

This is an open access article distributed under the terms of the Creative Commons Attribution IGO License IARC's preferred IGO license is the non-commercial: <https://creativecommons.org/licenses/by-nc/3.0/igo/legalcode> which permits non-commercial unrestricted use, distribution and reproduction in any medium, provided that the original work is properly cited. In any reproduction of this article there should not be any suggestion that IARC/WHO or the article endorse any specific organization or products. The use of the IARC/WHO logo is not permitted. This notice should be preserved along with the article's URL.

History: Received 4 Dec 2017; Accepted 2 Mar 2018; Online 25 Mar 2018

Correspondence to: Zdenko Herceg, IARC, 150 Cours Albert Thomas, Lyon 69008, France, Tel.: +33 4 72 73 83 98, Fax: +33 4 72 73 83 22, E-mail: hercegz@iarc.fr; or Jeongseon Kim, National Cancer Center, 323 Ilsan-ro, Ilsandong-gu, Goyang, Gyeonggi-do 10408, Republic of Korea, Tel.: +82 31 920 2570, Fax: +82 31 920 2579, E-mail: jskim@ncc.re.kr

What's new?

To investigate why gastric cancer is more prevalent in some countries than others, these authors asked how *H. pylori* infection affects gene methylation. Here, they compared genomes of normal gastric mucosa from gastric cancer cases and controls, with and without *H. pylori* infection. They noted changes in methylation correlating with *H. pylori* infection, including current and prior infections, and identified 438 differently methylated regions associated with *H. pylori* infections. Interestingly, 152 differently methylated regions correlated with increased gastric cancer risk regardless of *H. pylori* status. Some of these methylation patterns could lead to biomarkers for detecting disease progression.

(25% in Europe and 28% in the United States) due to late detection of the disease.⁴ Infection with the bacterium *Helicobacter pylori* (*H. pylori*), which is classified as carcinogenic to humans,⁵ is thought to be the single most common cause of GC and is estimated to be responsible for more than two thirds of cases.⁶ In several countries, *H. pylori* eradication is part of a national preventive strategy; however, *H. pylori* infection remains highly prevalent in many parts of the world, and the impact of past infection on GC risk is unclear.

Deregulation of epigenetic mechanisms is one of the hallmarks of cancer,⁷ and the epigenome can act as a molecular sensor of exposure to environmental factors, including infectious agents. *H. pylori* and the inflammatory process it triggers may promote GC development by inducing deregulation of the DNA methylome.^{8,9} A number of studies have conducted DNA methylation analysis in GC tumours and their adjacent tissues using candidate gene¹⁰ or genome-wide approaches,^{11–13} however, these studies were limited to cancer tissue, and the effects of *H. pylori* were not considered. Three studies have identified methylation signatures in normal gastric mucosa in response to *H. pylori* infection, but these involved only several genes or partial DNA methylome coverage^{14–16} or did not include GC patients.¹⁶ Two studies, which tested three candidate genes to identify risk markers for gastric metachronous cancer, were focused on the progression of cancer after excluding the effect of *H. pylori* infection.^{17,18} Furthermore, *H. pylori* alone might not be sufficient to drive GC development.¹⁹ In fact, only a fraction of infected individuals develop GC.²⁰ Therefore, other factors likely influence GC tumorigenesis. For example, several studies have proposed that germline mutations might be associated with susceptibility to GC risk.^{21,22} However, the aetiology of most of these associations remains largely unknown.

In our study, we aimed to analyse a set of normal gastric mucosa from cases and controls representing various *H. pylori* and GC statuses with both genome- and epigenome-wide coverage. Specifically, we were able to investigate the impact of both current *H. pylori* infection and epigenetic memory of past (eradicated) infection on aberrant DNA methylation, to identify DNA methylation signatures that predict GC risk and to determine whether the methylation changes induced in normal mucosa by infection and those induced by field effects in cancer have any common mechanisms. We also assessed the extent to which the identified methylation alterations were influenced by

genetic variation. Finally, we validated the results using a set of 180 samples, focusing on targeted molecular signatures.

Materials and Methods**Study population and data collection**

Participants were recruited at the National Cancer Centre Hospital in the Republic of Korea between March 2011 and June 2014. Case samples of individuals who had been histologically confirmed as early GC patients within the preceding 3 months at the hospital's Centre for Gastric Cancer were collected. Early GC was defined as GC confined to the mucosa or submucosa, regardless of the lymph node metastasis status. Control samples were collected from individuals selected via health-screening examinations conducted by the Centre for Cancer Prevention and Detection at the same hospital.

In total, 1,308 initial biopsy samples were collected (427 case and 881 control samples). After endoscopy and examination of the stomach, five further biopsy samples of the gastric mucosa were taken from each subject according to the Sydney system, and a biopsy sample in the greater curvature, at least 3 cm away from each tumour, was used for the methylation analysis (Figs. 1a and 1b).²³ For this analysis, subjects were stratified by cancer status and *H. pylori* infection status [current (HP+), negative (HP–) or past infection (HPpast)] and were then matched based on their 10-year age group, sex and Lauren classification of GC. As a discovery set, 84 subjects (42 cases and 42 controls, with the case and control groups containing the same number of subjects for each *H. pylori* status) were selected for the methylation analysis. For the validation set, we selected subjects from the same study population as that used for the initial methylation analysis. To increase the sample size, we used a case-to-control ratio of 1:2. A total of 180 subjects (60 cases and 120 controls) were randomly selected, and pyrosequencing was performed for biological validation of the discovery set results. Technical confirmation was achieved using 25 subjects who were included in both the discovery and validation sets. Demographic, lifestyle and medical history data were collected from the participants via a self-administered questionnaire. The *H. pylori* infection status was determined by a rapid urease test, a serological test and histological evaluation. For the rapid urease test, one biopsy specimen was taken from the greater curvature of the corpus. For the histological evaluation, four biopsy specimens were collected from the lesser curvature of the corpus and antrum. The *H. pylori* status was determined via Wright–Giemsa

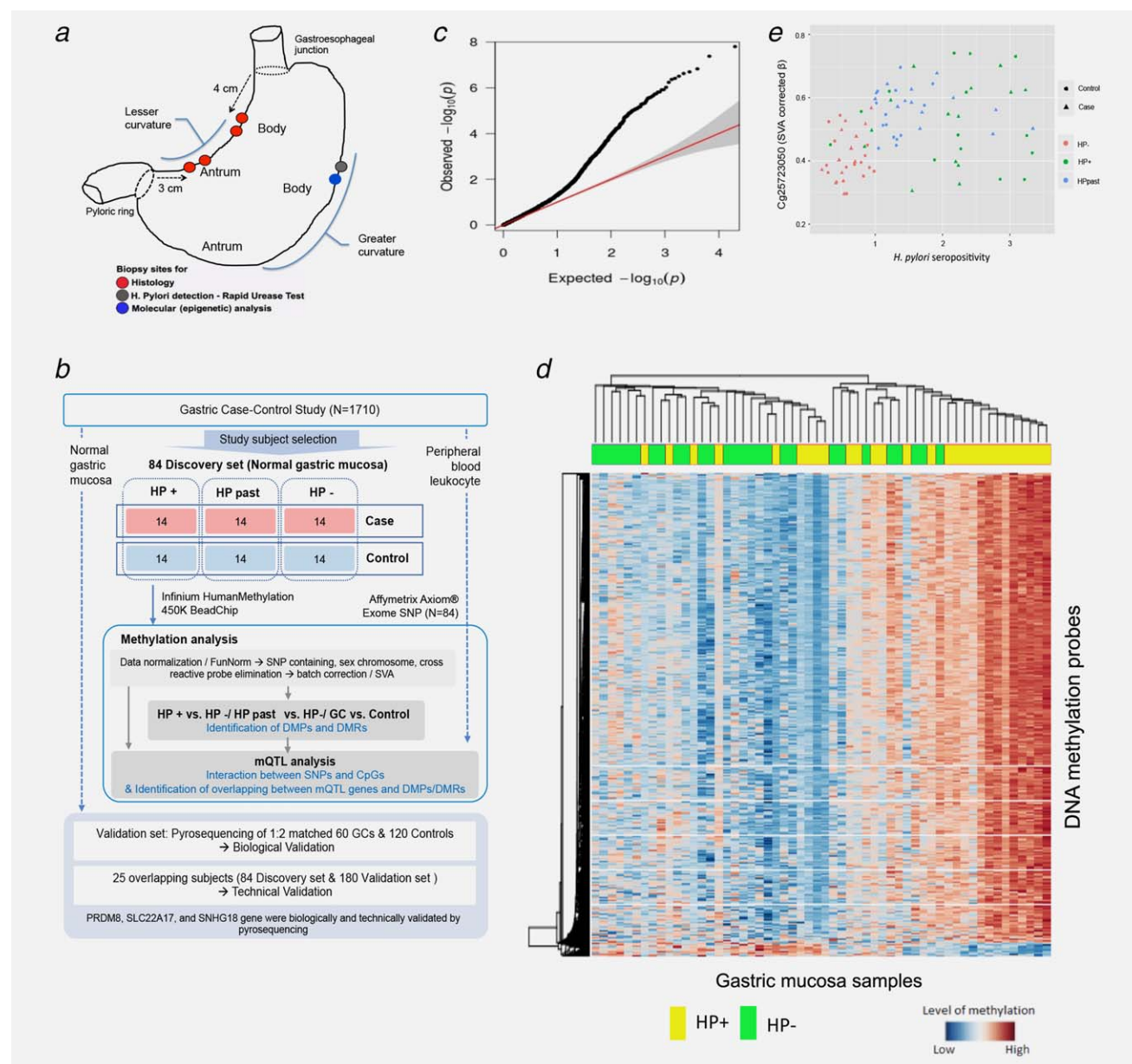


Figure 1. DNA methylome profiling of normal gastric mucosa by *H. pylori* infection status. (a) Gastric mucosa biopsy samples for molecular (epigenetic) analysis were obtained from the greater curvature of the gastric body (blue circle). (b) Flow chart illustrating the overall design of the study. Subjects were stratified by *H. pylori* infection status [current (HP+), negative (HP-), or past (past HP)] and cancer status (case or control) and were matched for age, sex and Lauren classification (for cases). (c) Quantile–quantile (Q–Q) plot of genome-wide DNA methylation between HP+ and HP-. The x-axis is the expected $-\log_{10} p$ values, and the y-axis is the observed $-\log_{10} p$ values. A total of 9,931 DMPs were identified without delta beta filtering ($\lambda = 1.27$). (d) Cluster heatmap analysis of DMPs with delta beta $\geq 20\%$. The rows represent probes for the 1,924 DMPs, and the columns represent individual HP+ and HP- samples. The cells are coloured according to the level of methylation. Among these DMPs, 52 CpG sites (2.7%) were hypomethylated, and 1,872 (97.3%) were hypermethylated. (e) The DMP with the lowest p values was cg25723050 (in the *CNIH3* gene). [Color figure can be viewed at wileyonlinelibrary.com]

staining of the biopsy specimens by a pathologist specialized in gastric cancer. Current infection was defined as at least one positive test result in the rapid urease test or histological evaluation of four biopsy sites. Past infection was defined as a positive result in only the serology test. Subjects who received eradication therapy were excluded. Therefore, past infections were naturally eradicated. Information on dietary intake was

obtained using a food frequency questionnaire, the reliability and validity of which have been previously reported.²⁴ Our study was approved by the Institutional Review Board at the National Cancer Centre (IRB number: NCCNCS-11-148) and by the Institutional Review Board at the International Agency for Research on Cancer (IARC, Lyon, France). All participants provided written informed consent.

DNA methylation analysis

DNA extracted from normal gastric mucosa was subjected to bisulphite conversion, and methylome profiling was performed using an Infinium Human Methylation 450 K BeadChip Kit (HM450 array; Illumina, San Diego, CA), which allows simultaneous interrogation of greater than 450,000 CpG sites, as described previously.^{25–27} The HM450 array data (IDAT files) were imported and processed using R/Bioconductor (<https://bioconductor.org/>). We excluded cross-reactive probes and probes for sites of known single-nucleotide polymorphisms (SNPs) with a minor allele frequency greater than 5%,²⁸ leaving a total of 452,162 probes for the analysis. The data were further normalized using the FunNorm function in the Bioconductor minfi package.^{29,30} The DNA methylation level β -values were logarithmically transformed to M-values to obtain data with a more normal distribution. Surrogate variable analysis (SVA) was performed on the methylome data to correct for potential batch effects and, especially, cell-type effects.³¹ To validate the findings obtained via methylome profiling, we used a pyrosequencing system (PSQ 96MA, Biotage, Uppsala, Sweden), as described previously.³²

We used a linear regression model after adjusting for covariates (smoking, alcohol consumption and folate consumption, as well as *H. pylori* seropositivity for the case-control comparison model) to identify differentially methylated positions (DMPs) using the Bioconductor limma package.³³ We used the Bioconductor DMRcate package with the recommended proximity-based criteria to identify differentially methylated regions (DMRs).³⁴ Statistically significant DMPs and DMRs were defined as those with a false discovery rate (FDR)-adjusted p values <0.05 . Identified DMRs were additionally filtered using cutoff values of maximum delta-beta $<5\%$ and mean delta-beta $<2\%$, with less than three CpGs constituting a DMR. The Bioconductor co-MET package was used to visualize the co-methylation patterns of several of the top DMRs.³⁵ Transcription factor binding site (TFBS) enrichment analysis was performed using the overlap of the identified DMP and ENCODE datasets, and transcription factor (TF) enrichment analysis was performed using the HOMER online tool (<http://homer.salk.edu/homer/motif/>) with a 50-bp window size.

Odds ratio (or) calculation and receiver operating characteristic curve construction

OR calculation and receiver operating characteristic (ROC) curve construction were performed using SAS 9.3 (SAS Institute, Cary, NC) and the PredictABEL R package, respectively.³⁶ We calculated ORs and their 95% confidence intervals (CIs) after cutting off the methylation levels of each gene by tertile based on the control population, and we also calculated the combined score of the three biomarker genes for GC risk. The two hypermethylated genes were given relative risk scores of 0, 1 and 2 with increasing tertiles of methylation levels, and the hypomethylated genes were given the opposite (decreasing) risk scores.

Methylation quantitative trait loci analysis

Genotypes were assessed using DNA from peripheral blood leukocytes and Axiom Exome319 Array Plates (Affymetrix, Santa Clara), which contain 318,983 SNPs. A quality-control procedure was used to eliminate poor-quality samples and SNP markers, and 96,252 markers passed this quality-control step. Genotype imputation was then performed, resulting in 713,348 markers in total.³⁷ To check for methylation sites that are potentially influenced by genetic variation, we aimed to identify methylation quantitative trait loci (mQTLs) to show the interactions between SNPs and methylation. Using the Matrix eQTL R package,³⁸ we examined the correlations within 5 kb upstream and 5 kb downstream of each polymorphism with a p values $<1e-5$. We excluded SNPs within 10 bp of the interrogated CpG sites to eliminate probe effects.

Results

DNA methylome profiling of normal gastric mucosa by *H. pylori* infection status

To investigate the epigenome-wide effects of *H. pylori* infection in gastric mucosa and to identify potential new biomarkers of GC risk, we performed genome-wide DNA methylation profiling of normal gastric mucosa from GC cases and controls (Figs. 1a and 1b). The general characteristics of the study participants included in the discovery and validation sets are summarized in Table 1. To investigate the stability/reversibility of aberrant DNA methylation changes after clearance of *H. pylori* infection, we also included samples of gastric mucosa from subjects with past (eradicated) *H. pylori* infection in the analysis (Table 1). Comparing the methylomes of the HP+ samples ($N=28$) with those of the HP– samples ($N=28$) revealed 9,931 DMPs ($\lambda = 1.27$) (Fig. 1c). Filtering these DMPs by a mean percentage difference in methylation (delta beta) $\geq 20\%$ yielded a final total of 1924 DMPs considered to be statistically significant (FDR <0.05) (Fig. 1d and Supporting Information Table S1). The example of the *CNIH3* gene in which highly significant DMPs associated with *H. pylori* were identified is shown in Figure 1e.

H. pylori-infected gastric mucosa exhibits marked genome-wide hypermethylation and differentially methylated regional clusters

Of the 1,924 filtered DMPs, 1,872 (97.3%) were hypermethylated (Fig. 1d and Supporting Information Table S1). Hypermethylated probes were highly enriched in promoter regions and were typically located close to (i.e., within 1 kb of) transcription start sites (TSS) (Figs. 2a and 2b). Hypermethylated DMPs were enriched in DNase I hypersensitive sites (DHSs) and enhancers (Fig. 2c). In addition, most hypermethylated probes were located within CpG islands, and this trend was more prominent when a higher delta beta cut-off point was applied (1,761 out of 1,872, 94% with delta-beta $\geq 20\%$) (Fig. 2d and Supporting Information Fig. S1). Overall, the methylation levels clearly differed by *H. pylori* status, and *H. pylori* infection

Table 1. General characteristics of the study subjects

	Discovery set (N = 84)			Validation set (N = 180)		
	Cases (n = 42)	Controls (n = 42)	p value	Cases (n = 60)	Controls (n = 120)	p value
<i>H. pylori</i> infection status, N						
Current	14	14		30	60	
Past (eradicated)	14	14		N/A	N/A	
Negative	14	14		30	60	
Lauren classification (cases only), N						
Intestinal	21	N/A		22	N/A	
Diffuse	18	N/A		28	N/A	
Mixed	3	N/A		8	N/A	
Missing	0	N/A		2	N/A	
Sex, N (%)						
Male	36 (85.7)	36 (85.7)		36 (60.0)	72 (60.0)	
Female	6 (14.3)	6 (14.3)		24 (40.0)	48 (40.0)	
Mean age \pm SD, years	52.0 \pm 11.1	52.2 \pm 9.5	0.924	53.5 \pm 9.74	53.6 \pm 8.45	0.920
Alcohol consumption status, N (%)						
Never	11 (26.2)	11 (26.2)	0.923	16 (26.7)	35 (29.2)	0.902
Former	3 (7.1)	4 (9.5)		3 (5.0)	7 (5.8)	
Current	28 (66.7)	27 (64.3)		41 (68.3)	78 (65.0)	
Smoking status, N (%)						
Never	11 (26.2)	16 (38.1)	0.075	28 (46.7)	65 (54.2)	0.094
Former	11 (26.2)	16 (38.1)		13 (21.7)	34 (28.3)	
Current	20 (47.6)	10 (23.8)		19 (31.7)	21 (17.5)	

N, Number; N/A, not applicable; SD, standard deviation.

appears to be strongly associated with hypermethylation across the methylome, especially in regulatory gene regions.

Pathway analyses using the genes located closest to the DMPs ($\Delta\beta \geq 20\%$) revealed enrichment in the neuroactive ligand-receptor interaction, cancer, axon guidance and Rap1 signalling pathways (Supporting Information Fig. S2). In addition, pathway analysis with DMP genes was performed after excluding the genes that are not expressed in the stomach (<https://www.proteinatlas.org/>) (Supporting Information Table S2). To further test whether the identified DMPs were enriched in TFBSs, we checked overlaps between the HP+ DMPs and the ENCODE datasets. The DMPs were significantly enriched in TFBSs compared with random probes or all probes in the 450 K array (Supporting Information Fig. S3). Thus, we performed a TF enrichment analysis to ascertain which specific TFs could underlie the methylation changes observed in our DMP list. Interestingly, the E2F3 motif was found to be present significantly more often in close proximity to the altered methylated sites, whereas this enrichment was not observed in the random probe set (Supporting Information Tables S3 and S4).

To identify highly correlated CpG regional clusters, we performed a DMR analysis that compared HP+ samples with HP- samples. DMR analysis is a type of dimension reduction that transforms single CpGs into regional clusters of highly

correlated CpGs. This analysis identified 438 DMRs (Supporting Information Table S5), suggesting the existence of regional clusters of differential methylation associated with *H. pylori* infection. To further visualize the correlations among CpGs within DMRs, we next plotted the top *H. pylori*-specific DMRs that mapped to different genomic loci (including the *NKX6-2*, *THBD*, *IGF2-AS*, *BARX2*, *NKX2-1*, *TRIL*, *SIM2* and *FOXQ1* genes) using the co-MET package. The co-MET panels showed a similar pattern of co-methylation among all top DMRs (Figs. 2e and 2f, Supporting Information Fig. S4, and data not shown). These findings suggest the coordinated hypermethylation of these loci. Interestingly, the co-MET panels revealed that all the TSS-proximal regions overlapped the hypermethylated CpG islands.

Persistence of *H. pylori*-induced aberrant DNA methylation in subjects with past *H. pylori* infection

Comparison of the HPpast (N = 28) with the HP- samples (N = 28) yielded 1086 DMPs (FDR < 0.05, $\Delta\beta \geq 10\%$) and 152 DMRs (FDR < 0.05) (Supporting Information Tables S6 and S7). The vast majority (98.4%) of the DMPs were hypermethylated, and the hypermethylated sites were enriched in DHSs, enhancers, and CpG island regions (Figs. 3a–3d). The most significant DMP annotated to the *CNIH3* gene. As was the case in the HP+ versus HP- analysis, the neuroactive ligand-receptor

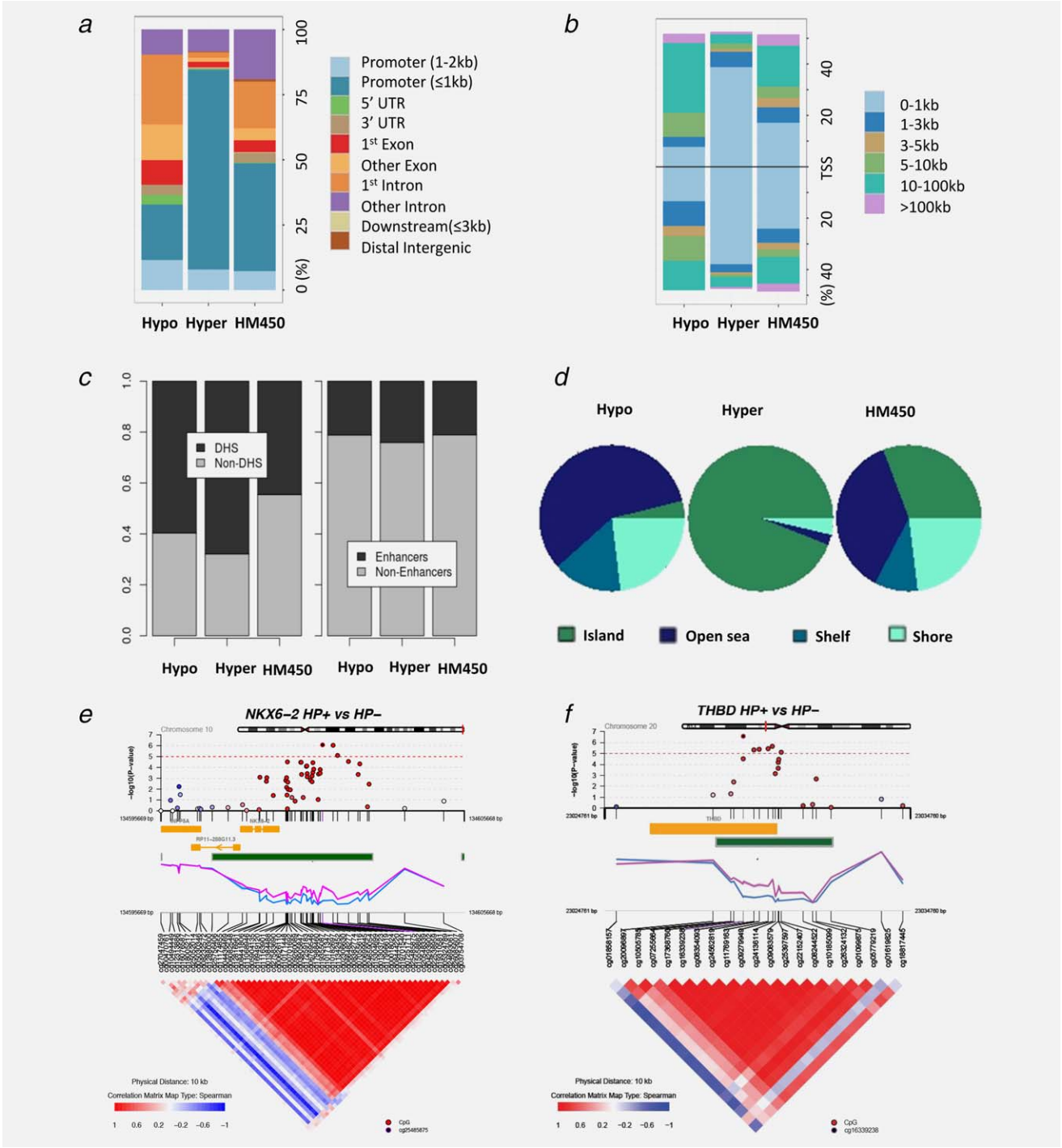


Figure 2. *H. pylori*-infected gastric mucosa exhibits marked genome-wide hypermethylation and differentially methylated regional clusters. (a–c) Hypermethylated probes were (a) enriched in promoter regions, (b) located short distances from TSSs, and (c) enriched in DHSs ($p < 2.2 \times 10^{-16}$) and enhancers ($p = 0.002$). The statistical significance was tested with binomial exact tests. (d) Most hypermethylated probes were located within CpG islands. (e and f) Methylation and co-methylation profiles of (e) *NKX6-2* and (f) *THBD*. The circles indicate the p values of the methylation differences between the HP+ and HP– groups per locus. The black circle represents the reference hit, and the other circles are coloured according to the Spearman correlation coefficients between the rest of the methylation values and the reference point, as indicated. The squares in the correlogram represent the pairwise correlations between CpG positions. The yellow bar represents the gene location and the green bar a CpG island. The blue and purple lines represent the methylation levels of the HP– and HP+ groups, respectively. All analyses were performed (and all graphical representations were generated) using SVA-corrected values with FDR < 0.05 and delta beta $> 20\%$. HM450, HM450 probes; Hyper, hypermethylated probes; Hypo, hypomethylated probes. [Color figure can be viewed at wileyonlinelibrary.com]

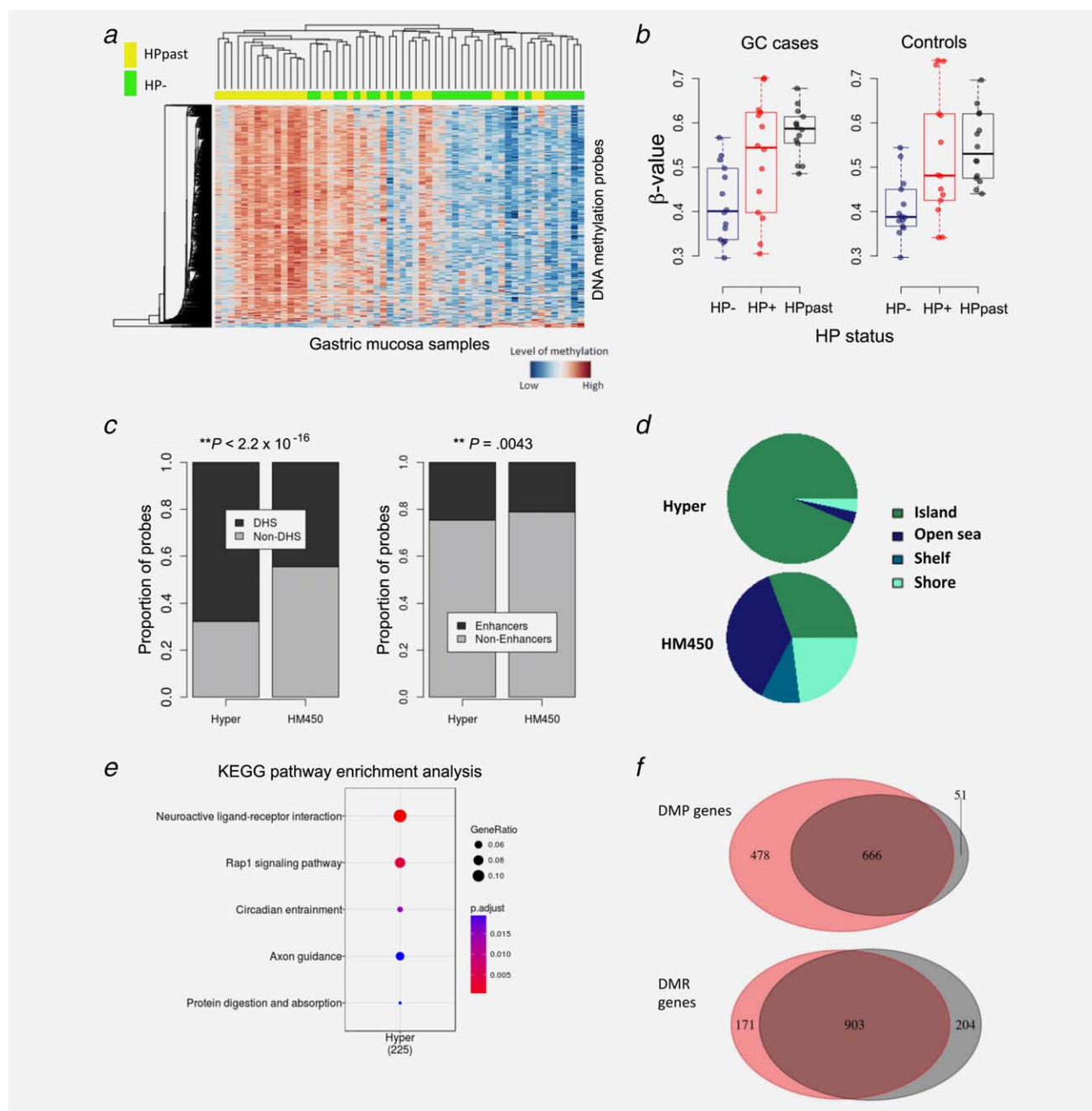


Figure 3. Methylation profiling of HPpast versus HP- samples of normal gastric mucosa. (a) Cluster heatmap analysis of DMPs (delta beta $\geq 10\%$). The rows represent probes for the 1,086 DMPs, and the columns represent individual HPpast and HP- samples. The cells are coloured according to the level of methylation. Among these DMPs, 17 CpG sites (1.6%) were hypomethylated, and 1,069 (98.4%) were hypermethylated. (b) The DMP with the lowest p values was cg25723050 (in the *CNIH3* gene). (c) Hypermethylated probes were enriched in DHSs and enhancers. Statistical significance was tested with binomial exact tests. (d) Most hypermethylated probes were located within CpG islands. (e) Pathway analyses were performed using the genes that were annotated closest to the identified DMPs (delta beta $\geq 20\%$) with the Kyoto Encyclopedia of Genes and Genomes (KEGG; <http://www.genome.jp/kegg/>). The top five pathways of the hypermethylated probes are shown. Hypomethylated probes showed no enrichment in any gene clusters. The gene ratio is the ratio between the number of DMP genes associated with a pathway and the number of all genes mapped to that pathway. (f) The red ovals represent HP+ DMP/R genes with delta beta $\geq 20\%$, and the grey ovals represent HPpast DMP/R genes with delta beta $\geq 10\%$. The similarity between the HPpast and HP+ methylomes was evidenced by a remarkable overlap between the findings of the two analyses (92.9 and 81.6% overlap of the HPpast DMPs and DMRs, respectively, with those of the HP+ group). All analyses were performed (and all graphical representations were generated) using SVA-corrected values with FDR < 0.05 and delta beta $> 10\%$. HM450, HM450 probes; Hyper, hypermethylated probes; p .adjust, FDR adjusted p values. [Color figure can be viewed at wileyonlinelibrary.com]

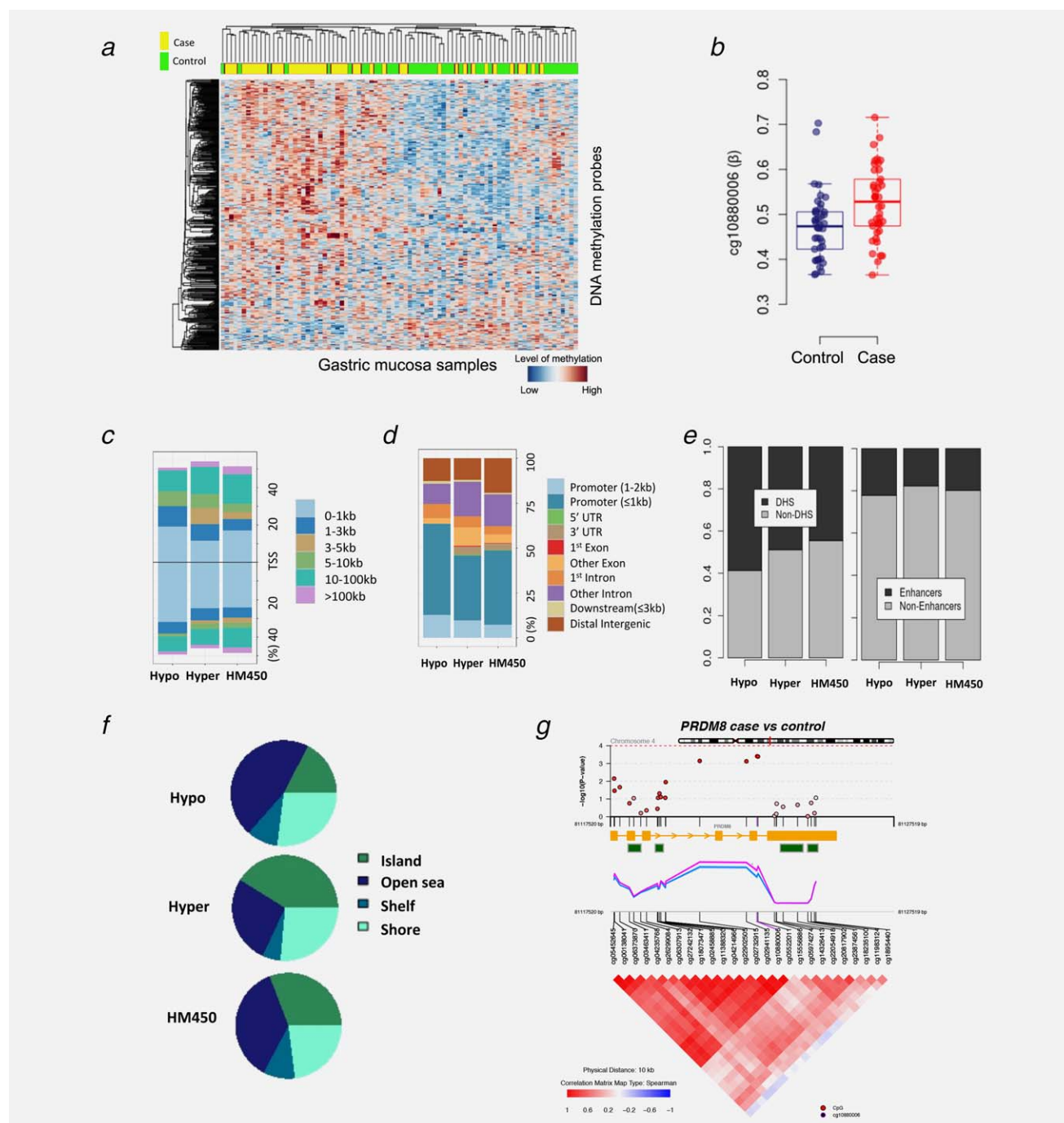


Figure 4. Methylation profiling of GC case samples versus control samples of normal gastric mucosa. (a) Cluster heatmap analysis of DMPs (delta beta $\geq 5\%$). The rows represent 601 CpG probes within DMRs, and the columns represent individual case and control samples. The cells are coloured according to the level of methylation. (b) The CpG site with the highest delta beta and lowest p values was cg10880006 (in the *PRDM8* gene). (c and d) CpG sites within DMRs were slightly enriched (but only in hypomethylated probes) within short distances from the TSS (c) and promoter regions (d) ($p = 0.023$). (e) DHSs were enriched in both hypomethylated probes ($p = 0.030$) and hypermethylated probes ($p = 0.046$). No significant enrichment in enhancer regions was found in these probes. (f) Distribution of CpG sites within DMRs. Most hypermethylated probes were located within CpG island regions ($p = 5.2 \times 10^{-7}$). (g) Co-methylation plot of *PRDM8*. All analyses were performed (and all graphical representations were generated) using SVA-corrected values with FDR < 0.05 and delta beta $> 5\%$. All the statistical tests were binomial, and their results were compared with the HM450 array data. HM450, HM450 probes; Hyper, hypermethylated probes; Hypo, hypomethylated probes. [Color figure can be viewed at wileyonlinelibrary.com]

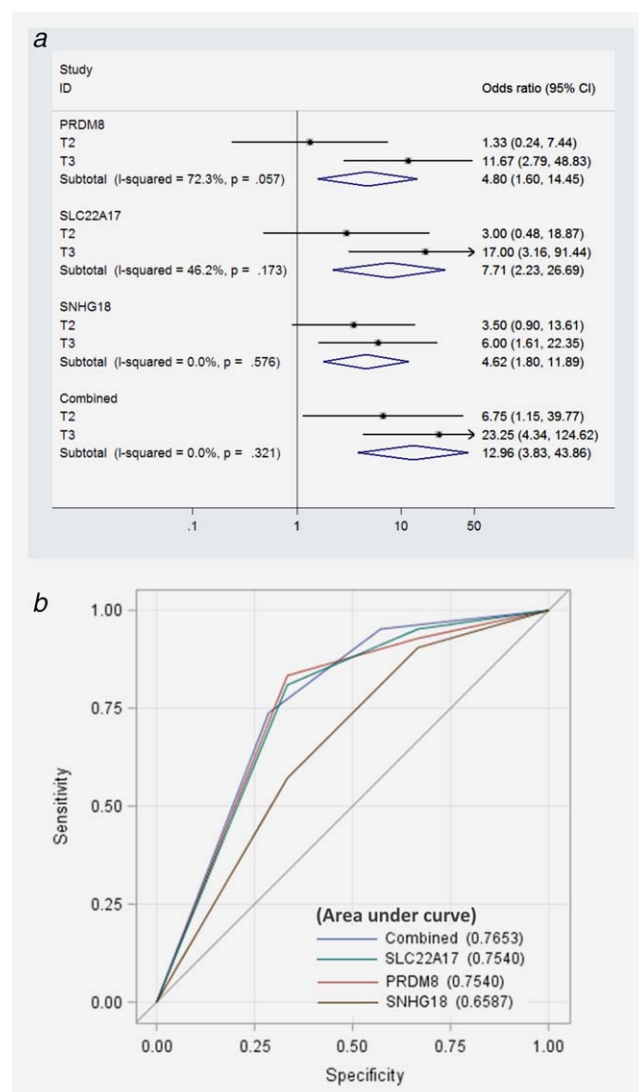


Figure 5. Putative biomarker genes for GC incidence and their combined methylation score. (a) Odds ratios and 95% CIs of GC by tertile (T) of methylation levels for three putative biomarker genes. (b) ROC curve for each gene and for their combined methylation score. [Color figure can be viewed at wileyonlinelibrary.com]

interaction and Rap1 signalling pathways were identified as the most significant (Fig. 3e). The similarity between the HPpast and HP+ methylomes was further evidenced by a remarkable overlap between the findings of the two analyses: 92.9 and 81.6% of the HPpast DMPs and DMRs, respectively, overlap with those of the HP+ group (Fig. 3f). These results show that *H. pylori*-associated epigenetic alterations persist after clearance of the infection (Fig. 3b), suggesting that *H. pylori* might leave an 'epigenetic memory' in gastric mucosa.

DNA methylome changes in normal gastric mucosa associated with case-control status

To investigate potential DNA methylation markers of GC risk independent of *H. pylori* status, we performed a methylome-wide analysis of the normal mucosa by GC case-control status.

No DMPs reached statistical significance after adjustment for multiple testing, likely because the target tissues analysed for GC status were non-cancerous in the first place. This finding suggests that the field effect of cancer on normal mucosa is minimal.

When the size of an effect is small, the statistical power must be increased to detect it, which can be achieved by increasing the sample size and/or reducing the dimension of the variables interrogated. We, therefore, performed a DMR analysis that decreased the matrix size of the HM450 array data, thus transforming a large number of individual CpGs into regional clusters of highly correlated CpG sites. Sixty-five DMRs (601 CpGs within DMRs, delta beta >5%) were identified (FDR < 0.05) (Figs. 4a–g and Supporting Information Table S8), among which the hypomethylated probes ($n = 65$) were slightly enriched in promoter regions, TSSs and DHSs (Figs. 4c–e), and the hypermethylated probes were significantly enriched in CpG island regions (Fig. 4f). Co-methylation was plotted using the highly significant CpG sites associated with case status that annotated to *PRDM8* (Figs. 4b and 4g). The resultant clusters are biologically interpretable because, in addition to being highly statistically correlated, the constituent CpG sites are also located in close genomic proximity and thus constitute regional (and not only statistical) clusters that function synchronously, as would be expected given their epigenetic nature.

To find suitable methylation biomarkers for GC, we identified the four genes with the highest delta-beta and lowest p values, namely, *PRDM8*, *SEPT5-GP1BB*, *SLC22A17* and *SNHG18* (Supporting Information Table S9), for validation by pyrosequencing. *SEPT5-GP1BB* was eliminated from the validation process due to its limited sequencing quality; the remaining three genes (two hypermethylated and one hypomethylated) (Supporting Information Table S10), showed similar DMR effect sizes as in the array-based approach and were therefore identified as potential biomarkers for GC risk (Fig. 5a and Supporting Information Table S11). We calculated ORs and their 95% CIs and found that their DNA methylation levels were highly associated with GC risk. The OR and 95% CI of the highest tertile of the combined score of the three genes was 23.25 (5.61–161.18) compared with the lowest tertile group. ROC curve analysis showed that the methylation score of *PRDM8*, *SLC22A17*, *SNHG18* and their combination might serve as biomarkers of GC, with area under the curve (AUC) values of 0.754, 0.754, 0.659 and 0.765, respectively (Fig. 5b). We validated the findings on these three genes using a larger sample set ($N = 180$, Table 1) and showed that their DNA methylation levels were highly associated with GC risk (Supporting Information Fig. S5). The methylation data of these genes were also technically confirmed using 25 randomly selected samples that were subjected to both 450 K array (84-sample discovery set) and pyrosequencing analyses (180-sample validation set) (data not shown). Therefore, although the overall DNA methylation patterns in normal gastric mucosa were not strikingly different in the cancer cases versus the controls, the normal gastric mucosa harbours specific biomarkers that can predict GC risk.

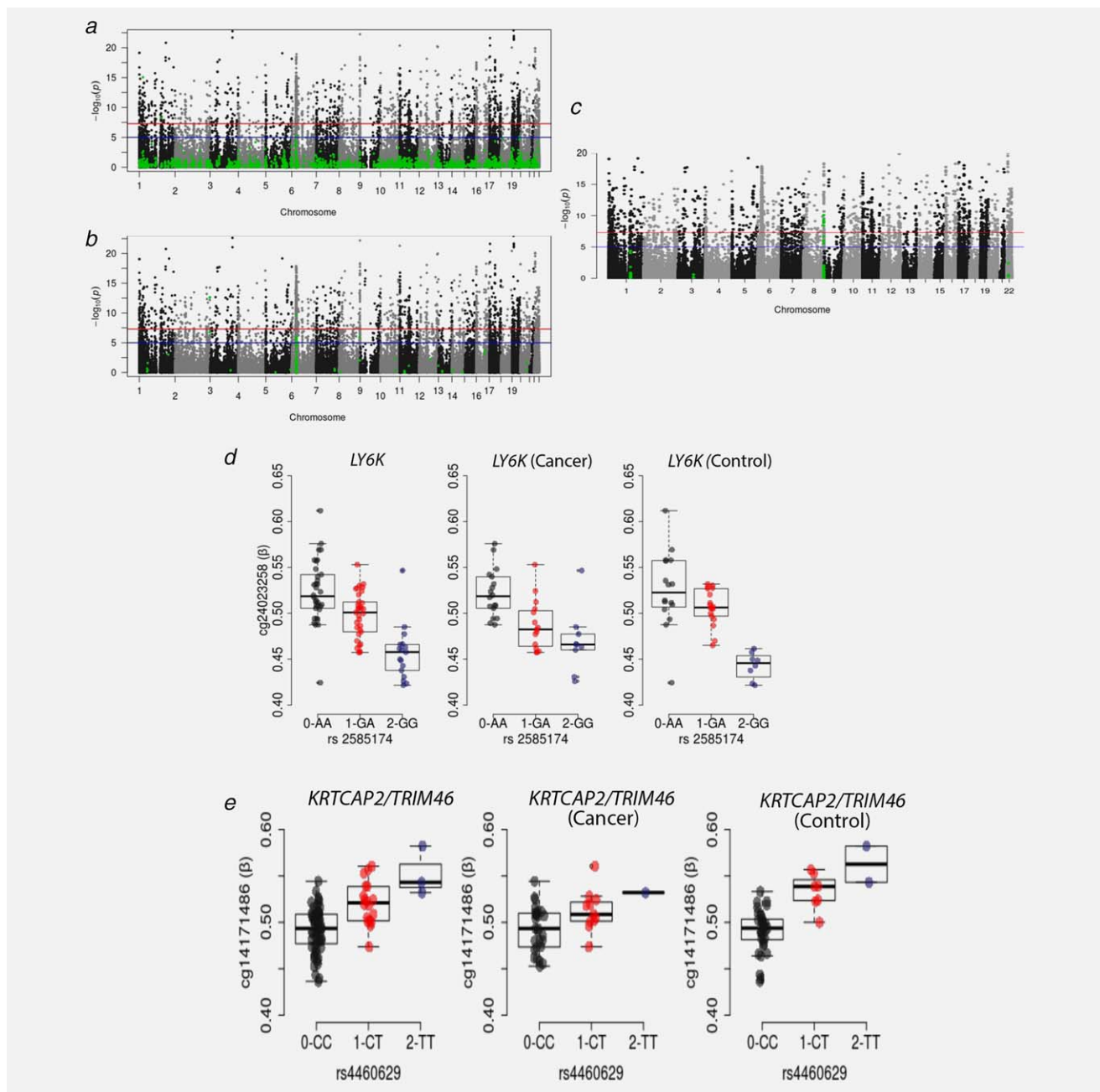


Figure 6. Manhattan plots of the p values of the mQTLs in normal gastric mucosa. (a–c) The mQTLs in each model were overlapped with differentially methylated sites by (a) *H. pylori* infection status ($\Delta\beta > 20\%$), (b) cancer status ($\Delta\beta > 5\%$), and (c) known GC-associated GWAS SNPs. The overlaps are indicated in green. (d and e) DNA methylation level stratified by the genotypes of *LY6K* and *KRTCAP2*, which overlapped with the genes. [Color figure can be viewed at wileyonlinelibrary.com]

mQTL analysis reveals a minimal impact of genetic variation on the *H. pylori*- and cancer-specific methylation changes

DNA methylation can be influenced by genetic variation, and specific methylation changes can be related to the SNP-associated risk of various diseases.³⁹ Therefore, we conducted an mQTL analysis to assess whether CpG methylation is influenced by nearby SNPs (i.e., SNPs located within a 10-kb window; cis-mQTL). SVA-corrected methylation

probes (combined with SNPs located within 5 kb on either side) with the same covariates in each DMP/DMR model identified a total of 2,325,652 pairs (formed by a combination of 337,218 SNPs and 216,590 CpG sites). The results were similar in the two models. A total of 101,247 pairs (4.3%; between 47,756 SNPs and 8,689 CpGs) and 101,397 pairs (4.3%; between 47,801 SNPs and 8,646 CpGs) were identified as mQTLs in the two models, respectively (p values $< 1e-5$).

The identified sites in the mQTL lists were matched to the DMPs in each model, but they overlapped significantly in only a few positions (Figs. 6a and 6b). The mQTL sites overlapped in three sites (two genes) with 20% delta beta HP+ DMPs, suggesting minimal genetic interference with *H. pylori*-associated methylation alterations. Cancer status-associated methylation changes also had some influence on genetic variation, but 32 sites (11 genes) in the 5% delta-beta CpG sites within DMRs overlapped with the mQTL sites, mainly in the *HLA-DPB2* gene regions. Interestingly, *PSCA*, which was shown to be associated with GC risk in a genome-wide association study (GWAS),⁴⁰ overlapped with the mQTL sites. To investigate whether the identified mQTLs were associated with the previously identified GWAS SNPs, we checked the overlaps between the GWAS and mQTL SNPs (Fig. 6c). We used previously published data to identify the GC-associated GWAS SNPs.^{40–43} Among the 66 known GWAS SNPs, 52 SNPs were available in our data, and the methylation levels of the CpGs located close to *LY6K* (lymphocyte antigen 6 family member K) and *KRTCAP2* (keratinocyte-associated protein 2) differed significantly by the genotype of the adjacent SNPs that had previously been identified to be associated with GC (Figs. 6d and 6e).

Discussion

In the present study, we characterized the genome-wide methylation profiles associated with *H. pylori* infection in a set of normal gastric mucosa originating from both GC patients and healthy individuals. A large number of DMPs (1,924) associated with *H. pylori* infection were identified, demonstrating the widespread deregulation of the methylome in normal gastric mucosa infected by *H. pylori*. A few previous studies have investigated the impact of *H. pylori* infection on DNA methylation in normal gastric mucosa.^{14–16} However, compared with these works, our study was not limited by sample size nor methylome coverage. Moreover, the present study is the first to apply dimension reduction approaches, which allowed identification of a large number of highly correlated CpG clusters that are associated with *H. pylori* infection.

The functional impact of the deregulated methylome induced by *H. pylori* infection remains to be established. Our findings that the majority of *H. pylori*-specific DMPs are strongly associated with hypermethylation across the genome and that the hypermethylated sites are highly enriched in the gene regulatory regions are consistent with the notion that the *H. pylori* might influence gene expression in gastric mucosa through aberrant gene silencing. This notion is further supported by our finding that hypermethylated sites associated with *H. pylori* are mostly enriched in CpG islands, which tend to occur predominantly in gene promoters. Furthermore, our findings that *H. pylori*-specific DMPs are also enriched in TF binding motifs reinforce the notion that *H. pylori* infection deregulates gene expression. Previous studies have shown that DNA methylation levels can be dynamically regulated at TFBSs within tissue-specific promoters during cell differentiation; therefore, the binding of TFs may play a role in modulating DNA methylation states at gene

promoters.⁴⁴ In this context, our finding that E2F3 motifs are enriched close to DMPs upon *H. pylori* infection suggests that this specific TF might play a role in methylation alteration. Additionally, E2F family TFs are known to be overexpressed in *H. pylori*-infected gastric mucosa.⁴⁵ Our results point to a novel working hypothesis according to which modulation of DNA methylation could be a downstream regulatory mechanism of E2F3 activation upon *H. pylori* infection.

Our results that *H. pylori*-specific DMPs are enriched in DHSs and enhancer chromatin marks, which tend to be occupied by cell-type-specific TFs,⁴⁶ suggest the existence of an alternative mechanism through which *H. pylori* might deregulate gene expression in gastric mucosa. The co-MET analysis revealed that aberrant co-methylation patterns are consistent across neighbouring CpG sites within individual gene loci, which agrees with the previous observation that aberrant methylation changes tend to occur in regional clusters instead of at isolated individual CpG sites. This finding also supports the notion that methylation changes at CpG-rich regions are likely to have an impact on gene expression regulation.⁴⁷ Furthermore, all TSS-proximal regions overlapped with hypermethylated CpG islands, consistent with the notion that aberrant hypermethylation located near the TSS could affect gene transcription.

Interestingly, our comparison of methylome data from gastric mucosa exhibiting current and past *H. pylori* infection revealed a strong overlap between the HPpast and HP+ methylomes, indicating that many of the genes targeted by *H. pylori* remain differentially methylated after *H. pylori* clearance. This result is consistent with the phenomenon known as ‘epigenetic memory’ that seems to be evident in subjects exposed to different environmental and lifestyle factors.^{27,48} Therefore, methylation changes may serve as an epigenetic memory system that ‘records’ and ‘transmits’ information from past exposures to environmental factors.⁴⁹ Several studies have revealed the long-lasting persistence of specific methylation changes associated with environmental exposures.^{27,50} This phenomenon might depend on multiple factors, including the genomic context (some CpGs remain aberrantly methylated for longer periods than others in response to the same risk exposure), the level and duration of exposure (infection), and the tissue composition (e.g., the level of *H. pylori*-associated differential methylation might reflect local tissue heterogeneity).

The finding that normal gastric mucosa harbours DMRs associated with case-control status suggests the presence of DNA methylation changes that can potentially serve as biomarkers of cancer risk. Despite the identification of DMRs, no statistically significant DMPs associated with case-control status were identified using the array data. However, our validation studies in a larger population did confirm the presence of differential methylation between cases and controls. This finding favours the possibility that normal gastric mucosa harbours DNA methylation changes highly associated with gastric cancer risk; however, further studies are needed to test if these

alterations constitute an increased susceptibility to GC or if they are merely markers of other processes leading to carcinogenesis.

Our mQTL analysis revealed that *H. pylori*-specific DMPs or CpG sites within GC-specific DMRs had a minor overlap with mQTLs, suggesting minimal genetic interference with *H. pylori*- or cancer-associated methylation alterations. However, probes within GC-specific DMRs mapped to *PSCA* overlapped with mQTLs, suggesting that methylation deregulation may be involved in *PSCA*-induced gastric carcinogenesis. Among 52 known GC-GWAS SNPs, the genotype of two SNPs close to the *LY6K* and *KRTCAP2* loci seemed to affect the methylation levels of adjacent CpG sites. *LY6K* and *KRTCAP2* are in the same linkage disequilibrium block with *PSCA*⁴⁰ and *MUC*,⁴³ respectively, which are believed to be strongly associated with GC risk. These results suggest that although the genetic influence on the methylation signatures described in the present work appears to be minimal, the previously identified GWAS-SNPs might act by modifying the methylation levels of proximal CpG sites.

Subjects with no signs of infection at the time of the study may have been infected before which may represent a limitation of our study. In addition, although the *H. pylori* infection status was classified according to the results of three tests (the rapid urease test and serological and histological evaluations), a small portion of gastric cancer patients in the HP- group may have had past or current *H. pylori* infections that were not detected due to false negative test results.

However, despite these weaknesses, our study design and validation are robust and our main findings are unlikely to be compromised by the limitations.

Conclusions

We comprehensively characterized the DNA methylome changes associated with current and past *H. pylori* infection in normal gastric mucosa from both GC patients and healthy individuals. We also identified DNA methylation markers of gastric cancer risk that are independent of the *H. pylori* status. These findings might have important implications for understanding the mechanisms through which *H. pylori* contributes to GC development as well as their potential reversibility in response to *H. pylori* clearance. Methylation-based biomarkers could be used for the follow-up of individuals at high risk for GC even after *H. pylori* eradication.

Availability of Data and Material

The dataset supporting the conclusions of this article is available in the GEO repository [GSE99553].

Acknowledgements

We thank Dr. Florence Le Calvez-Kelm and Mr. Geoffroy Durand for their assistance with the HM450 experiments and Dr. Hernandez Vargas for his advice regarding the data analyses. We thank IARC employees Elizabeth Page, who provided administrative assistance, and Jessica Cox, who provided technical editing services.

References

1. Ferlay J, Soerjomataram I, Dikshit R, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer* 2015;136:E359–86.
2. Okada E, Ukawa S, Nakamura K, et al. Demographic and lifestyle factors and survival among patients with esophageal and gastric cancer: the Biobank Japan Project. *J Epidemiol* 2017;27:S43–35.
3. Oh C-M, Won Y-J, Jung K-W, et al. Cancer statistics in Korea: incidence, mortality, survival, and prevalence in 2013. *Cancer Res Treat* 2016;48:436.
4. Heemskerk VH, Lentze F, Hulsewe KW, et al. Gastric carcinoma: review of the results of treatment in a community teaching hospital. *World J Surg Oncol* 2007;5:81.
5. Coglian VJ, Baan R, Straif K, et al. Preventable exposures associated with human cancers. *J Natl Cancer Inst* 2011;103:1827–39.
6. Parkin DM. The global health burden of infection-associated cancers in the year 2002. *Int J Cancer* 2006;118:3030–44.
7. Jones PA, Issa JP, Baylin S. Targeting the cancer epigenome for therapy. *Nat Rev Genet* 2016;17:630–41.
8. Hattori N, Ushijima T. Epigenetic impact of infection on carcinogenesis: mechanisms and applications. *Genome Med* 2016;8:10.
9. Niwa T, Tsukamoto T, Toyoda T, et al. Inflammatory processes triggered by *Helicobacter pylori* infection cause aberrant DNA methylation in gastric epithelial cells. *Cancer Res* 2010;70:1430–40.
10. Balassiano K, Lima S, Jenab M, et al. Aberrant DNA methylation of cancer-associated genes in gastric cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC-EURGAST). *Cancer Lett* 2011;311:85–95.
11. Liu Z, Zhang J, Gao Y, et al. Large-scale characterization of DNA methylation changes in human gastric carcinomas with and without metastasis. *Clin Cancer Res* 2014;20:4598–612.
12. Cheng Y, Yan Z, Liu Y, et al. Analysis of DNA methylation patterns associated with the gastric cancer genome. *Oncol Lett* 2014;7:1021–6.
13. Zouridis H, Deng N, Ivanova T, et al. Methylation subtypes and large-scale epigenetic alterations in gastric cancer. *Sci Transl Med* 2012;4:156ra140.
14. Shin CM, Kim N, Jung Y, et al. Role of *Helicobacter pylori* infection in aberrant DNA methylation along multistep gastric carcinogenesis. *Cancer Sci* 2010;101:1337–46.
15. Shin CM, Kim N, Jung Y, et al. Genome-wide DNA methylation profiles in noncancerous gastric mucosae with regard to *Helicobacter pylori* infection and the presence of gastric cancer. *Helicobacter* 2011;16:179–88.
16. Zhang Y, Zhang XR, Park JL, et al. Genome-wide DNA methylation profiles altered by *Helicobacter pylori* in gastric mucosa and blood leukocyte DNA. *Oncotarget* 2016;7:37132–44.
17. Asada K, Nakajima T, Shimazu T, et al. Demonstration of the usefulness of epigenetic cancer risk prediction by a multicentre prospective cohort study. *Gut* 2015;64:388–96.
18. Maeda M, Nakajima T, Oda I, 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.
19. Graham DY. *Helicobacter pylori* update: gastric cancer, reliable therapy, and possible benefits. *Gastroenterology* 2015;148:719–31.
20. Wroblewski LE, Peek RM, Jr, Wilson KT. *Helicobacter pylori* and gastric cancer: factors that modulate disease risk. *Clin Microbiol Rev* 2010;23:713–39.
21. Gaston D, Hansford S, Oliveira C, et al. Germline mutations in MAP3K6 are associated with familial gastric cancer. *PLoS Genet* 2014;10:e1004669.
22. Oliveira C, Senz J, Kaurah P, et al. Germline CDH1 deletions in hereditary diffuse gastric cancer families. *Hum Mol Genet* 2009;18:1545–55.
23. Dixon MF, Genta RM, Yardley JH, Classification and grading of gastritis, et al. The updated Sydney System. International Workshop on the Histopathology of Gastritis, Houston 1994. *Am J Surg Pathol* 1996;20:1161–81.
24. Ahn Y, Kwon E, Shim JE, et al. Validation and reproducibility of food frequency questionnaire for Korean genome epidemiologic study. *Eur J Clin Nutr* 2007;61:1435–41.
25. Hernandez-Vargas H, Castelinio J, Silver JM, et al. Exposure to aflatoxin B1 in utero is associated with DNA methylation in white blood cells of

- infants in The Gambia. *Int J Epidemiol* 2015;44:1238–48.
26. Martin M, Ancey PB, Cros MP, et al. Dynamic imbalance between cancer cell subpopulations induced by transforming growth factor beta (TGF-beta) is associated with a DNA methylome switch. *BMC Genom* 2014;15:435.
 27. Ambatipudi S, Cuenin C, Hernandez-Vargas H, et al. Tobacco smoking-associated genome-wide DNA methylation changes in the EPIC study. *Epigenomics* 2016;8:599–618.
 28. Chen YA, Lemire M, Choufani S, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics* 2013;8:203–9.
 29. Fortin J-P, Labbe A, Lemire M, et al. Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome Biol* 2014;15:503.
 30. Aryee MJ, Jaffe AE, Corrada-Bravo H, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* 2014;30:1363–9.
 31. McGregor K, Bernatsky S, Colmegna I, et al. An evaluation of methods correcting for cell-type heterogeneity in DNA methylation studies. *Genome Biol* 2016;17:84.
 32. Vaissiere T, Hung RJ, Zaridze D, et al. Quantitative analysis of DNA methylation profiles in lung cancer identifies aberrant DNA methylation of specific genes and its association with gender and cancer risk factors. *Cancer Res* 2009;69:243–52.
 33. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015;43:e47.
 34. Peters JT, Buckley JM, Statham LA, Pidsley R, Samaras K, Lord R, Clark JS, Molloy LP. De novo identification of differentially methylated regions in the human genome. *Epigenet Chrom* 2015;8:6.
 35. Martin TC, Yet I, Tsai P-C, et al. coMET: visualisation of regional epigenome-wide association scan results and DNA co-methylation patterns. *BMC Bioinform* 2015;16:131.
 36. Kundu S, Aulchenko YS, Van Duijn CM, et al. PredictABEL: an R package for the assessment of risk prediction models. *Eur J Epidemiol* 2011;26:261.
 37. Yang S, Park Y, Lee J, Choi JJ, Kim YW, Ryu KW, Sung J, Kim J. Effects of soy product intake and interleukin genetic polymorphisms on early gastric cancer risk in Korea: a case-control study. *Cancer Res Treat* 2017;49:1044–56.
 38. Shabalín AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* 2012;28:1353–8.
 39. Bell JT, Spector TD. A twin approach to unraveling epigenetics. *Trends Genet* 2011;27:116–25.
 40. Sakamoto H, Yoshimura K, Saeki N, et al. Genetic variation in PSCA is associated with susceptibility to diffuse-type gastric cancer. *Nat Genet* 2008;40:730–40.
 41. Shi Y, Hu Z, Wu C, et al. A genome-wide association study identifies new susceptibility loci for non-cardia gastric cancer at 3q13.31 and 5p13.1. *Nat Genet* 2011;43:1215–8.
 42. Abnet CC, Freedman ND, Hu N, et al. A shared susceptibility locus in PLCE1 at 10q23 for gastric adenocarcinoma and esophageal squamous cell carcinoma. *Nat Genet* 2010;42:764–7.
 43. Hu N, Wang Z, Song X, et al. Genome-wide association study of gastric adenocarcinoma in Asia: a comparison of associations between cardia and non-cardia tumours. *Gut* 2016;65:1611–8.
 44. Marchal C, Miotto B. Emerging concept in DNA methylation: role of transcription factors in shaping DNA methylation patterns. *J Cell Physiol* 2015;230:743–51.
 45. Isomoto H, Furusu H, Shin M, et al. Enhanced expression of transcription factor E2F in *Helicobacter pylori*-infected gastric mucosa. *Helicobacter* 2002;7:152–62.
 46. Stadler MB, Murr R, Burger L, et al. DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* 2011;480:490–5.
 47. Weber M, Davies JJ, Wittig D, et al. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat Genet* 2005;37:853–62.
 48. Guida F, Sandanger TM, Castagné R, et al. Dynamics of smoking-induced genome-wide methylation changes with time since smoking cessation. *Hum Mol Genet* 2015;24:2349–59.
 49. Hecceg Z, Vaissiere T. Epigenetic mechanisms and cancer: an interface between the environment and the genome. *Epigenetics* 2011;6:804–19.
 50. Vineis P, Chatziioannou A, Cunliffe VT, et al. Epigenetic memory in response to environmental stressors. *FASEB J* 2017;31:2241–51.