Classification of Epstein-Barr virus positive gastric cancers by definition of DNA methylation epigenotypes

Keisuke Matsusaka¹,², Atsushi Kaneda¹,⁵*, Genta Nagae¹, Tetsuo Ushiku², Yasuko Kikuchi¹,³, Rumi Hino², Hiroshi Uozaki², Yasuyuki Seto³, Kenzo Takada⁴, Hiroyuki Aburatani¹, Masashi Fukayama²

Author affiliation

¹Genome Science Division, Research Center for Advanced Science and Technology (RCAST), ²Department of Pathology and ³Department of Gastrointestinal Surgery, Metabolic Care and Endocrine Surgery, Graduate School of Medicine, The University of Tokyo, Japan; ⁴Department of Tumor Virology, Institute for Genetic Medicine, Hokkaido University, Japan; ⁵PRESTO, Japan Science and Technology Agency (JST)

Running Title: Methylation in EBV(+) gastric cancer

Financial support

Grants-in-Aids for Scientific Research B 22390255 (AK), S 20221009 (HA), and A 20249022 (MF) from Ministry of Education, Culture, Sports, Science and Technology of Japan; JST PRESTO program (AK).

*Corresponding author: Atsushi Kaneda

Genome Science Division, RCAST, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8904, Japan
Disclosure of Potential Conflicts of Interest

The authors have no conflict of interest.

Word count: 4,980 words

Number of figures/tables: 5 figures and 1 table
Abstract

Epstein-Barr virus (EBV) is associated with Burkitt's lymphoma, nasopharyngeal carcinoma, opportunistic lymphomas in immunocompromised hosts, and a fraction of gastric cancers. Aberrant promoter methylation accompanies human gastric carcinogenesis, though the contribution of EBV to such somatic methylation changes has not been fully clarified. We analyzed promoter methylation in gastric cancer cases using illumina's Infinium beadarray; hierarchical clustering analysis classified gastric cancers into three subgroups: EBV(-)/low methylation, EBV(-)/high methylation, and EBV(+)/high methylation. The three epigenotypes were characterized by three groups of genes: genes methylated specifically in the EBV(+) tumors (e.g. CXXC4, TIMP2, PLXND1), genes methylated both in EBV(+) and EBV(-)/high tumors (e.g. COL9A2, EYA1, ZNF365), and genes methylated in all of the gastric cancers (e.g. AMPH, SORCS3, AJAP1). Polycomb repressive complex (PRC)-target genes in embryonic stem cells were significantly enriched among EBV(-)/high methylation genes and commonly methylated gastric cancer genes (P=2x10^{-15} and 2x10^{-34}, respectively), but not among EBV(+) tumor-specific methylation genes (P=0.2), suggesting a different cause for EBV(+)-associated de novo methylation. When recombinant EBV was introduced into the EBV(-)/low methylation epigenotype gastric cancer cell, MKN7, three independently established subclones displayed increases in DNA methylation. The promoters targeted by methylation were mostly shared among the three subclones, and the new methylation changes caused gene repression. In summary, DNA methylation profiling classified gastric cancer into three epigenotypes, and EBV(+) gastric cancers showed distinct methylation patterns likely attributable to EBV infection.
Introduction

Gastric cancer is the second leading cause of cancer-related deaths in the world (1). It has been reported that two pathogens, *Helicobacter pylori* (*H. pylori*) and Epstein-Barr virus (EBV), participates in gastric cancer development (2, 3). *H. pylori* is a gram-negative spiral-shaped bacterium, infected in the stomach of about half of the world’s population (4, 5), and the recent prospective cohort studies have demonstrated that *H. pylori* infection played an essential role in gastric carcinogenesis (6). The other pathogen, EBV, belongs to the *Herpesviridae* causing infectious mononucleosis in initial infection, and more than 90% adult people finally become EBV carriers (7). EBV is involved in several kinds of malignant tumors, such as Burkitt’s lymphoma (8), nasopharyngeal carcinoma (9), and opportunistic lymphoma in immunocompromised host (10, 11). EBV-positive (EBV(+)) gastric cancer was discovered in 1990 (3), and proved to distribute all over the world without regional or racial differences at the rate of 7-15% (12, 13).

During multistep carcinogenesis, gastric cancer is thought to arise through accumulation of epigenetic alterations and genetic alterations including *p53* and *CDH1* mutation (14-16). Aberrant DNA methylation of gene promoter region is one of the most important epigenetic alterations resulting in gene silencing, and gastric cancer is one of tumors exhibiting high frequency of aberrant promoter methylation e.g. in *CDKN1A/p16*, *RUNX3*, and *LOX* (17, 18). It was reported that aberrant methylation in *MLH1* had close correlation with microsatellite instability in gastric cancer similar to colorectal cancer (19, 20).

Kang et al. reported that EBV(+) gastric cancer was involved in aberrant promoter methylation more frequently than EBV-negative (EBV(-)) gastric cancer (21). While frequency of *MLH1* methylation was significantly less in EBV(+) than EBV(-) gastric cancer (21-23), *TP73* and *HOXA10* were reported to be methylated more in EBV(+) gastric cancer (24, 25). The number of
analyzed genes, however, has been limited to known cancer-associated genes in previous reports of EBV(+) gastric cancer. It was also reported that an EBV(+) gastric cancer tissue had monoclonal EBV genome in each cancer cell as an episomal latent form without integration into the host genome (26, 27). The monoclonality suggested that EBV infection might occur at a very early stage of carcinogenesis and that EBV infection might play a role in cancer development, including induction of the specific aberrant DNA methylation phenotype if any.

To clarify methylation epigenotypes specific to EBV(+) gastric cancer, we analyzed DNA methylation status of promoter regions in 51 clinical gastric cancers, including 11 EBV(+) cases, using Infinium HumanMethylation27 BeadChip (Infinium, illumina) for 27,578 individual CpG dinucleotides, which covers 14,495 gene promoter regions. Gastric cancer was clustered into three epigenotypes by unsupervised two-way hierarchical clustering: low- and high-methylation epigenotypes in EBV(-) cancer, and EBV(+) specific markedly higher methylation epigenotype. Using Akata system of recombinant EBV (rEBV) infection into monolayer cells in vitro (28), EBV infection per se was shown to induce the EBV(+) specific methylation epigenotype in EBV(-) gastric cancer cell.
Materials and Methods

Detailed information of Materials and Methods is described in Supplementary Information.

Clinical samples and cell lines. Primary gastric cancer samples were obtained from patients undergoing gastrectomy at Department of Gastrointestinal Surgery, University of Tokyo, with written informed consents. These samples were immediately frozen with liquid nitrogen and stored at -80°C. The frozen materials were microscopically examined for determination of cancer cell contents by two independent pathologists, and were dissected to enrich cancer cells when necessary. Fifty-one samples containing >40% of cancer cells were used. DNA was extracted by QIAquick DNA mini kit (QIAGEN, Valencia, CA). Gastric cancer cell line MKN7 was obtained from Riken BioResource Center Cell Bank (Tsukuba, Japan), AGS from ATCC (Manassas, VA), and SNU719 from Korean Cell Line Bank (Seoul, Korea). The cell lines were authenticated by the cell banks using short tandem repeat PCR. A xenograft tumor, KT, was previously established (29). DNA of normal gastric mucosa (NGM) and peripheral blood cell (PBC) were purchased by The Coriell Cell Repositories (Camden, NJ). This study was certified by Ethics Committee in the University of Tokyo.

Infinium assays. Infinium (illumina, San Diego, CA) contains 27,578 CpG sites covering 14,495 genes. In each CpG site, ratio of fluorescent signal is measured by a methylated probe relative to sum of methylated and unmethylated probes, so-called β-value. The β-value, 0.00–1.00, reflects the methylation level of individual CpG site. Bisulfite conversion, whole genome amplification, labeling, hybridization and scanning were performed according to the manufacturer’s protocols. According to the previously proposed classification (30), Infinium probes were classified into three categories, high-CpG, intermediate-CpG, and low-CpG probes, based on CpG ratio and GC content within 500-bp region (from -250 bp to +249 bp) around the probe site. Genes in X and Y
chromosomes were excluded to avoid gender difference, so the whole genes analyzed were 13,897. Infinium data is available at GEO datasets (GSE31789).

Pyrosequencing analysis. Quantitative validation for methylation was carried out by pyrosequencing as described previously (31). Primers were designed using Pyro Q-CpG Software (QIAGEN), to amplify bisulfite-treated DNA region containing several CpG sites. Primer sequences are listed in Supplementary Table S1. Methylation control samples (0%, 25%, 50%, 75% and 100%) were prepared as previously described (32), and used to check quantitativity of pyrosequencing assays (Supplementary Fig. S1).

Gene Ontology analysis. Gene annotation enrichment analysis was done for Gene Ontology (biological process, cellular component, and molecular function) using the Functional Annotation tool at DAVID Bioinformatics Resources (33).

Recombinant EBV infection (Akata System). The floating cell line, Akata, derived from EBV(+) Burkitt’s lymphoma was previously modified to produce recombinant EBV (rEBV), in which the neomycin resistance gene was inserted into BXLF1, and was used as a virus donor cell line (28). The rEBV(+) Akata cells were replaced into serum-free medium and stimulated by serum IgG cross-linking to switch from latent to lytic infection. After removal of serum IgG by washing twice with PBS, Akata cells were added to the dish of monolayer gastric cancer cell line, MKN7 or AGS, to co-culture for 3 days. The dish was thoroughly washed four times with PBS on day 3 to remove floating Akata cells. The rEBV-infected gastric cancer cells were then selected with geneticin. Three rEBV(+) clones were established independently in different dishes to avoid selecting the same clones.

In Situ Hybridization. To determine EBV infection status, in situ hybridization targeting EBER (EBER-ISH) was performed as reported (27), using formalin-fixed paraffin-embedded
specimens in all clinical samples and blocks of cell lines.

**Immunohistochemistry.** Formalin-fixed paraffin-embedded cell block was analyzed by immunohistochemistry using anti-Cytokeratin antibody (#M3515, DAKO Japan, Tokyo, Japan) as epithelial cell marker and anti-CD45 antibody (#M0701, DAKO) as leukocyte marker.

**Plasmid construction and transfection.** LMP2A cDNA was a kind gift from Prof. Paul J. Farrell, Imperial College, UK. As for small RNAs, EBER1 and EBER2, pcDNA3 containing 10 tandem repeats of EBER1 and EBER2 open reading frames was previously established (34). MKN7 was transfected with pcDNA3 or pcDNA5/TO vector (Invitrogen) containing cDNA of LMP2A, EBNA1, EBER1/2 and BARF0 or mock vector, using lipofectamine 2000 (Invitrogen), and selected with 200 µg/mL geneticin or hygromycin B. Cells were cultured for 6 weeks after transfection, and collected for extraction of genomic DNA, total RNA, and protein.

**Expression microarray analysis.** mRNA expression in gastric cancer cell lines was analyzed using Affymetrix GeneChip Human Genome U133 plus 2.0 oligonucleotide arrays (Fremont, CA). For global normalization, the average signal in an array was made equal to 100. Expression array data is available at GEO datasets (GSE31789).

**RT-PCR.** RT-PCR analysis was performed for EBV latent genes (LMP2A, EBNA1, EBER1, EBER2 and BARF0) and ACTB. Ten microliters of each PCR product were separated on 2% agarose gels and stained with ethidium bromide. Primer sequences are listed in Supplementary Table S2.

**Quantitative Real-Time PCR.** Real-time PCR analysis using iCycler Thermal Cycler (Bio-Rad Laboratories) was performed as previously described (18). For real-time RT-PCR, the quantity of mRNA of each gene was normalized to that of ACTB. For PCR of genomic DNA, the quantity of BZLF1 region in EBV genome was normalized to that of CRP region in host genome (35). Primer sequences are listed in Supplementary Tables S2 and S3.
**Immunoblot analysis.** Aliquots of protein were subjected to SDS/PAGE followed by immunoblot analysis using antibodies against DNMT1 (N-16, Santa Cruz) and Actin (I-19, Santa Cruz), and the resulting immunoblots were visualized using Amersham ECL Plus (GE Healthcare).

**Statistical analysis.** Statistical analyses were performed using $\chi^2$ test, $t$-test, ANOVA, or Kolmogorov-Smirnov test. Unsupervised two-way hierarchical clustering was performed based on correlation distance and average linkage clustering algorithm in sample directions and probe directions using Cluster 3.0 software (36). Heat map was drawn by Java TreeView software (37).
Result

**Probe selection in Infinium analysis.** To analyze Infinium data, the following criteria were set up; (a) probe with the highest CpG ratio was selected for one promoter when multiple probes were designed; (b) standard deviation of β-value concerning 51 gastric cancer and 2 NGM samples was calculated for each probe, and probes with standard deviation <0.05 were excluded for hierarchical clustering; (c) genes with intermediate- and high-CpG probes were used in hierarchical clustering to analyze promoter with high CpG density e.g. CpG island, and genes with low-CpG probe were separately shown in heatmap. Through (a)-(c), 4,249 probes were selected (Fig. 1).

**Unsupervised two-way hierarchical clustering.** Unsupervised hierarchical clustering analysis of 51 clinical gastric cancer and 2 NGM samples classified gastric cancer into three subgroups: low- and high-methylation epigenotypes (Low- and High-epigenotype hereafter) in EBV(-) cases, and markedly higher methylation epigenotype that was completely matched to EBV(+) cases (EBV(+)-epigenotype hereafter) (P=3.0×10^{-6}, Fig. 1, Table 1). Two NGM samples were less methylated and clustered in a fourth subgroup together with two normal PBC samples, and the PBC samples showed different methylation pattern in low-CpG probes compared to other 53 stomach samples.

The EBV(+)-epigenotype had the following features (Fig. 1, Fig. 2, Table 1). (a) EBV(+)-epigenotype had ~270 excessively methylated genes which were not methylated in Low- nor High-epigenotypes. (b) Genes methylated in EBV(-) gastric cancers were mostly methylated in EBV(+) cases, except MLH1. That is, methylation specific to EBV(-) cancer was hardly detected. (c) Whereas High-epigenotype showed frequent methylation of MLH1 (46%), EBV(+)-epigenotype did not show MLH1 methylation at all (P=0.037, Table 1). (d) PRC-target genes reported in ES cell (38) were enriched in genes methylated in Low- and High-epigenotypes, but not in genes methylated
specifically in EBV(+) epigenotype (Fig. 1, Fig. 2).

Clinicopathological information was summarized in Table 1, and EBER-ISH results of clinical samples were shown in Table 1 and representatively in Fig. 3A.

The methylation profiles of 3 gastric cancer cell lines, MKN7, AGS, and SNU719, and one xenograft tumor, KT, were also analyzed by Infinium (Fig. 1). SNU719 and KT, derived from EBV(+) gastric cancer, showed markedly higher methylation epigenotype similar to EBV(+) clinical gastric cancer samples, indicating that EBV(+) epigenotype was not epigenotype due to contamination of infiltrated blood cells but due to EBV(+) gastric cancer cells themselves. Interestingly, though AGS was not EBV(+) cell (Fig. 3B, Supplementary Fig. S2), AGS was considered to be EBV(+) epigenotype because its markedly higher methylation pattern was similar to SNU719 and KT (Fig. 1, Fig. 2A).

**Extraction of marker genes for three epigenotypes.** Three groups of marker genes characterizing three epigenotypes were extracted (See detailed criteria in Supplementary Table S4): 72 Common-marker genes, methylated commonly in Low-, High-, and EBV(+) epigenotypes but unmethylated in NGM; 79 High-marker genes, methylated in EBV(+) and High-epigenotypes but unmethylated in Low-epigenotype and NGM; 53 EBV(+) marker genes, methylated in EBV(+) epigenotype specifically (Fig. 2A, Supplementary Tables S4-S5).

Whereas the ratio of PRC target genes among the whole 13,897 genes was 8.7% (1,211/13,897), PRC target genes were significantly enriched in Common- and High-marker genes, 50.0% (36/72, \(P=2\times10^{-34}\)) and 34.2% (27/79, \(P=2\times10^{-15}\)), respectively (Fig. 2A-B). PRC target genes in EBV(+) epigenotype were, however, only 15.1% and not significantly enriched at all (8/53 genes, \(P=0.2\)), suggesting a possible methylation mechanism different from Common- and High-marker genes (Fig. 2B).
Validation by pyrosequencing. Methylation status obtained by Infinium was quantitatively validated by pyrosequencing. Nine marker genes were randomly chosen to include three markers each for Common-, High- and EBV(+) marker groups (Fig. 2C, Supplementary Table S1). While each Infinium probe determines methylation status at single CpG site, pyrosequencing data not only validated the methylation at probe site, but also guaranteed the methylation status of surrounding multiple CpG sites as a DNA region.

rEBV infection experiment. To analyze whether EBV infection per se could induce EBV(+) epigenotype, rEBV was infected into two gastric cancer cell lines, MKN7 and AGS, using Akata system. While AGS was considered to be EBV(+) epigenotype, MKN7 was considered to be Low-epigenotype because High- and EBV(+) markers were generally unmethylated and Common-markers were methylated in MKN7. It was confirmed that the three rEBV-infected clones obtained were not contaminated with Akata cells (Fig. 3C, Supplementary Fig. S3). These clones (MKN7_EB#1, #2, #3, and AGS_EB#1, #2, #3) were compared with the parent cells (MKN7_p and AGS_p) and mock cells cultured for the same period of 18 weeks without rEBV infection (MKN7_EB(-) and AGS_EB(-)). EBV existence in rEBV-infected clones was confirmed by EBER-ISH (Fig. 3A, Supplementary Fig. S2), and by real-time PCR for EBV genome (Fig. 3D). Expression of EBV latent genes in the three clones was confirmed by real-time RT-PCR (Fig. 3E).

DNA methylation profiles of the clones with/without rEBV infection were analyzed using Infinium. All the MKN7_EB#1, #2 and #3 showed marked de novo DNA methylation into EBV(+) and High-marker genes that were not methylated in MKN7_p (Fig. 4A), thus were considered to acquire EBV(+) epigenotype. The methylation status was again validated by pyrosequencing (Fig. 4B). 2,215, 1,917 and 2,073 genes showed de novo methylation in MKN7_EB#1, #2 and #3, respectively, and were well overlapped (Fig. 5A), indicating that this methylation induction is a
Interestingly, \textit{de novo} methylation was induced in High-markers as well as in EBV(+)markers, but was not in \textit{MLH1} (Fig. 4A). These results supported the features in clinical samples that EBV(+)epigenotype gastric cancer showed methylation in all of Common-, High-, and EBV(+)markers, but not in \textit{MLH1}. The \textit{de novo} methylation was not detected in MKN7\_EB(-), confirming that the methylation was not due to 18-week culture during clone establishment but due to EBV infection. As for AGS, most of EBV(+) and High-marker genes had already been methylated in AGS\_p, and the methylation profile did not change in any of clones with/without rEBV infection.

\textbf{Evaluation of gene silencing.} To analyze whether the \textit{de novo} methylation contributed to gene repression, gene expression changes and methylation increase in rEBV(+) clones compared to MKN7\_EB(-) were analyzed (Fig. 5B). In all the clones of MKN7\_EB#1, #2, and #3, there was a significant correlation between decreased expression and increased $\beta$-value ($P<10^{-10}$, Kolmogorov-Smirnov test).

\textbf{Induction of EBV latent genes.} EBV latent genes, \textit{LMP2A}, \textit{EBNA1}, \textit{EBER1/2} and \textit{BARF0}, were reported to be expressed in EBV(+) gastric cancer (39, 40). MKN7 was transfected with expressing vector containing each EBV latent gene, and expression of the induced gene was confirmed by RT-PCR (Supplementary Fig. S4A). DNA methylation analysis by Infinium showed no remarkable changes of DNA methylation profile in any EBV latent gene-induced MKN7, compared to mock-transfected cells (Supplementary Fig. S4B). To analyze DNA methyltransferase level, western blotting analysis was performed. DNMT1 level in rEBV(+) MKN7 clone was increased to approximately 2-fold compared to MKN7\_EB(-) (Supplementary Fig. S4C), which is relevant to the markedly high methylation epigenotype. But increase of DNMT1 level was not detected in any EBV
latent gene-induced MKN7 compared to mock MKN7 cells (Supplementary Fig. S4D).
Discussion

We here clarified through unsupervised hierarchical clustering analysis using methylation data of thousands of genes, that clinical gastric cancer was classified into three epigenotypes, and that EBV(+) cases completely matched to the markedly higher methylation epigenotype, EBV(+) epigenotype. The feature of EBV(+) epigenotype was as follows. (a) There were genes methylated specifically in EBV(+) epigenotype, e.g. EBV(+) markers. (b) Genes methylated in EBV(-) gastric cancer were also methylated in EBV(+) cases, e.g. Common- and High-markers. In other words, there was no phenotype with EBV(+) marker methylation without Common- or High-marker methylation. (c) A rare exceptional gene was MLH1, which was preferentially methylated in High-epigenotype, but not methylated in EBV(+) epigenotype at all.

Aberrant promoter methylation in cancer has been observed preferentially in PRC-target genes in ES cells, suggesting replacement of reversible gene repression in a stem cell origin to permanent silencing in cancer (41). To evaluate whether the replacement to promoter methylation occurred similarly in EBV(+) gastric cancer, we analyzed enrichment of PRC-target genes in methylation markers. Whereas Common- and High-marker genes showed significant enrichment of PRC-target genes as reported (41), EBV(+) marker genes had no significant enrichment of PRC-target genes. DNA methylation was thus considered to not only replace PRC target gene repression to permanent silencing, but also extend to non-PRC-target genes in EBV(+) gastric cancer. This excessive methylation in EBV(+) gastric cancer was also suggested to imply a unique mechanism of aberrant methylation.

The methylation of MKN7 was similar to Low-epigenotype, and EBV infection induced marked de novo methylation both in EBV(+) marker and High-marker genes clearly, resulting in resemblance of EBV(+) epigenotype. The de novo methylation caused gene repression as well.
MLH1 was not methylated in MKN7, and MLH1 methylation was not induced in EBV-infected MKN7 clones (Fig. 4A), which also resembled EBV(+) epigenotype. These data indicated that EBV infection *per se* could cause markedly higher methylation epigenotype in EBV(+) gastric cancer, and may partly account for the reason why there was no case in which EBV(+) marker genes were independently methylated without High-marker gene methylation, and also why there was no MLH1 methylation in EBV(+) epigenotype.

Marker genes included previously reported tumor suppressor genes, e.g. *HIC1, CDKN1C* and *LOX* (Supplementary Table S4) (42-44). Methylation of these genes was not observed in MKN7_p, but induced in EBV-infected MKN7 clones. Moreover, EBV(+) marker *CDKN1C* was not a PRC target gene in ES cells (38). Though EBV infection in the non-neoplastic epithelium of the stomach was detected rarely, monoclonality of EBV genome suggested that the phase of EBV infection into the epithelial cells should occur at the initial or a very early stage of carcinogenesis (26, 27). These data and reports may suggest that EBV infection might play a role in gastric carcinogenesis by causally inducing excessive DNA methylation, even into non-PRC target EBV(+) marker genes e.g. *CDKN1C*.

In other EBV-associated neoplasms, e.g. nasopharyngeal carcinoma (45, 46) and natural killer cell lymphoma (47), DNA hypermethylation in the limited number of promoter regions has been reported. DNA methylation of the double-strand DNA virus genome itself was also observed in the host cells, not only for EBV, but also other oncogenic viruses e.g. hepatitis B virus (HBV) and human papilloma virus (HPV) (48). While methylation of viral genome may be induced by defense reaction of host cells against foreign invasion, viruses could take advantage by suppressing adverse expression of viral genes to escape from host immunity and maintain latent infection, therefore maintain virus(+) cancer cells. The host defense system might induce aberrant methylation in host
cellular genome, too, as a side-effect.

In the latent EBV infection, only *EBNA1, EBER1, EBER2, LMP2A* and *BARF0* genes are expressed (40). To gain insight which gene expression could contribute to methylation induction, each latent gene was introduced in MKN7 and cells were cultured for 6 weeks. Overexpression of any single gene did not induce *de novo* methylation, at least at the CpG sites analyzed by Infinium. Several possibilities are raised. Firstly, whereas rEBV(+) clones were established through 18-week culture, 6 weeks may not be long enough to acquire *de novo* methylation. Secondly, combination of 2-5 latent genes might be necessary to induce methylation. Thirdly, some mechanism in host cells may also have to be activated through virus infection. Considering these possibilities, further analyses, e.g. infection of rEBV with a specific gene knock-out, might clarify which gene expression should be critical in methylation induction. Interestingly, an EBV(-) cell line AGS showed EBV(+) epigenotype, and rEBV infection could not induce any more methylation. This might suggest that host cells themselves have methylation mechanism and causes other than EBV infection could also trigger the high methylation epigenotype, though it might be rare and mostly triggered by EBV infection. In any case, the host genome methylation cannot be explained simply by methylation expansion from inserted foreign DNA to surrounding sequences on a chromosome, since EBV virus exists as an extrachromosomal circular molecule in the nucleus of cancer cell (40).

In summary, gastric cancer was classified into three DNA methylation epigenotypes, EBV-positive gastric cancers showed markedly high methylation epigenotype expanding to the non-PRC target genes, and EBV infection *per se* could cause the methylation of EBV(+) epigenotype.
Acknowledgements

We thank Prof. Paul J. Farrell for vector supply, and Hiroko Meguro, Kyoko Fujinaka, Kaori Shiina, Aiko Nishimoto, Kei Sakuma, and Maya Isogai for technical help. This work was supported by Grants-in-Aids for Scientific Research (B) 22390255, (S) 20221009 and (A) 20249022 from Ministry of Education, Culture, Sports, Science and Technology of Japan, and by JST PRESTO program.
References


19. Leung SY, Yuen ST, Chung LP, Chu KM, Chan AS, Ho JC. hMLH1 promoter methylation and lack of hMLH1 expression in sporadic gastric carcinomas with high-frequency microsatellite instability. Cancer Res 1999; 59: 159-64.


36. Laboratory of DNA Information Analysis, Human Genome Center, Institute of Medical Science,
Matsusaka K, et al.


Table 1. Clinicopathological features in clinical gastric cancer samples

<table>
<thead>
<tr>
<th>Clinical features</th>
<th>All cases</th>
<th>Low-epigenotype</th>
<th>High-epigenotype</th>
<th>EBV(+) epigenotype</th>
<th>Outlier</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong># of samples</strong></td>
<td>51 (100%)</td>
<td>7 (14%)</td>
<td>13 (25%)</td>
<td>11 (22%)</td>
<td>20 (39%)</td>
<td></td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>38 (75%)</td>
<td>6 (86%)</td>
<td>7 (54%)</td>
<td>8 (73%)</td>
<td>17 (85%)</td>
<td>0.61</td>
</tr>
<tr>
<td>Female</td>
<td>13 (25%)</td>
<td>1 (14%)</td>
<td>6 (46%)</td>
<td>3 (27%)</td>
<td>3 (15%)</td>
<td>0.42</td>
</tr>
<tr>
<td><strong>Age (y)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean ± SD</td>
<td>69.7 ±/− 10.5</td>
<td>61.1 ±/− 13.4</td>
<td>69.0 ±/− 8.8</td>
<td>70.9 ±/− 9.9</td>
<td>69.3 ±/− 10.7</td>
<td>0.062</td>
</tr>
<tr>
<td><strong>Tumor location in the stomach</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Upper</td>
<td>26 (51%)</td>
<td>3 (43%)</td>
<td>3 (23%)</td>
<td>4 (36%)</td>
<td>16 (80%)</td>
<td>0.50</td>
</tr>
<tr>
<td>Middle</td>
<td>12 (24%)</td>
<td>3 (43%)</td>
<td>3 (23%)</td>
<td>5 (45%)</td>
<td>0 (0%)</td>
<td>0.27</td>
</tr>
<tr>
<td>Lower</td>
<td>13 (25%)</td>
<td>1 (14%)</td>
<td>7 (54%)</td>
<td>1 (9%)</td>
<td>4 (20%)</td>
<td></td>
</tr>
<tr>
<td><strong>Histology</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intestinal type</td>
<td>24 (47%)</td>
<td>1 (14%)</td>
<td>8 (62%)</td>
<td>1 (9%)</td>
<td>14 (70%)</td>
<td>0.051</td>
</tr>
<tr>
<td>Diffuse type</td>
<td>27 (53%)</td>
<td>6 (86%)</td>
<td>5 (38%)</td>
<td>10 (91%)</td>
<td>6 (30%)</td>
<td>*0.013</td>
</tr>
<tr>
<td><strong>Invasion Depth</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>m, sm</td>
<td>8 (16%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>4 (36%)</td>
<td>4 (20%)</td>
<td>0.13</td>
</tr>
<tr>
<td>mp</td>
<td>9 (18%)</td>
<td>0 (0%)</td>
<td>5 (38%)</td>
<td>1 (9%)</td>
<td>3 (15%)</td>
<td>0.15</td>
</tr>
<tr>
<td>ss, se</td>
<td>34 (67%)</td>
<td>7 (79%)</td>
<td>8 (62%)</td>
<td>6 (55%)</td>
<td>13 (65%)</td>
<td></td>
</tr>
<tr>
<td><strong>Lymph node metastasis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(+)</td>
<td>30 (59%)</td>
<td>4 (57%)</td>
<td>7 (54%)</td>
<td>7 (64%)</td>
<td>12 (60%)</td>
<td>0.94</td>
</tr>
<tr>
<td>(-)</td>
<td>21 (41%)</td>
<td>3 (43%)</td>
<td>6 (46%)</td>
<td>4 (36%)</td>
<td>8 (40%)</td>
<td>0.70</td>
</tr>
<tr>
<td><strong>EBER in situ hybridization</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(+)</td>
<td>11 (22%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>11 (100%)</td>
<td>0 (0%)</td>
<td>*3.0×10⁻⁶</td>
</tr>
<tr>
<td>(-)</td>
<td>40 (78%)</td>
<td>7 (100%)</td>
<td>13 (100%)</td>
<td>0 (0%)</td>
<td>20 (100%)</td>
<td>*4.0×10⁻⁵</td>
</tr>
<tr>
<td><strong>MLH1 methylation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(+)</td>
<td>7 (14%)</td>
<td>0 (0%)</td>
<td>6 (46%)</td>
<td>0 (0%)</td>
<td>1 (5%)</td>
<td>*0.037</td>
</tr>
<tr>
<td>(-)</td>
<td>44 (86%)</td>
<td>7 (100%)</td>
<td>7 (54%)</td>
<td>11 (100%)</td>
<td>19 (100%)</td>
<td>*0.016</td>
</tr>
</tbody>
</table>

SD, standard deviation. m, mucosa. sm, submucosa mp, musculus propria. ss, subserosa. se, serosa. *P-value <0.05. L, Low. H, High. E, EBV(+). EBV(+) epigenotype significantly correlated to EBV(+), MLH1 methylation(-), and diffuse type histology.
Figure legends

Fig. 1. Unsupervised two-way hierarchical clustering of gastric cancer. Upper heatmap: 4,249 probes extracted from high-CpG and intermediated-CpG probes were used for clustering. Middle heatmap: 915 probes from low-CpG probes were additionally shown. Bottom: clinicopathological factors (summarized in Table 1). Gender: pale blue, male; dark red, female. Age (years): light orange, <64; orange, 65-74; brown, >75. Histology: pale blue, intestinal type; dark red, diffuse type. Invasion Depth: light orange, mucosa/submucosa; orange, muscularis propria; brown, subserosa or deeper. LN (lymph node) metastasis: dark red, positive; pale blue, negative. Gray: N/A. Right: methylation control samples (0%-100%), and PRC-target genes in ES cells (black bar) (38) were shown. Clinical gastric cancer samples were classified into three subgroups: low and high methylation epigenotypes (Low- and High-epigenotypes) for EBV(-) gastric cancer, and markedly higher methylation epigenotype that completely matched EBV(+) cases (EBV(+)-epigenotype). EBV(+) cell line and xenograft, SNU719 and KT, and EBV(-) cell lines, MKN7 and AGS, were also shown. MLH1 methylation (black) was frequently detected in High-epigenotype and not detected in EBV(+) cases at all (P<0.05, χ² test). Cases not belonging to any epigenotype were denominated “outlier” hereafter (Fig. 2).

Fig. 2. (A) Marker extraction. EBV(+)-markers were 53 genes methylated in EBV(+)-epigenotype only. High-markers were 79 genes methylated in EBV(+)- and High-epigenotypes, but not in Low-epigenotype or NGM. Common-markers were 72 genes methylated commonly in EBV(+)-, High-, and Low-epigenotypes. Right, PRC-target genes in ES cells (black bar) (38). (B) The ratio of PRC target genes. PRC-target genes were significantly enriched in High-markers (P=2×10⁻¹⁵) and Common-markers (P=2×10⁻³⁴), but not in EBV(+)-markers (P=0.2), compared to the ratio in the whole 13,897 genes. (C) Validation of 9 markers
by pyrosequencing. *Upper*, Infinium result. *Lower*, pyrosequencing result. While Infinium probe determines methylation status of single CpG site, pyrosequencing data not only validated the methylation status of the probe site (*red* number in each marker), but also analyzed multiple surrounding CpG sites quantitatively (See also Supplementary Fig. S1).

**Fig. 3.** EBV infection. (A) HE staining and EBER-ISH for gastric cancer tissue and non-neoplastic lesion. EBER-ISH showed EBV existence in cancer cells. Tumor stroma of EBV(+) gastric cancer showed lymphocyte infiltration, and these lymphocytes were negative for EBER-ISH (*upper right*). (B) EBER-ISH for cell lines. While SNU719 was positive for EBV, MKN7 parental cells (MKN7_p) and AGS parental cells (AGS_p) were negative. Clones infected with rEBV were representatively shown (MKN7_EB#1 and AGS_EB#1), and other clones were shown in Supplementary Fig. S2. (C) Immunohistochemical analysis of EBV(+) MKN7 clones and Akata cells. While cell block of Akata cells was positive for CD45 and negative for Cytokeratin, MKN7_EB#1 was positive for Cytokeratin and negative for CD45, indicating no contamination of Akata cells in rEBV-infected clones. Other clones were shown in Supplementary Fig. S3. (D) Real-time PCR for EBV genome. Copy number of rEBV genomes was measured quantitatively by real-time PCR of *BZLF* gene, normalized to *CRP* of host genome. (E) Real-time RT-PCR of EBV genes. Expression of EBV genes, *EBER1*, *EBER2* and *EBNA1*, was measured quantitatively, normalized to *ACTB*. Their expression in infected clones was confirmed.

**Fig. 4.** Induction of de novo methylation by rEBV infection. *p*, parental cell. (-), rEBV(-) control cell cultured for the same period without rEBV infection. #1-3, three rEBV(+) clones. (A) Infinium result. MKN7_p showed methylation generally in Common-markers, but not in High- or EBV(+)markers,
and thus MKN7 was regarded as Low-epigenotype cell line. While there was no methylation change in MKN7_(-) clone, methylation was induced in most of unmethylated High- and EBV(+) markers in MKN7_#1, #2 and #3 clones. In an EBV(+) epigenotype cell line AGS, Common-, High- and EBV(+) markers were methylated already in AGS_p, and rEBV infection did not induce any more methylation. (B) Validation of genes with de novo methylation by pyrosequencing. Upper, Infinium result. Lower, pyrosequencing result. De novo methylation of High- and EBV(+) markers in MKN7_#1, #2 and #3 clones was validated about 9 genes analyzed in Fig. 2C.

Fig. 5. (A) Overlapping of genes with de novo methylation in MKN7_#1, #2 and #3. Genes whose $\beta$-value in MKN7_p was <0.2 and showing increase of $\beta$-value by >0.2 compared to MKN7_p, were counted. (B) Enrichment of genes with de novo methylation in repressed genes. From the whole 13,897 genes analyzed, 6,170 genes were extracted, by excluding genes not on expression array and genes showing GeneChip score <100 in all of MKN7_#1, #2 and #3 clones. The 6,170 genes were sorted by the expression ratio of EBV(+) clone to MKN7_#1, #2 and #3 clones. In descending order. Genes with $\beta$-value increase by >0.5 compared to MKN7_#1, #2, #3; Kolmogorov-Smirnov test), indicating that EBV-induced methylation correlated to gene repression significantly.
Fifty-one gastric cancers, two normal gastric mucosa (NGM), and two peripheral blood cells (PBC)
Matsusaka_Fig 3

(A) HE, EBER-ISH

(B) SNU719, MKN7-p, MKN7-EB\#1, AGS-p, AGS-EB\#1

(C) HE, Cytokeratin, CD45

(D) BZLF1 / CRP

(E) EBER1 / ACTB, EBER2 / ACTB, EBNA1 / ACTB
Matsusaka_Fig 4
Matsusaka_Fig 5

A

MKN7-EB#1 (2,215)
191
157

348

123

1,519

MKN7-EB#2 (1,917)

MKN7-EB#3 (2,073)
122

B

<table>
<thead>
<tr>
<th></th>
<th>MKN7-EB#1</th>
<th>MKN7-EB#2</th>
<th>MKN7-EB#3</th>
</tr>
</thead>
<tbody>
<tr>
<td>GeneChip</td>
<td>!<strong>!</strong>*</td>
<td>!**!</td>
<td>!*!</td>
</tr>
<tr>
<td>Infinium</td>
<td>!**!</td>
<td>!*!</td>
<td>!*!</td>
</tr>
</tbody>
</table>

Fold change (log2)

\[ \beta\text{-value increase} \]

\[ > +0.5 \]

\[ P < 10^{-15} \]

\[ P = 4 \times 10^{-15} \]

\[ P = 9 \times 10^{-11} \]
Classification of Epstein-Barr virus positive gastric cancers by definition of DNA methylation epigenotypes

Keisuke Matsusaka, Atsushi Kaneda, Genta Nagae, et al.

Cancer Res  Published OnlineFirst October 11, 2011.