ORIGINAL ARTICLE

Adenocarcinoma of the stomach and esophagogastric junction with low DNA methylation show poor prognoses

Masayuki Urabe[1,2,3] · Keisuke Matsusaka[3,4] · Tetsuo Ushiku[2] · Masaki Fukuyo[3,5] · Bahityar Rahmutulla[3] · Hiroharu Yamashita[1] · Yasuyuki Seto[1] · Masashi Fukayama[2] · Atsushi Kaneda[3]

Abstract

Background Gastric cancer (GC) is characterized by unique DNA methylation epigenotypes (MEs). However, MEs including adenocarcinomas of the esophagogastric junction (AEG) and background non-neoplastic columnar mucosae (NM) remain to be clarified.

Methods We analyzed the genome-wide DNA MEs of AEG, GC, and background NM using the Infinium 450 k beadarray, followed by quantitative pyrosequencing validation. Large-scale data from The Cancer Genome Atlas (TCGA) were also reviewed.

Results Unsupervised two-way hierarchical clustering using Infinium data of 21 AEG, 30 GC, and 11 NM revealed four DNA MEs: extremely high-ME (E-HME), high-ME (HME), low-ME (LME), and extremely low-ME (E-LME). Promoter methylation levels were validated by pyrosequencing in 146 samples. Non-inflammatory normal mucosae were clustered into E-LME, whereas gastric or esophagogastric junction mucosae with chronic inflammatory changes caused by either *Helicobacter pylori* infection or reflux esophagitis were clustered together into LME, suggesting that inflammation status determined DNA MEs regardless of the cause. Three cases of Barrett's-related adenocarcinoma were clustered into HME. Among 94 patients whose tumors could be clustered into one of four MEs, 11 patients with E-LME cancers showed significantly shorter overall survival than that in the other MEs, even with the multivariate Cox regression estimate. TCGA data also showed enrichment of AEG in HME and a poorer prognosis in E-LME.

Conclusions E-LME cases, newly confirmed in this study, form a unique subtype with poor prognosis that is not associated with inflammation-associated elevation of DNA methylation levels. LME could be acquired via chronic inflammation, regardless of the cause, and AEG might preferentially show HME.

Keywords Adenocarcinoma of the esophagogastric junction · Gastric cancer · DNA methylation epigenotype

---

Masayuki Urabe and Keisuke Matsusaka contributed equally to this manuscript.

✉ Atsushi Kaneda
kaneda@chiba-u.jp

[1] Department of Gastrointestinal Surgery, Graduate School of Medicine, the University of Tokyo, Tokyo, Japan

[2] Department of Pathology, Graduate School of Medicine, the University of Tokyo, Tokyo, Japan

[3] Department of Molecular Oncology, Graduate School of Medicine, Chiba University, Inohana 1-8-1, Chuo-Ku, Chiba 260-8670, Japan

[4] Department of Pathology, Chiba University Hospital, Chiba, Japan

[5] Department of Genome Research and Development, Kazusa DNA Research Institute, Chiba, Japan

Introduction

The development of gastric cancer (GC), a leading cause of cancer-related mortality worldwide [1–3], has an attributable association with two infectious pathogens; *Helicobacter pylori* (*H. pylori*) and Epstein-Barr virus (EBV). *H. pylori* infection in the gastric mucosa is a crucial etiological risk factor in GC, and is classified as a definite carcinogen by the World Health Organization [4, 5]. Accumulating DNA methylation represents chronic mucosal inflammation induced by *H. pylori* and results in the silencing of tumor suppressor genes. It is subsequently related to gastric carcinogenesis [6–9]. Indeed, a recent prospective study has demonstrated that the methylation levels of some marker genes in normal gastric mucosa are associated with an increased risk

of developing metachronous gastric cancers [10, 11]. As for EBV infection, we previously conducted DNA methylation analysis on a genome-wide scale in clinical samples of GC using the Infinium HumanMethylation 27 k BeadArray (Illumina, San Diego, CA, USA). The GC cases were clustered into three DNA methylation epigenotypes (MEs), i.e., extremely high-ME, high-ME, and low-ME; moreover, cases not belonging to any MEs were defined as "outliers." We found that EBV-positive GC showed prominently high levels of genome-wide DNA methylation [12], consistent with the results reported in previous publications [13–17], which have been followed and verified by the latest large-scale comprehensive landmark studies on GC [18–20].

Although the molecular basis of GC has been extensively examined, that of adenocarcinoma of the esophagogastric junction (AEG) and background non-neoplastic columnar mucosae (NM) remains unclear. The worldwide incidence of AEG has grown at an alarming rate in contrast to the declining frequency of GC [2, 3, 21–23]. Owing to its unique location on the boundary between the esophagus and stomach, AEG is considered to have a somewhat heterogeneous background: AEG is assumed to partially share the same *H. pylori*/EBV-associated carcinogenesis with GC, although it also substantially comprises Barrett's-related adenocarcinoma (BA), which is causally ascribed to increased body weight, gastroesophageal reflux disease, and absence of *H. pylori* colonization [24–26]. A past study comparing the mutational profiles between GC and esophagogastric junction (EGJ) tumors showed that 49% (57/117) of recurrently mutated genes were unique to EGJ tumors [27]. A recent DNA methylation analysis of AEG reported *SLC46A3* and *cg09177106* as prediction marker genes for the tissue origin of AEG [28]. To gain insight into DNA MEs of AEG and their similarity to or difference from GC, comprehensive DNA methylome analyses of GC and AEG should be conducted with respect to their tumorigenic backgrounds.

To address this issue, this study quantitatively estimated the DNA methylation patterns of 62 fresh-frozen clinical samples, including 21 GC, 30 AEG, and 11 background NM (ten non-neoplastic gastric mucosae (NGM) and one non-neoplastic Barrett's mucosa (NBM)) using the Infinium HumanMethylation 450 k beadarray (Illumina) for more than 480,000 individual CpG dinucleotides. The results of the Infinium methylation assay were validated using pyrosequencing. Based on the pyrosequencing results, we also obtained the epigenetic profiles of additional cases, including 38 background NM (35 NGM and three NBM) and 46 GC/AEG (including four BA). Further, we assessed the correlation between clinical-pathological factors and distinguished MEs. Finally, we reviewed our results using large-scale data from The Cancer Genome Atlas (TCGA).

Materials and methods

The flowchart of the study is shown in Supplementary Fig. 1.

Ethics statement

All samples were retrieved from the archives of the Department of Pathology, the University of Tokyo Hospital, and were approved for use in this study by the Ethics Committee of the University of Tokyo and Chiba University.

Clinical samples

We obtained 21 GC, 30 AEG, and 11 background NM (ten NGM and one NBM) samples from patients who underwent surgical resection at the University of Tokyo (from October 2003 to May 2015) with written informed consent. These samples were immediately frozen in liquid nitrogen and kept at $-80\ °C$ until further use. Two independent observers (M.U. and K.M.) microscopically examined the fresh-frozen cancer tissues to determine tumor cellularity. All tumor samples were microscopically confirmed to contain $\geq 40\%$ of cancer cells and were used for further Infinium analysis and subsequent quantitative validation by pyrosequencing. DNA was extracted using the QIAamp DNA MicroKit (QIAGEN, Valencia, CA, USA). For columnar mucosal samples without any cancer components, we histopathologically evaluated the presence/absence of *H. pylori*-associated gastritis (atrophic changes, intestinal metaplasia, and infiltration of inflammatory cells, which are associated with *H. pylori* colonization on Giemsa staining) (Fig. 1b–d) and Barrett's esophagus (characterized by mucosal distortion, mild inflammation, and remnants of proper esophageal glands) (Fig. 1e). Two observers (M.U. and K.M.) performed these assessments, along with one more gastroenterological pathologist (T.U.). Mucosal samples were dissected to enrich more than 30% of epithelial cells if needed.

We also extracted genomic DNA from 43 fresh-frozen samples (35 NGM and eight GC/AEG) and 41 formalin-fixed paraffin-embedded (FFPE) samples (three NBM and 38 GC/AEG including four BA) from patients who underwent surgical resection at the University of Tokyo (from October 2001 to May 2015, with written informed consent). DNA was extracted using a QIAamp DNA MicroKit (QIAGEN). We performed pyrosequencing analysis to determine the epigenetic profiles of each specimen. The clinical data of the enrolled patients were obtained by reviewing medical records.

Immunohistochemistry and in situ hybridization using tissue microarray technology

Tissue microarrays were constructed from 94 tumor cases (45 GC and 49 AEG) using a manual tissue array (Beecher Instruments, Inc., Sun Prairie, WI, USA). For immunohistochemical analysis, 4 μm sections were cut from tissue microarray blocks and tissue blocks just before use. Immunohistochemistry was performed using the Ventana Benchmark XT autostainer (Ventana Medical Systems Inc./Roche Diagnostics, Tucson, AZ, USA), involving a labeled streptavidin–biotin–peroxidase method, followed by visualization with 3,3'-diaminobenzidine.

The primary antibodies used included mouse monoclonal anti-MLH1 (clone ES05, dilution 1:50; Novocastra Laboratories Ltd., Newcastle, UK) and anti-p53 (DO-7, dilution 1:50; Novocastra Laboratories Ltd.). The presence or absence of MLH1 immunostaining was evaluated in the nuclei (Fig. 1f–g). The results of p53 immunohistochemistry were determined as "aberrant" when < 1% of tumor nuclei were positive for p53 (null pattern) or when more than 50% of the neoplastic cells were positive for p53 (overexpression pattern); the remaining results were considered as "wild type." EBV-encoded small RNA in situ hybridization (EBER-ISH) was performed on paraffin sections using a fluorescein isothiocyanate-labeled peptide nucleic acid probe (Y5200; Dako) and anti-FITC (V0403, dilution 1:200; Dako, Glostrup, Denmark).

Infinium methylation assay

The DNA methylation status of the tumor and non-tumor mucosal tissues was established using Illumina Infinium HumanMethylation450 BeadChips, as previously described [12]. Methylation levels at each CpG site were represented as the fluorescence signal ratio, called a $β$-value (ranging from 0.00 to 1.00). Bisulfite conversion of 500 ng genomic DNA from each sample was performed using the Zymo EZ DNA Methylation Kit (Zymo Research, Irvine, CA, USA). Whole-genome amplification, labeling, hybridization, and scanning were performed according to the manufacturer's protocols. Genes in X and Y chromosomes were excluded to avoid gender differences. SNU-719 cells derived from EBV-positive gastric cancer were obtained from the Korean Cell Line Bank (Seoul, Korea). A xenograft tumor, KT, derived from EBV-positive gastric cancer, was previously established [29]. The methylation control samples were prepared as previously reported [30]. Infinium data were submitted to Gene Expression Omnibus (GEO) (accession no. GSE 207,846). DNA methylation profiles of two NFGM samples (083 N and 431 N), four gastric cancer samples (065 T, 125 T, 121 T, and 018 T), SNU719, and KT were previously analyzed using Infinium 450 k (GSM2363429–2363434 and GSM2633597–2633598 in GSE89269) [31]. DNA methylation profiles of embryonic stem cells (ESC) and blood cells were obtained from GEO datasets (GSM867942 and GSM868038 in GSE30654) [32].

Pyrosequencing analysis

Methylation status was quantitatively validated by pyrosequencing, as previously reported [12]. For each category of markers, multiple genes were randomly chosen to prepare representative marker genes to classify tissue samples into MEs (Supplementary Table 1). Primers were designed to include no or only one CpG site in their sequence using Pyro Q-CpG Software (QIAGEN) to amplify bisulfite-treated DNA regions containing several CpG sites. For the C of a CpG site within a primer sequence, a nucleotide that does not anneal to C or U was chosen, e.g., adenosine (A). Briefly, the biotinylated PCR product was bound to streptavidin sepharose beads HP (Amersham Biosciences, Sweden), washed, and denatured using 0.2 mol/L NaOH. After adding 0.3 μmol/L sequencing primers to the purified, single-stranded PCR product, pyrosequencing was performed using the PyroMark Q96 ID System (QIAGEN) according to the manufacturer's instructions. The primer sequences and conditions are shown in Supplementary Table 2.

TCGA data

Cancer genomics data from TCGA were downloaded from TCGA Firehose portal (Broad Institute TCGA Genome Data Analysis Center (2016): Firehose stddata_2016_01_28 run. Broad Institute of MIT and Harvard, https://doi.org/10.7908/C11G0KM9; https://gdac.broadinstitute.org/).

Statistical analysis

Unsupervised two-way hierarchical clustering analyses were performed using Cluster 3.0 software (http://bonsai.hgc.jp/~mdehoon/software/cluster/). Heatmaps were drawn using Java TreeView software (http://jtreeview.sourceforge.net/). Clinicopathological data were compared using the Chi-squared exact test for categorical variables. The survival curves were drawn using the Kaplan–Meier method. The log-rank test analyzed differences between curves. Univariate and multivariate Cox regression was also performed for survival analysis; multivariate Cox regression incorporated all variables on univariate analysis. Overall survival was defined as survival until death from any cause. Gene annotation enrichment analysis was performed for Gene Ontology (GO; Biological process) and KEGG pathway using
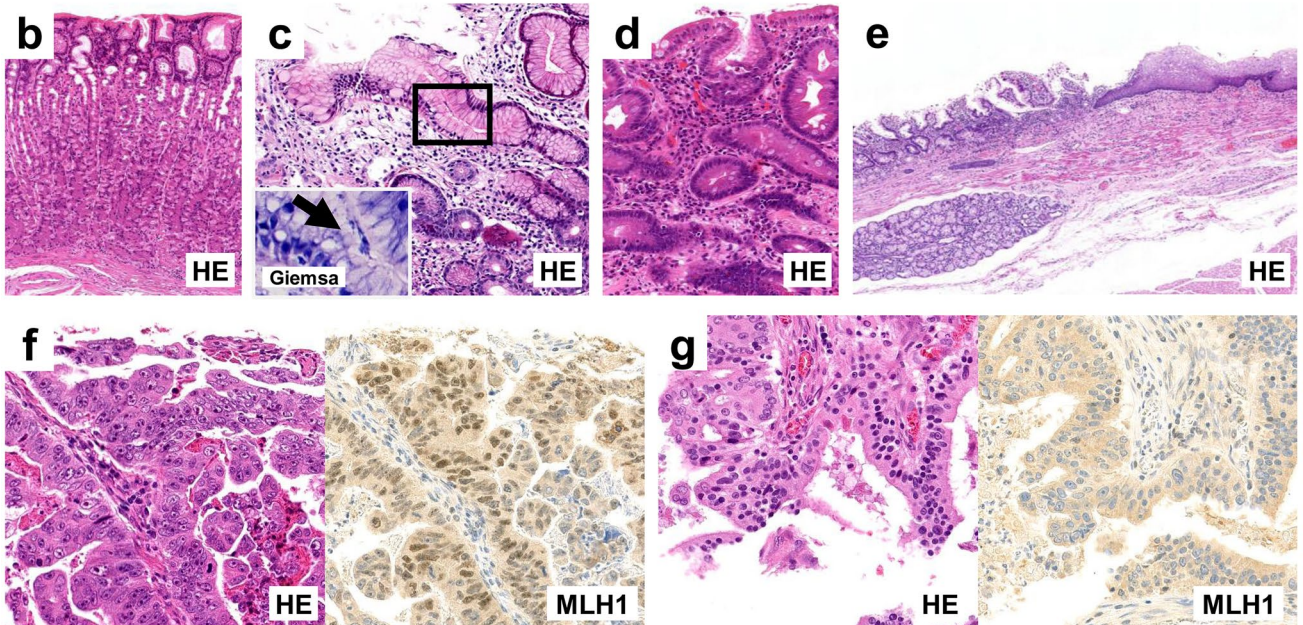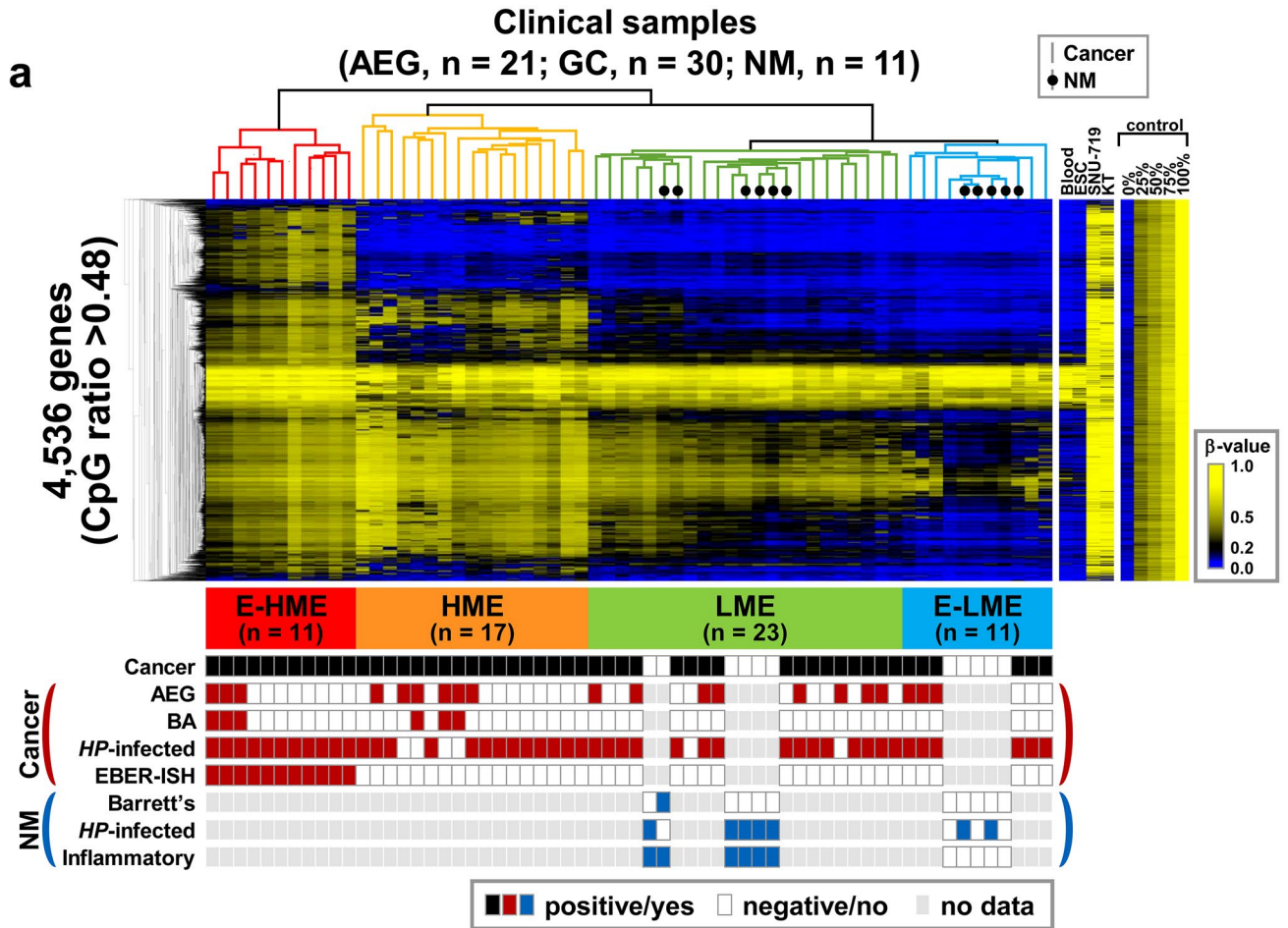
a Clinical samples (AEG, n = 21; GC, n = 30; NM, n = 11)

| Cancer
● NM

4,536 genes (CpG ratio >0.48)

Blood ESC SNU-719 KT

control
0% 25% 50% 75% 100%

β-value
1.0
0.5
0.2
0.0

E-HME (n = 11) | HME (n = 17) | LME (n = 23) | E-LME (n = 11)

Cancer

Cancer
AEG
BA
HP-infected
EBER-ISH

NM
Barrett's
HP-infected
Inflammatory

■ ■ ■ positive/yes □ negative/no ■ no data

b HE

c HE Giemsa

d HE

e HE

f HE MLH1

g HE MLH1

◄**Fig. 1** Unsupervised two-way hierarchical clustering using Infinium 450 k and microscopic appearance of the clinical samples. **a** Upper left heat map: in total, 4536 probes extracted from high- and intermediate-CpG probes were used for clustering. *Horizontal*, cases; *Vertical*, genes. Black dots indicate background NM on the sample dendrogram. Upper middle heat map: peripheral blood cells (Blood), embryonic stem cells (ESC), EBV-positive cells (SNU-719), and xenograft (KT) are shown as controls. Upper right heat map: methylation control samples (0–100%). Unsupervised two-way hierarchical clustering revealed four distinct MEs among 62 clinical samples (AEG, $n = 21$; GC, $n = 30$; background NM, $n = 11$): E-HME ($n = 11$), HME ($n = 17$), LME ($n = 23$), and E-LME ($n = 11$). Bottom column, clinicopathologic factors. *Cancer*: black, yes; white, no (background NM). *AEG*: dark red, yes; white, no ($=$ GC). *BA*: dark red, yes (Barrett's-related adenocarcinoma); white, no (Barrett's-unrelated AEG or GC). *HP-infected*: dark red, positive by Giemsa stain; white, negative. *EBER-ISH*: dark red, positive; white, negative. *Barrett's*: dark blue, Barrett's mucosa; white, gastric mucosa. *HP-infected*: dark blue, positive by Giemsa stain; white, negative. *Inflammatory*: dark blue, chronic gastritis; white, no inflammatory cell infiltration or chronic inflammatory changes. *HP*, *Helicobacter pylori*. Histological photographs (**b–g**). **b** Normal gastric mucosa (fundic gland) without any atrophic changes or inflammatory findings (hematoxylin and eosin (HE) stain). **c** Active gastritic mucosa with *H. pylori* colonization (inset, arrow, Giemsa stain) in the foveolar epithelium (HE stain). **d** Markedly atrophic gastric mucosa with complete type intestinal metaplasia (HE stain). **e** Barrett's mucosa characterized by mucosal distortion, mild inflammation, and remnants of proper esophageal glands in the submucosa, adjacent to the esophageal squamous epithelium (HE stain). **f** A specimen with positive MLH1 expression in cancer cell nuclei (HE stain and MLH1-immunohistochemistry). **g** A specimen with deficient MLH1 expression in cancer cell nuclei (HE stain and MLH1-immunohistochemistry)

the Functional Annotation Tool at DAVID Bioinformatics Resources 6.8 (https://david.ncifcrf.gov/). Values of $P < 0.05$ were considered significant in all statistical analyses. Statistical analyses were performed using JMP Pro version 12.2.0 (SAS Institute, Cary, NC) or BellCurve for Excel (Social Survey Research Information Co., Ltd., Tokyo, Japan).

Results

Probe selection for genome-wide DNA methylation analysis using infinium

The following criteria were applied for analyzing Infinium data: (i) the probe with the highest CpG ratio was selected for one promoter when multiple probes were designed; (ii) standard deviations (SD) of the β-value in 62 fresh frozen clinical samples (30 AEG, 21 GC, 11 background NM) were calculated for each probe, and probes with SD less than 0.1 were excluded for hierarchical clustering; (iii) probes with intermediate- and high-CpG ratio were used in hierarchical

clustering to analyze promoters with high CpG density (e.g., CpG islands). Based on these criteria, 4536 probes were selected.

Unsupervised two-way hierarchical clustering of GC, AEG, and background NM

Unsupervised hierarchical clustering analysis classified a total of 62 fresh-frozen samples into four distinguished MEs without any outliers (Fig. 1a): Extremely high ME (E-HME), high ME (HME), low ME (LME), and extremely low ME (E-LME). Some implicative findings were noted in this clustering analysis: (i) AEG and GC were clustered together in each ME, including E-HME. (ii) Of 20 AEG, three cases were pathologically diagnosed as BA and clustered into HME without MLH1-deficient cases. (iii) Background NM cases were also clustered into LME or E-LME, the former showing genome-wide aberrant methylation levels equal to those in some LME cancer cases. One case of mucosa with Barrett's-related inflammation was clustered into LME. Five mucosae with microscopic inflammation derived from *H. pylori* infection were clustered into LME. (iv) Five gastric mucosae in E-LME included two cases of *H. pylori*-infected epithelium. However, in histological assessment, all fresh-frozen mucosal tissues, whether accompanied by *H. pylori*, involved infiltration of several inflammatory cells and showed no atrophic changes.

Validation of DNA methylation by pyrosequencing

Marker genes characterizing four MEs were extracted from probes included in the Infinium assays (Fig. 2a. *See* Supplementary Table 1 for the list of marker genes and the extraction criteria) as follows: 568 E-HME marker genes (specifically methylated in E-HME), 358 HME marker genes (methylated in E-HME and HME but unmethylated in LME or E-LME), 319 LME marker genes (methylated in E-HME, HME, and LME but unmethylated in E-LME), and 153 E-LME marker genes (methylated in E-HME, HME, LME, and E-LME tumors but unmethylated in E-LME mucosa). Significant enrichment of GO-terms "cell adhesion" (GO:0007155) and "regulation of transcription from RNA polymerase II promoter" (GO:0006357) was observed in the 153 E-LME marker genes (Supplementary Table 3).

Pyrosequencing could guarantee the methylation status of each Infinium probe site and the surrounding multiple CpG sites. Using the E-HME, HME, and LME marker genes, the methylation status obtained with Infinium was quantitatively
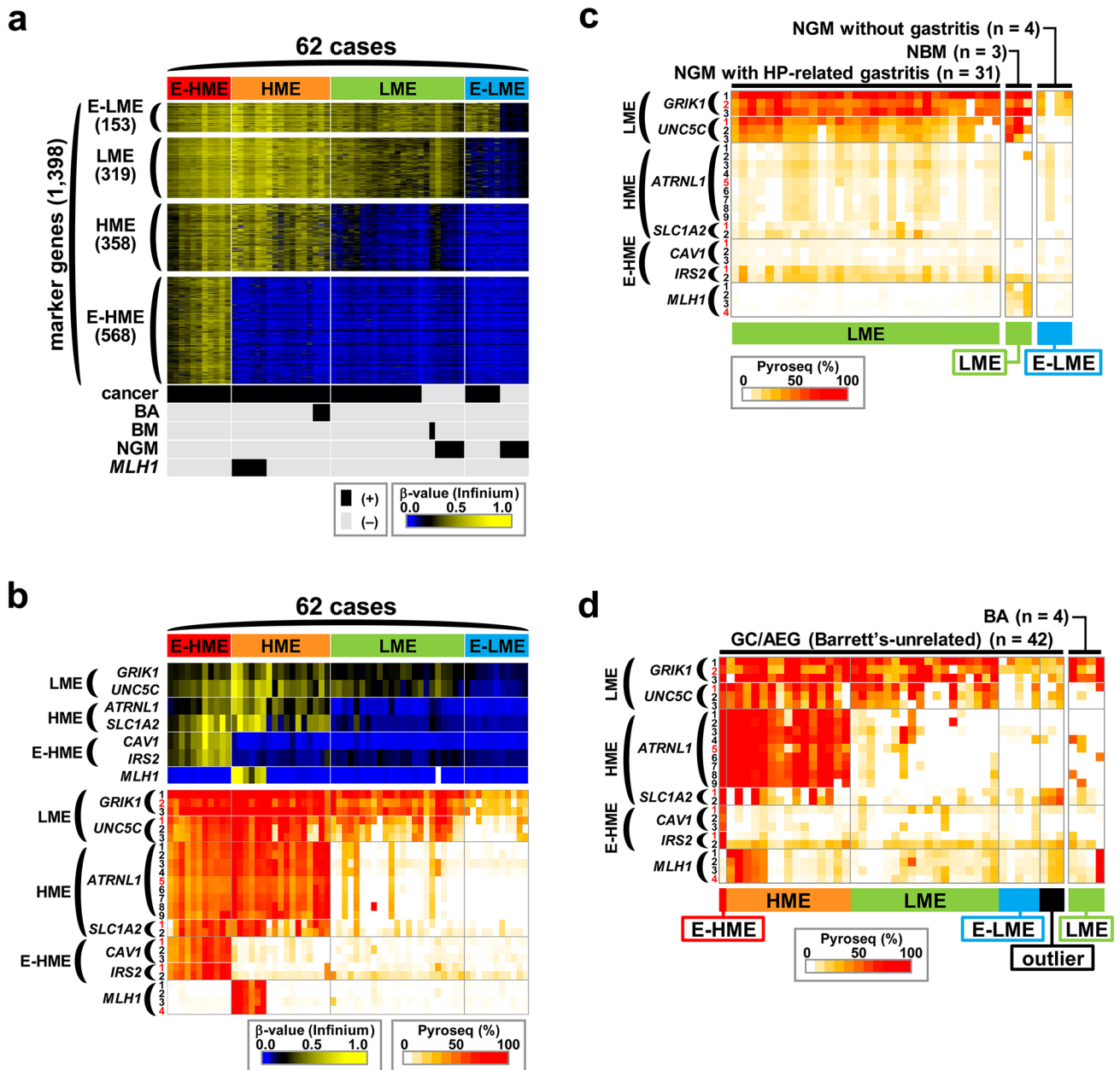
Fig. 2 Marker gene extraction and validation by pyrosequencing. **a** In total, 1398 marker genes characterizing four MEs were extracted from probes included in Infinium assays; 568 E-HME marker genes (specifically methylated in E-HME), 358 HME markers (methylated in E-HME and HME but unmethylated in LME or E-LME), 319 LME markers (methylated in E-HME, HME, and LME but unmethylated in E-LME), and 513 E-LME markers (methylated in E-HME, HME, LME, and E-LME tumors but unmethylated in E-LME mucosa). *BA*, Barrett's-related adenocarcinoma; *NBM*, non-neoplastic Barrett's mucosa; *NGM*, non-neoplastic gastric mucosa. **b** Validation for Infinium data using pyrosequencing of approximately seven marker genes. *Upper*, Infinium result; *lower*, pyrosequencing results. Horizontal cases correspond to those in Fig. 1a. Pyrosequencing resulted in highly quantitative validation of sequential DNA methylation status in Infinium probe-targeted CpG sites and surrounding CpG sites. *Red* number, Infinium probe-targeted CpG sites. **c** Pyrosequencing data for 38 non-cancerous epithelial tissues: NGM with *H. pylori*-associated gastritis ($n = 31$), BM ($n = 3$) and NGM without gastritis ($n = 4$). Based on our panel method, 31 gastritis NGM and 3 BM were classified as LME, and four non-gastritis NGM were classified as E-LME, without any exceptions. **d** Pyrosequencing data for 46 cancer tissues: Barrett's-unrelated GC/AEG ($n = 42$) and BA ($n = 4$). Our panel method classified one Barrett's-unrelated cancer as E-HME, 15 Barrett's-unrelated cancer as HME, 18 Barrett's-unrelated cancer and four BA as LME, and 5 Barrett's-unrelated cancer as E-LME

validated by pyrosequencing. Pyrosequencing primers were designed for six representative marker genes (Fig. 2b (upper) and Supplementary Table 2), in addition to *MLH1* (known to be a preferentially methylated gene in some HME cases [12]), to segregate four MEs: two marker genes (*CAV1* and *IRS2*) for the E-HME marker group, two markers (*ATRNL1* and *SLC1A2*) for the HME marker group and two markers (*GRIK1* and *UNC5C*) for the LME marker group. As shown in Fig. 2b (lower), the pyrosequencing results validated Infinium data.

Among 35 NGM, 31 samples had histopathological findings consistent with those of *H. pylori* gastritis; the remaining four samples did not exhibit any inflammation induced by *H. pylori* infection. According to our panel method mentioned below, all 31 gastritis mucosal samples were classified as LME, and no tissues were classified as E-LME. Three NBM samples were also distributed into LME. Four non-inflammatory mucosa samples were all classified as E-LME (Fig. 2c).

Among cancerous specimens, 42 samples were successfully determined as one of four MEs (Fig. 2d). Four BA cases were all subtyped as LME. One BA sample showed aberrant DNA methylation (> 70%) at all four CpG sites in *MLH1,* whereas HME marker genes were scarcely methylated. Of 42 non-BA adenocarcinomas, 39 cases were divided into one E-HME, 15 HME, 18 LME, and five E-LME.

Panel-based epigenotyping of 84 clinical samples using pyrosequencing data

To determine MEs quickly without complicated hierarchical clustering from large-scale Infinium data, we developed a novel model to classify clinical samples into four established MEs using pyrosequencing data based on a panel system (Fig. 3a). Samples with methylation levels > 25% by pyrosequencing data were considered "methylated." Three panels were evaluated as follows: (i) at least one of *GRIK1* and *UNC5C* is methylated, (ii) at least one of *ATRNL1* and *SLC1A2* is methylated, and (iii) at least one of *CAV1* and *IRS2* is methylated. If a sample fulfilled all these criteria, it was diagnosed as E-HME. If a sample met (i) and (ii) but not (iii), it was interpreted as HME, regardless of *MLH1* methylation. If a sample fulfilled (i) but neither (ii) nor (iii), it was diagnosed as LME only when none of the markers were methylated in *CAV1* and *IRS2*. It was interpreted as E-LME, if a sample fulfilled none of these three criteria, i.e. only when none of the markers were methylated in *ATRNL1*, *SLC1A2*, *CAV1,* and *IRS2*.

When the panel method was applied to the validation pyrosequencing data of 62 clinical samples used in the Infinium assay, all samples were successfully classified into one of the four established MEs, with 94% positive accuracy

(58/62). Based on these steps, we thus sought to characterize 38 non-tumor mucosa (35 NGM and three NBM) and 46 GC/AEG cases (including four BA cases).

Relationship between MEs and clinicopathological factors in 94 GC/AEG cases

From the above process, 94 GC/AEG cases were divided into four MEs as follows: 51 cases with hierarchical clustering using Infinium data and 43 cases with epigenotyping using pyrosequencing data. The association between MEs and clinicopathological factors among 94 tumors is shown in Table 1 and Supplementary Table 4. Immunohistochemical deficiency of MLH1 was seen only in the HME group. In the E-HME group, aberrant expression of p53 was relatively less frequent.

Survival analysis

The median follow-up period for 94 patients was 45.8 months. We analyzed overall survival according to four MEs and found significant differences ($P = 0.03$, Fig. 3b). When comparing the E-LME group and other MEs in a dichotomous fashion, patients in the E-LME group showed a significantly shortened prognosis ($P = 0.003$, Fig. 3c). Moreover, univariate Cox regression analysis revealed that lymph node metastasis, lymphatic invasion, venous invasion, and the E-LME profile were significantly associated with shortened overall survival (Table 2). Subsequent multivariate Cox regression analysis displayed an independent prognostic value of E-LME (Hazard ratio 3.77, 95% CI 1.63–8.74; $P = 0.0020$, Table 2).

Unsupervised two-way hierarchical clustering using TCGA data

Unsupervised two-way hierarchical clustering analysis in 561 cases, including stomach adenocarcinoma (symbolized as STAD in TCGA), esophageal adenocarcinoma (ESAD) (symbolized as ESCA in TCGA, excluding squamous cell carcinoma), and our original cases (62 GC/AEG cases) was performed on the 4,837 probes used in Fig. 1a (Fig. 4a). Gene probes were extracted with SD less than 0.05. The association between MEs and clinicopathological factors among 430 tumors is shown in Supplementary Table 5. Consistent with our results, STAD/ESAD combined with our GC/AED cases were successfully clustered into four MEs. In TCGA data, STAD cases were categorized into four molecular subtypes; *EBV*, EBV-positive; *MSI*, microsatellite instability; *CIN*, chromosomal instability; *GS*, genomically stable [19]. HME included MSI-high (MSI-H) and MSS/MSI-low (MSI-L) cases subdivided into HME_MSI-H and HME_MSS. HME_MSS and E-LME included ESAD significantly more than other MEs (Fig. 4b and Supplementary Table 5). As E-HME and
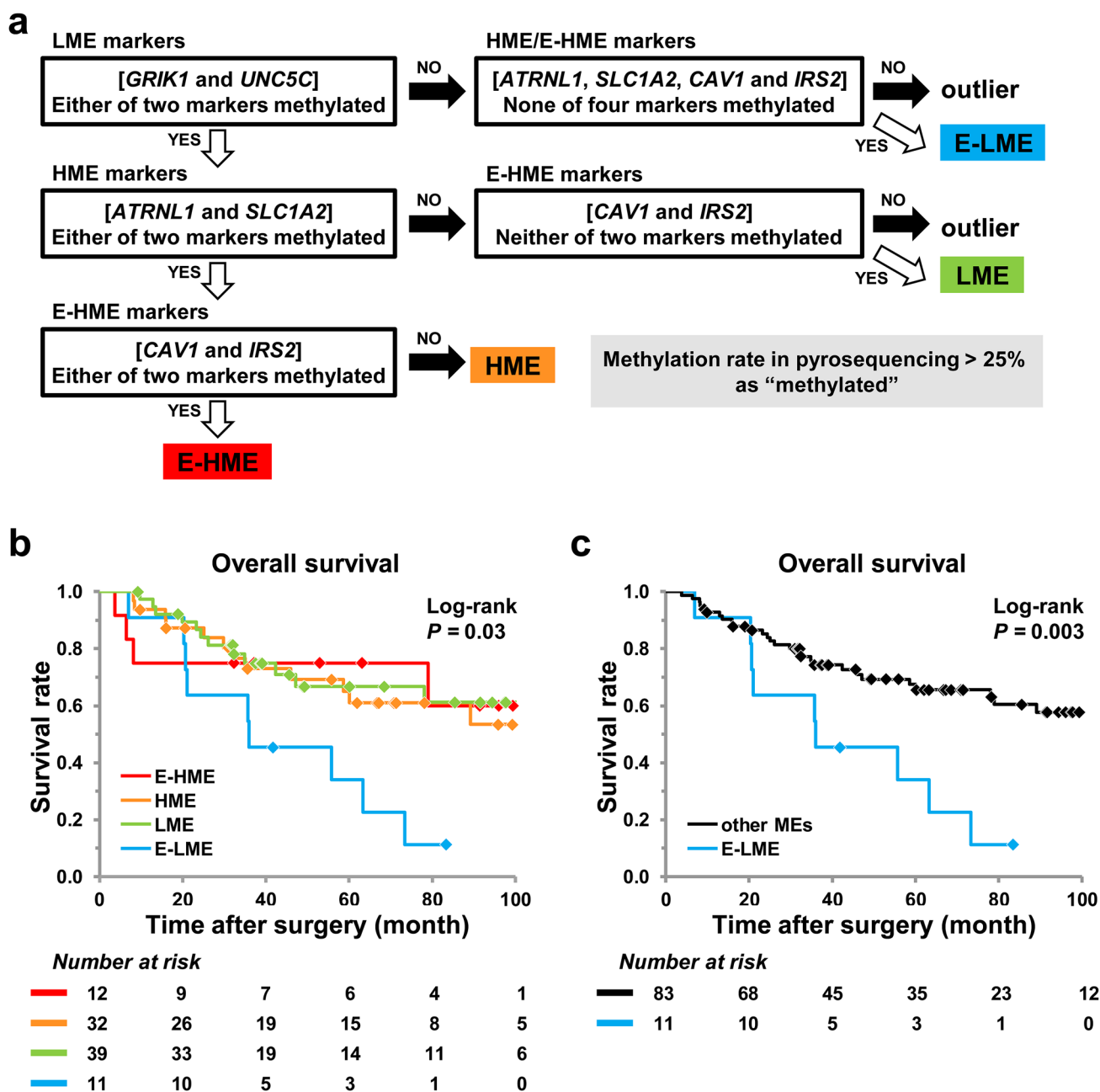
Fig. 3 A novel model to classify clinical samples into four established epigenotypes based on pyrosequencing data. **a** A panel model to classify clinical samples into four established epigenotypes using pyrosequencing data to determine cancer epigenotypes without hierarchical clustering from Infinium assays. **b** Kaplan–Meier curves for overall survival among four epigenotypes. **c** Kaplan–Meier curves for overall survival in a dichotomous manner (E-LME *vs*. other MEs)

HME_MSI-H almost corresponded with *EBV* and *MSI*, we evaluated the molecular subtypes of other MEs (HME_MSS, LME, and H-LME). LME included significantly more *GS* than the total ratio (Fig. 4c). However, the ratio of *CIN* and *GS* in HME_MSS and E-LME showed no significant difference. Overall survival was analyzed according to E-LME *vs*. other MEs in STAD data in Stages I and II (*P* = 0.03, Fig. 4d).

Discussion

We performed genome-wide DNA methylation analysis in GC and AEG, combined with background NM, based on our previous results [12], and obtained suggestive results.

First, Infinium analysis divided tissues from AEG and background NM into four DNA MEs. Some background

Table 1 Relationship between four MEs and clinicopathological factors

Variables	Total	E-HME	HME	LME	E-LME	P-value
# of cases	94	12	32	39	11	
Age at surgery (years)						
≤64	30	3	9	13	5	0.7
≥65	64	9	23	26	6	
Sex						
Male	77	10	23	34	10	0.3
Female	17	2	9	5	1	
Location						
Antrum	8	1	6	1	0	0.2
Body	37	8	9	15	5	
EGJ	49	3	17	23	6	
Lauren histology						
Intestinal	49	6	21	14	8	0.2
Mixed	40	5	10	22	3	
Diffuse	5	1	1	3	0	
Muscular invasion						
No	18	5	3	8	2	0.1
Yes	76	7	29	31	9	
Nodal metastasis						
No	34	6	12	12	4	0.7
Yes	60	6	20	27	7	
Barrett's-related						
No	87	12	29	35	11	0.5
Yes	7	0	3	4	0	
Lymphatic invasion						
No	60	4	20	28	8	0.2
Yes	34	8	12	11	3	
Venous invasion						
No	78	8	27	33	10	0.7
Yes	16	4	5	6	1	
Helicobacter pylori-infected						
No	13	0	4	7	2	0.2
Yes	81	12	35	25	9	
EBER-ISH						
Negative	82	0	32	39	11	<0.001[*]
Positive	12	12	0	0	0	
IHC for MLH1						
Proficient	84	12	22	39	11	<0.001[*]
Deficient	10	0	10	0	0	
IHC for TP53						
Wild type	36	9	13	10	4	0.02[*]
Aberrant	58	3	19	29	7	

EGJ esophagogastric junction, *EBER-ISH* Epstein–Barr virus-encoded small RNA in situ hybridization, *IHC* immunohistochemistry, *E-HME* extremely high methylation epigenotype, *HME* high methylation epigenotype, *LME* low methylation epigenotype, *E-LME*, extremely low methylation epigenotype

[*]$P < 0.05$, χ^2 test

NM and GC/AEG showed genome-wide methylation levels partially equal to those of the cancers in LME and E-LME.

Second, the DNA methylation profiles of background NM seemed to be divided, corresponding to their microscopic

Table 2 Univariate and multivariate Cox regression analyses for overall survival among 94 patients

| | Overall survival | | | |
| | Univariate | | Multivariate | |
	HR (95% CI)	*P*-value	HR (95% CI)	*P*-value
Age at surgery (years)				
≤ 64 vs. ≥ 65	1.42 (0.69–2.94)	0.3	1.71 (0.79–3.71)	0.18
Sex				
Male vs. Female	1.41 (0.59–3.38)	0.4	1.03 (0.37–2.86)	0.96
Tumor location				
Antrum vs. Body vs. EGJ	1.27 (0.77–2.12)	0.4	1.22 (0.67–2.24)	0.51
Lauren histology				
Intestinal vs. Mixed vs. Diffuse	1.13 (0.81–1.58)	0.5	1.06 (0.75–1.52)	0.73
Muscular invasion				
No vs. Yes	2.94 (0.90–9.58)	0.074	1.22 (0.3–4.89)	0.78
Nodal metastasis				
No vs. Yes	2.62 (1.58–8.28)	0.0023*	1.89 (0.68–5.23)	0.22
Lymphatic invasion				
No vs. Yes	4.12 (1.72–9.90)	0.0015*	2.52 (0.82–7.74)	0.11
Venous invasion				
No vs. Yes	2.96 (1.39–6.33)	0.047*	1.66 (0.19–14.5)	0.65
Epigenotype				
Other than E-LME vs. E-LME	2.96 (1.39–6.33)	0.005*	3.77 (1.63–8.74)	0.0020*

HR hazard ratio, *CI* confidence interval, *EGJ* esophagogastric junction, *E-LME* extremely low methylation epigenotype

*$P < 0.05$

chronic inflammatory changes. This finding suggested that chronic continuous inflammation, regardless of the cause (*H. pylori* infection or gastric juice reflux over EGJ), could play an essential role in the accumulation of aberrant DNA methylation in both gastric and EGJ mucosa.

We also conducted pyrosequencing using additional clinical samples. Although the information from these experiments was limited, some insights supported the results from the Infinium analysis. As mentioned above, gastritis appeared to induce LME in cancer-free mucosa. Regarding Barrett's mucosa and existing cancer in EGJ, the genes in the LME marker group showed high DNA methylation levels. Overall, whether for GC or AEG, non-specific inflammation (not necessarily accompanied by *H. pylori* infection) could be a crucial player in forming the epigenetic cancer field. Curiously, one BA showed *MLH1* methylation without HME, which, even in similar cases, had not been described in any previous publications. This finding may suggest a novel pattern of DNA methylation induction. However, the results were highly anecdotal, and further collections of larger-sized data are needed before any definitive conclusions can be drawn.

On the conceptual background that inflammation may induce accumulation of DNA methylation related to carcinogenesis in the esophagogastric columnar epithelium, methylation levels of normal mucosa can be helpful

prognosticators for AEG. A recent prospective study underlined that estimation of methylation in three marker genes (*miR-124a-3*, *EMX1*, and *NKX6-1*) was clinically valuable in predicting the development of metachronous gastric cancer [10, 11]; they found *miR-124a-3* to be the most potent predictor among the three genes. Our research indicated that MIR124-3 was semiquantitatively similar to LME marker genes (Supplementary Fig. 2), which can be estimated as the basis for the predictive impact of *miR-124a-3*. Similarly, our study's combination of 319 genes extracted as LME markers (Supplementary Table 1) may be promising as valuable prognostic indicators in AEG.

The development of E-LME tumors, however, may be difficult to predict using mucosal DNA methylation levels, as they naturally show low methylation levels. Of 11 E-LME tumors in our cohort, nine had *H. pylori* gastritis in their background mucosa (Table 1), suggesting that E-LME cancers might arise from scarcely methylated epithelial cells even when accompanied by *H. pylori* colonization. E-LME might be a unique and specific subtype in GC and AEG, not associated with the elevation of DNA methylation levels derived from inflammation. Consequently, one concern is that background mucosal DNA methylation levels could not be used for evaluating susceptibility to E-LME tumors, the long-term outcome of which was significantly shortened even in multivariate analysis (Fig. 3b–c, Table 2). Although
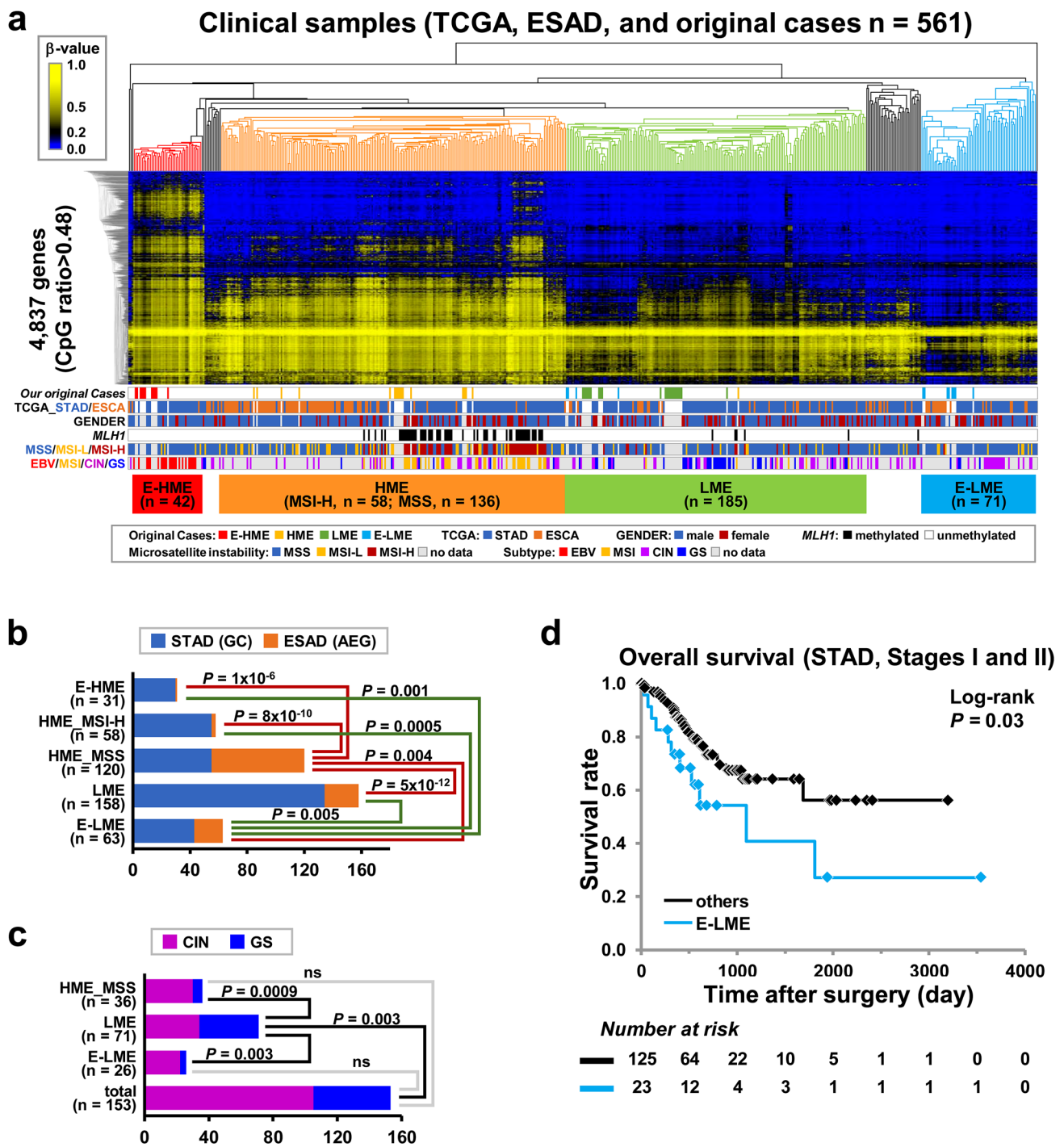
Fig. 4 DNA methylation analysis in TCGA data of GC (STAD) and adenocarcinoma of esophageal cancer (ESAD) combined with that of our original samples. **a** Unsupervised two-way hierarchical clustering revealed four distinct MEs among 561 clinical samples: E-HME ($n=42$), HME with MSI-H ($n=58$), HME with MSS/MSI-L ($n=136$), LME ($n=185$) and E-LME ($n=71$). In total, 4,837 probes extracted from high- and intermediate-CpG probes were used for clustering. The information for each sample is presented in the lower columns. **b** Ratios of STAD or ESCA are shown by bar graphs. **c** Ratios of molecular subtypes (chromosomal instability (CIN) and genomically stable (GS)) are shown in the bar graph. **d** Ratios of gender are shown by bar graph. **e** Kaplan–Meier curves for overall survival in a dichotomous manner (E-LME *vs.* the other MEs)

several past reports have highlighted the predictive impact of EBV infection or *MLH1* promoter methylation in GC [33–35], their prognosticating potential in cancer without aberrant DNA methylation is unclear, and it is thus difficult to provide plausible explanations for our results. One promising possibility is that the 153 E-LME marker genes extracted might be more valuable prognostic indicators for esophagogastric adenocarcinoma, including E-LME tumors, compared with LME markers. However, further investigation (if possible, prospective) is needed to confirm this suggestion.

Finally, large-scale data from TCGA was reviewed to compensate for our small number of cases. STAD and ESAD combined with our original data were successfully clustered into four MEs. E-LME, a novel ME in this report, included STAD and ESAD, and the leading molecular basis of E-LME was CIN. Interestingly HME without MSI (HME_MSS) was enriched in ESAD. On the contrary, HME_MSI was relatively rare in ESAD.

The overall survival of E-LME was significantly poorer than that of other MEs in early-stage GC cases, i.e., Stages I and II. As STAD cases with Stages III and IV and most ESCA cases in TCGA data (other than Japanese countries) generally had poor prognoses, we could not identify any significant differences in these advanced cases. When overall survival of 94 Japanese AEG/GC cases was analyzed separately for 49 AEG and 45 GC cases, both AEG and GC tended to show poorer prognosis in E-LME cases than other ME cases (Supplementary Fig. 3), though the sample size and the number of events were small. For precise evaluation of E-LME in AEG and GC, more cases at a similar therapeutic level are desired to be accumulated for the analysis.

In summary, we performed genome-wide DNA methylation analysis among GC, AEG, and background NM. These could be divided into four distinguished MEs. Background NM occasionally revealed genome-wide accumulation of DNA methylation comparable with that in cancers. Histopathological inflammation, caused by *H. pylori* infection or reflux esophagitis, could play an essential role in the accumulation of aberrant DNA methylation in gastric and esophagogastric columnar mucosae. E-LME, a novel ME, was possibly unrelated to the accumulation of aberrant DNA methylation and exhibited a poor prognosis.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 Countries. CA Cancer J Clin. 2021;71:209–49.
2. Ferro A, Peleteiro B, Malvezzi M, Bosetti C, Bertuccio P, Levi F, et al. Worldwide trends in gastric cancer mortality (1980–2011), with predictions to 2015, and incidence by subtype. Eur J Cancer. 2014;50:1330–44.
3. Bertuccio P, Chatenoud L, Levi F, Praud D, Ferlay J, Negri E, et al. Recent patterns in gastric cancer: a global overview. Int J Cancer. 2009;125:666–73.
4. Uemura N, Okamoto S, Yamamoto S, Matsumura N, Yamaguchi S, Yamakido M, et al. Helicobacter pylori infection and the development of gastric cancer. N Engl J Med. 2001;345:784–9.
5. . IARC working group on the evaluation of carcinogenic risks to humans: some industrial chemicals. Lyon, 15–22 February 1994. IARC Monogr Eval Carcinog Risks Hum. 1994;60:1–560.
6. 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):995.
7. Kaneda A, Kaminishi M, Yanagihara K, Sugimura T, Ushijima T. Identification of silencing of nine genes in human gastric cancers. Cancer Res. 2002;62:6645–50.
8. Chan AO, Lam SK, Wong BC, Wong WM, Yuen MF, Yeung YH, et al. Promoter methylation of E-cadherin gene in gastric mucosa associated with helicobacter pylori infection and in gastric cancer. Gut. 2003;52:502–6.
9. Kang GH, Lee HJ, Hwang KS, Lee S, Kim JH, Kim JS. Aberrant CpG island hypermethylation of chronic gastritis, in relation to aging, gender, intestinal metaplasia, and chronic inflammation. Am J Pathol. 2003;163:1551–6.
10. 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.
11. 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.
12. Matsusaka K, Kaneda A, Nagae G, Ushiku T, Kikuchi Y, Hino R, et al. Classification of Epstein-Barr virus-positive gastric cancers by definition of DNA methylation epigenotypes. Cancer Res. 2011;71:7187–97.
13. Chong JM, Sakuma K, Sudo M, Ushiku T, Uozaki H, Shibahara J, et al. Global and non-random CpG-island methylation in gastric carcinoma associated with Epstein-Barr virus. Cancer Sci. 2003;94:76–80.
14. Sakuma K, Chong JM, Sudo M, Ushiku T, Inoue Y, Shibahara J, et al. High-density methylation of p14ARF and p16INK4A in Epstein-Barr virus-associated gastric carcinoma. Int J Cancer. 2004;112:273–8.

15. Ushiku T, Chong JM, Uozaki H, Hino R, Chang MS, Sudo M, et al. p73 gene promoter methylation in Epstein-Barr virus-associated gastric carcinoma. Int J Cancer. 2007;120:60–6.

16. Chang MS, Uozaki H, Chong JM, Ushiku T, Sakuma K, Ishikawa S, et al. CpG island methylation status in gastric carcinoma with and without infection of Epstein-Barr virus. Clin Cancer Res. 2006;12:2995–3002.

17. Kang GH, Lee S, Kim WH, Lee HW, Kim JC, Rhyu MG, et al. Epstein-barr virus-positive gastric carcinoma demonstrates frequent aberrant methylation of multiple genes and constitutes CpG island methylator phenotype-positive gastric carcinoma. Am J Pathol. 2002;160:787–94.

18. Wang K, Yuen ST, Xu J, Lee SP, Yan HH, Shi ST, et al. Whole-genome sequencing and comprehensive molecular profiling identify new driver mutations in gastric cancer. Nat Genet. 2014;46:573–82.

19. Cancer genome atlas research N. Comprehensive molecular characterization of gastric adenocarcinoma. Nature. 2014;513:202–9.

20. Cancer genome atlas research N, Analysis working group: Asan U, Agency BCC, Brigham, Women's H, Broad I, et al. Integrated genomic characterization of oesophageal carcinoma. Nature. 2017;541:169–75.

21. Vial M, Grande L, Pera M. Epidemiology of adenocarcinoma of the esophagus, gastric cardia, and upper gastric third. Recent Results Cancer Res. 2010;182:1–17.

22. Kusano C, Gotoda T, Khor CJ, Katai H, Kato H, Taniguchi H, et al. Changing trends in the proportion of adenocarcinoma of the esophagogastric junction in a large tertiary referral center in Japan. J Gastroenterol Hepatol. 2008;23:1662–5.

23. Nicolas C, Sylvain M, Come L, Jean F, Anne-Marie B, Valerie J. Trends in gastric cancer incidence: a period and birth cohort analysis in a well-defined French population. Gastric Cancer. 2016;19:508–14.

24. de Jonge PJ, van Blankenstein M, Grady WM, Kuipers EJ. Barrett's oesophagus: epidemiology, cancer risk and implications for management. Gut. 2014;63:191–202.

25. Rokkas T, Pistiolas D, Sechopoulos P, Robotis I, Margantinis G. Relationship between helicobacter pylori infection and esophageal neoplasia: a meta-analysis. Clin Gastroenterol Hepatol. 2007;5:1413–7.

26. Urabe M, Ushiku T, Shinozaki-Ushiku A, Iwasaki A, Yamazawa S, Yamashita H, et al. Adenocarcinoma of the esophagogastric junction and its background mucosal pathology: a comparative analysis according to Siewert classification in a Japanese cohort. Cancer Med. 2018;7:5145–54.

27. Chong IY, Cunningham D, Barber LJ, Campbell J, Chen L, Kozarewa I, et al. The genomic landscape of oesophagogastric junctional adenocarcinoma. J Pathol. 2013;231:301–10.

28. Zhang CD, Takeshima H, Sekine S, Yamashita S, Liu YY, Hattori N, et al. Prediction of tissue origin of adenocarcinomas in the esophagogastric junction by DNA methylation. Gastric Cancer. 2021;25:336–46.

29. Iwasaki Y, Chong JM, Hayashi Y, Ikeno R, Arai K, Kitamura M, et al. Establishment and characterization of a human Epstein-Barr virus-associated gastric carcinoma in SCID mice. J Virol. 1998;72:8321–6.

30. 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.

31. Matsusaka K, Funata S, Fukuyo M, Seto Y, Aburatani H, Fukayama M, et al. Epstein-Barr virus infection induces genome-wide de novo DNA methylation in non-neoplastic gastric epithelial cells. J Pathol. 2017;242:391–9.

32. Nazor KL, Altun G, Lynch C, Tran H, Harness JV, Slavin I, et al. Recurrent variations in DNA methylation in human pluripotent stem cells and their differentiated derivatives. Cell Stem Cell. 2012;10:620–34.

33. Camargo MC, Kim WH, Chiaravalli AM, Kim KM, Corvalan AH, Matsuo K, et al. Improved survival of gastric cancer with tumour Epstein-Barr virus positivity: an international pooled analysis. Gut. 2014;63:236–43.

34. Li YZ, Yang YS, Lu YY, Herman JG, Brock MV, Zhao P, et al. Predictive value of CHFR and MLH1 methylation in human gastric cancer. Gastric Cancer. 2015;18:280–7.

35. Shigeyasu K, Nagasaka T, Mori Y, Yokomichi N, Kawai T, Fuji T, et al. Clinical significance of MLH1 methylation and CpG island methylator phenotype as prognostic markers in patients with gastric cancer. Plos One. 2015;10: e0130409.