ORIGINAL ARTICLE

Altered gene expression due to aberrant DNA methylation correlates with responsiveness to anti-EGFR antibody treatment

Yasufumi Otsuki[1,2] | Kota Ouchi[1,2] | Shin Takahashi[1,2] | Keiju Sasaki[1,2] | Yasuhiro Sakamoto[3] | Akira Okita[3] | Chikashi Ishioka[1,2,4]

[1]Department of Clinical Oncology, Institute of Development, Aging and Cancer, Tohoku University, Miyagi, Japan

[2]Department of Medical Oncology, Tohoku University Hospital, Miyagi, Japan

[3]Department of Medical Oncology, Osaki Citizen Hospital, Miyagi, Japan

[4]Department of Clinical Oncology, Tohoku University Graduate School of Medicine, Miyagi, Japan

Correspondence
Chikashi Ishioka, Department of Clinical Oncology, Institute of Development, Aging and Cancer, Tohoku University, 4-1 Seiryo-machi, Aobaku, Sendai 980-8575, Japan.
Email: chikashi@tohoku.ac.jp

Abstract

The cetuximab gene expression signature and DNA methylation status of colorectal cancer (CRC) are predictive of the therapeutic effects of anti-epidermal growth factor receptor (EGFR) antibody therapy. As DNA methylation is a means of regulating gene expression, it may play an important role in the expression of cetuximab signature genes. This study aims to determine the effects of aberrant DNA methylation on the regulation of cetuximab signature gene expression. Comprehensive DNA methylation and gene expression data were retrieved from CRC patients in three tumor tissue (TT) cohorts and three normal colorectal mucosa/tumor tissue paired (NCM-TT) cohorts. Of the 231 cetuximab signature genes, 57 exhibited an inverse correlation between the methylation of promoter CpG sites and gene expression level in multiple cohorts. About two-thirds of the promoter CpG sites associated with the 57 genes exhibited this correlation. In all 57 gene promoter regions, the methylation levels in NCMs did not differ according to comparisons based on cetuximab signature or DNA methylation status classification of matched TTs. Thus, the altered expression of 57 genes was caused by aberrant DNA methylation during carcinogenesis. Analysis of the association between cetuximab signature or DNA methylation status and progression-free survival (PFS) of anti-EGFR antibody agents in the same cohort showed that DNA methylation status was most associated with PFS. In conclusion, we found that aberrant DNA methylation regulates specific gene expression in cetuximab signature during carcinogenesis, suggesting that it is one of the important determinants of sensitivity to anti-EGFR antibody agents.

KEYWORDS
colorectal neoplasms, DNA methylation, ErbB receptors, genetic promoter regions, transcriptomes

---

---

1 | INTRODUCTION

The incidence of colorectal cancer (CRC) has increased in recent years, accounting for the third-largest number of cancers and second-largest number of deaths worldwide.[1] Treatment with anti-EGFR antibodies (e.g., cetuximab) is recommended for patients with *RAS* wild-type metastatic CRC (mCRC).[2–5] However, less than half of these patients respond to anti-EGFR antibody treatment. The predictive factors that determine the responsiveness of CRC to anti-EGFR antibody therapy are under investigation.

DNA methylation is a major epigenetic regulatory mechanism. The methylation of CpG islands in the promoter region upstream of a transcription start site (TSS) suppresses gene expression.[6,7] In CRC, about 20% of the cases are positive for the CpG island methylator phenotype (CIMP). In these cases, aberrant DNA methylation accumulates at many loci in CpG islands, resulting in specific clinical features. Meta-analyses have concluded that CIMP is a poor prognostic factor for CRC.[8,9] Ouchi et al. reported that genome-wide DNA methylation status can extract more hypermethylated CRC than classical CIMP markers and is a predictor of the efficacy of anti-EGFR antibody therapy, with low-methylated CRC (LMCC), but not highly methylated CRC (HMCC), exhibiting sensitivity to anti-EGFR antibodies.[10]

The cetuximab signature, previously reported by Schutte et al., is a gene set that predicts the therapeutic effect of cetuximab as determined by a comprehensive gene expression analysis using a patient-derived CRC xenograft model. The cetuximab signature consists of a set of genes that are upregulated in responders (103 genes) and non-responders (138 genes).[11] The mechanism underlying the regulation of cetuximab signature genes is unknown.

Based on these previous observations, we hypothesized that aberrant DNA methylation of genes in the cetuximab signature alters their expression, thereby determining the therapeutic effect of anti-EGFR antibodies. This study aims to determine the significance of aberrant DNA methylation in the regulation of cetuximab signature gene expression using comprehensive DNA methylation and gene expression data from multiple cohorts.

2 | MATERIALS AND METHODS

2.1 | Patient cohorts

The three tumor tissue (TT) cohorts used in this study are as follows: (1) TUH cohort, mCRC patients treated with anti-EGFR antibody at Tohoku University Hospital (TUH) and National Cancer Center Hospital; (2) TCGA cohort, CRC patient data obtained from The Cancer Genome Atlas (TCGA, https://portal.gdc.cancer.gov/); and (3) RBWH cohort, data from CRC patients at the Royal Brisbane and Women's Hospital (RBWH) obtained from ArrayExpress (https://www.ebi.ac.uk/arrayexpress/).[12]

The three normal colorectal mucosa tumor tissue (NCM-TT) paired cohorts also used in this study are as follows: (1) OCH paired cohort, mCRC patients treated with anti-EGFR antibody therapies at Osaki Citizen Hospital (OCH); (2) TCGA paired cohort, CRC patient data obtained from TCGA; and (3) RBWH paired cohort, data from CRC patients at RBWH obtained from ArrayExpress.[12]

The study protocol complies with the Declaration of Helsinki and the Ethical Guidelines for Medical and Health Research Involving Human Subjects. This study was approved by the ethics committee of Tohoku University, National Cancer Center Hospital, and Osaki Citizen Hospital. Patients who could not provide consent for this study due to death or other reasons were handled by opt-out.

2.2 | Comprehensive gene expression analysis

In the TUH cohort, we used the Whole Human Genome Oligo Microarray kit (4 × 44 K) (Agilent Technologies).[13] The raw data were normalized to a signal value of the 75th percentile of each probe. In the OCH paired cohorts, we used the Clariom™ Human D Pico Assay (ThermoFisher Scientific). The raw data were normalized using Signal Space Transformation-Robust Multiarray Analysis (sst-RMA).[14,15] Microarray data are available from GSE104645 for the TUH cohort and from GSE185770 for the OCH paired cohort.

In the TCGA and TCGA paired cohort, fragments per kilobase of exon per million mapped reads (FPKM) normalized data were obtained from the Genomic Data Commons (GDC) data portal (https://portal.gdc.cancer.gov/) in June 2020 and converted to transcripts per million (TPM) for analysis.[16] In the RBWH and RBWH paired cohort, processed data were obtained from the E-MTAB-8148 dataset in ArrayExpress in December 2020.

In the TUH, TCGA, and RBWH cohorts, baseline shift was performed to set the median of all samples to 0 for each probe. In the OCH, TCGA, and RBWH paired cohorts, the expression ratio was defined as [(expression level in TT) / (expression level in NCM)].

2.3 | Classification based on cetuximab signature

Expression data for the 241 genes in the cetuximab signature were extracted from the normalized comprehensive gene expression data for each cohort (Table S1). For each patient, Pearson's correlation coefficient for the cetuximab signature was calculated. Patients with positive correlations to the cetuximab signature were classified as responders, while the others were classified as non-responders.

2.4 | Gene mutation analysis

DNA direct sequencing was performed on the *KRAS* gene (codons 12 and 13) and *BRAF* gene (codon 600) in the TUH and OCH paired cohorts, and for *KRAS* and *NRAS* genes (codons 59, 61, 117, and 146)

in the OCH paired cohort.[17] For direct DNA sequencing, TaKaRa EX Taq (Takara Bio) was used to amplify the genes by PCR. PCR primers and conditions are shown in Tables S2 and S3. In the TUH cohort, the Luminex Assay using GENOSEARCH Mu-PACK (MBL) was performed for the *KRAS* gene (codons 61 and 146) and *NRAS* gene (codons 12, 13, and 61).

In the TCGA and TCGA paired cohorts, mutation data for the *KRAS* and *NRAS* genes (codons 12, 13, 59, 61, 117, 146) and *BRAF* gene (codon 600) were obtained from cBioportal (https://www. cbioportal.org/). In the RBWH and RBWH paired cohorts, mutation data for the *KRAS* gene (codons 12 and 13) and *BRAF* gene (codon 600) were obtained from the ArrayExpress E-MTAB-8148 dataset.

2.5 | Comprehensive DNA methylation analysis

Infinium Human Methylation 450 BeadChip (Illumina) and Infinium Methylation EPIC BeadChip (Illumina) were used for the TUH cohort and OCH paired cohort, respectively.[10] BeadChips were scanned using the BeadStation or iScan system. Methylation data with $p \geq 0.05$ were excluded. All samples achieved >95% CpG coverage. The raw data were normalized using the internal controls of the GenomeStudio software package. The methylation level at each CpG site was calculated using the Methylation Module attached to BeadStudio or GenomeStudio as follows: [β value: the number of methylated probes (intensity of the methylated signal) / (intensity of the methylated probes + intensity of the unmethylated probes) number of methylated probes + number of unmethylated probes].

In the TCGA and TCGA paired cohorts, normalized β values calculated using Infinium Human Methylation 450 BeadChip were obtained from the GDC data portal. In the RBWH and RBWH paired cohorts, normalized β values calculated using Infinium Human Methylation 450 BeadChip were obtained from the ArrayExpress E-MTAB-7036 dataset.

We defined probes with β value ≥0.3 as methylation positive and probes with β value <0.3 as methylation negative, as described by Fennell et al.[12] To exclude CpG sites with low β-value variability from the analysis, CpG sites with <10% or >90% methylation-positive cases in each cohort were excluded in the TT cohorts. To exclude CpG sites with little changes in β value between NCM and TT, CpG sites with <10% of cases with an absolute Δβ value [(β value of TT) − (β value of NCM)] ≥0.2 in each cohort were excluded in the NCM-TT paired cohorts.

2.6 | Classification based on DNA methylation status

Ouchi et al. reported 16 CpG sites used to determine the genome-wide DNA methylation status (Table S4).[18] Accordingly, we defined

the methylation status using these 16 CpG as follows: HMCC, ≥8 methylation-positive sites; LMCC, ≤7 methylation-positive sites.

2.7 | Evaluation of therapeutic effect

In the TUH cohort, progression-free survival (PFS) on anti-EGFR antibody therapy was used to evaluate therapeutic effect. Progression-free survival was defined from the start date of anti-EGFR antibody therapy to the date of imaging or clinical progression.

2.8 | Statistical analysis

Difference in methylation levels between samples in the NCM cohort were analyzed using the exactRankTests package in R. CpG sites with $p < 0.05$ for the Wilcoxon rank sum test and false discovery rate (FDR) <0.05 using the Benjamini & Hochberg method were defined as significant.

Statistical analysis of patient background and PFS in the TUH cohort were performed using JMP Pro 16 (SAS). Patient background was tested for significance using χ^2 test (or Fisher's exact probability test) and Wilcoxon rank sum test. The primary site was defined as right-sided colon for "cecum, ascending and transverse" and left-sided colon for "descending and sigmoid." The cetuximab signature and DNA methylation status were compared using Fisher's exact probability test. Survival curves were generated using the Kaplan-Meier method, and significant differences were identified using the log-rank test. Univariate and multivariate analyses using the Cox proportional hazards model were performed to identify the background factors contributing to PFS.

2.9 | Integrated analysis of gene expression data and DNA methylation data

The gene promoter region was defined as 0-200 bases (TSS200) and 200-1500 bases (TSS1500) upstream of the transcription start point in the annotation of UCSC_REFGENE_GROUP. In the TT cohorts, we analyzed the correlation between β values for the promoter region and expression levels of each gene in the cetuximab signature. Correlation analysis was performed using Pearson's correlation coefficient, with a significance level of $p < 0.05$. In the NCM-TT paired cohorts, we determined the correlation between changes in methylation level in the promoter region (Δβ value) and expression levels (expression ratio) between NCM and TT in the same patient. Correlation analysis was performed using Pearson's correlation coefficient, and absolute values of the correlation coefficient ≥0.3 were considered significant.

Details of Materials and Methods are provided in Supplementary Appendices.

3 | RESULTS

Flowcharts outlining the analysis of the results in 3.3-3.7 and 3.8-3.11 are shown in Figure 1A and B, respectively. The total number of patients in each cohort was as follows: TUH, 97; TCGA, 175; RBWH, 112; OCH paired, 9; TCGA paired, 13; and RBWH paired, 22.

3.1 | Classification based on cetuximab signature in each cohort

Based on the cetuximab signature, 43 (44.3%), 79 (45.1%), 56 (50.0%), 4 (44.4%), 5 (38.5%), and 15 (68.2%) patients were classified as non-responders in the TUH, TCGA, RBWH, OCH paired, TCGA paired, and RBWH paired cohorts, respectively (Figure S1).

3.2 | Classification based on DNA methylation status in each cohort

Based on the DNA methylation status at 16 CpG sites, 26 (26.8%), 25 (14.3%), 18 (16.1%), 4 (44.4%), 5 (38.5%), and 14 (63.6%) patients

were classified as HMCC in the TUH, TCGA, RBWH, OCH paired, TCGA paired, RBWH paired cohorts, respectively (Figure S1).

3.3 | Association between DNA methylation status and cetuximab signature classification in the TT cohorts

We analyzed the association between cetuximab signature and DNA methylation status in the TT cohorts. In the TUH, TCGA, and RBWH cohorts, 80.8%, 64.0%, and 77.8% of HMCC patients were classified as non-responders, respectively. HMCC patients tended to be more included in the non-responder group (Tables S5, S6, and S7).

3.4 | Correlation between promoter region β values and expression levels of cetuximab signature genes

To clarify the relationship between cetuximab signature genes and aberrant DNA methylation in the TT cohort, we analyzed the correlation between promoter region β values and gene expression levels in the TT cohorts. Target genes for analysis included 231 of the 241
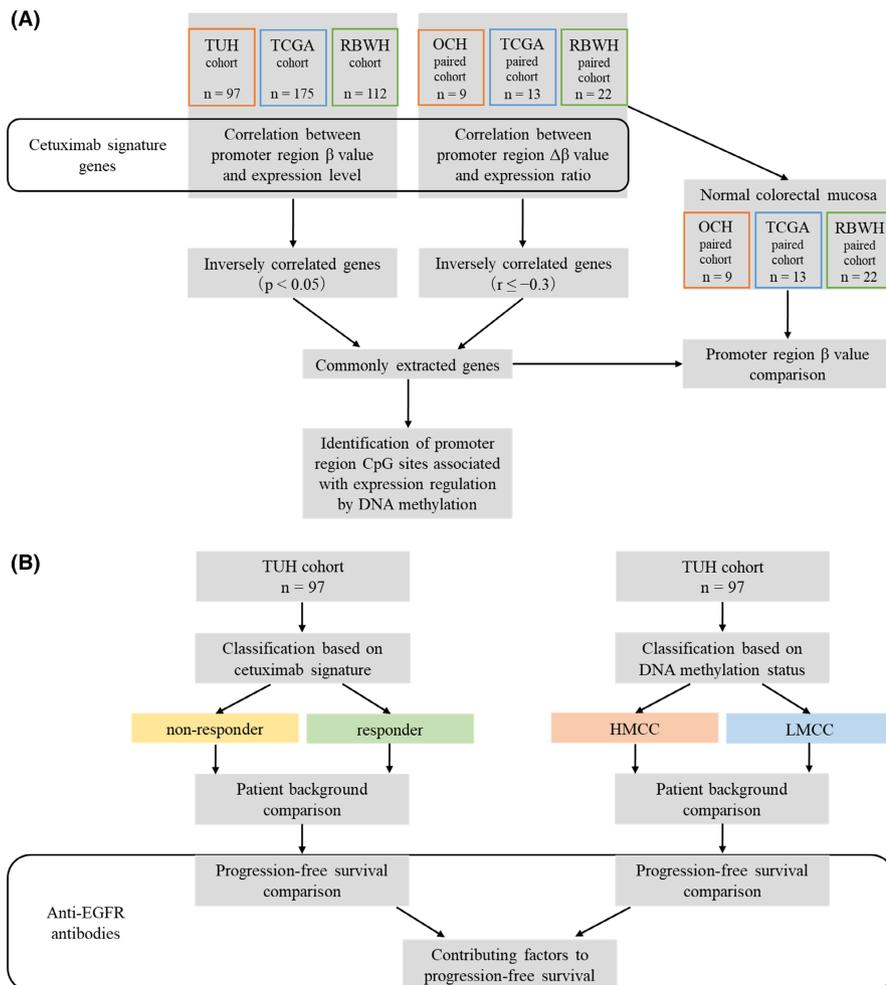

FIGURE 1 Flowcharts of analyses. A, Flowchart for analysis of the relationship between genes involved in the cetuximab signature and methylation regulation. B, Flowchart of the factors contributing to progression-free survival (PFS) in patients receiving anti-EGFR antibody therapy in the TUH cohort. EGFR, epidermal growth factor receptor; HMCC, highly methylated colorectal cancer; LMCC, low-methylated colorectal cancer; OCH, Osaki Citizen Hospital; RBWH, Royal Brisbane and Women's Hospital; TCGA, The Cancer Genome Atlas; TUH, Tohoku University Hospital

cetuximab signature genes, excluding seven genes on the X chromosome (*MAGEA11*, *MAP7D2*, *PLXNB3*, *SCML2*, *SMARCA1*, *SYTL5*, *TSPAN6*) and three genes (*ANXA10*, *KYNU*, *LYZ*) whose promoter regions were not probed in the methylation array.

A significant inverse correlation was observed between promoter region β values and gene expression levels in the following cohorts: TUH (51 genes), TCGA (59 genes), and RBWH (32 genes). Heatmaps of representative promoter region β values per gene and expression levels are shown for each cohort in Figure 2A–C.

In the TT cohorts, a total of 88 genes exhibited a significant inverse correlation between the promoter region β value and expression levels in at least one of three cohorts (Table S8).
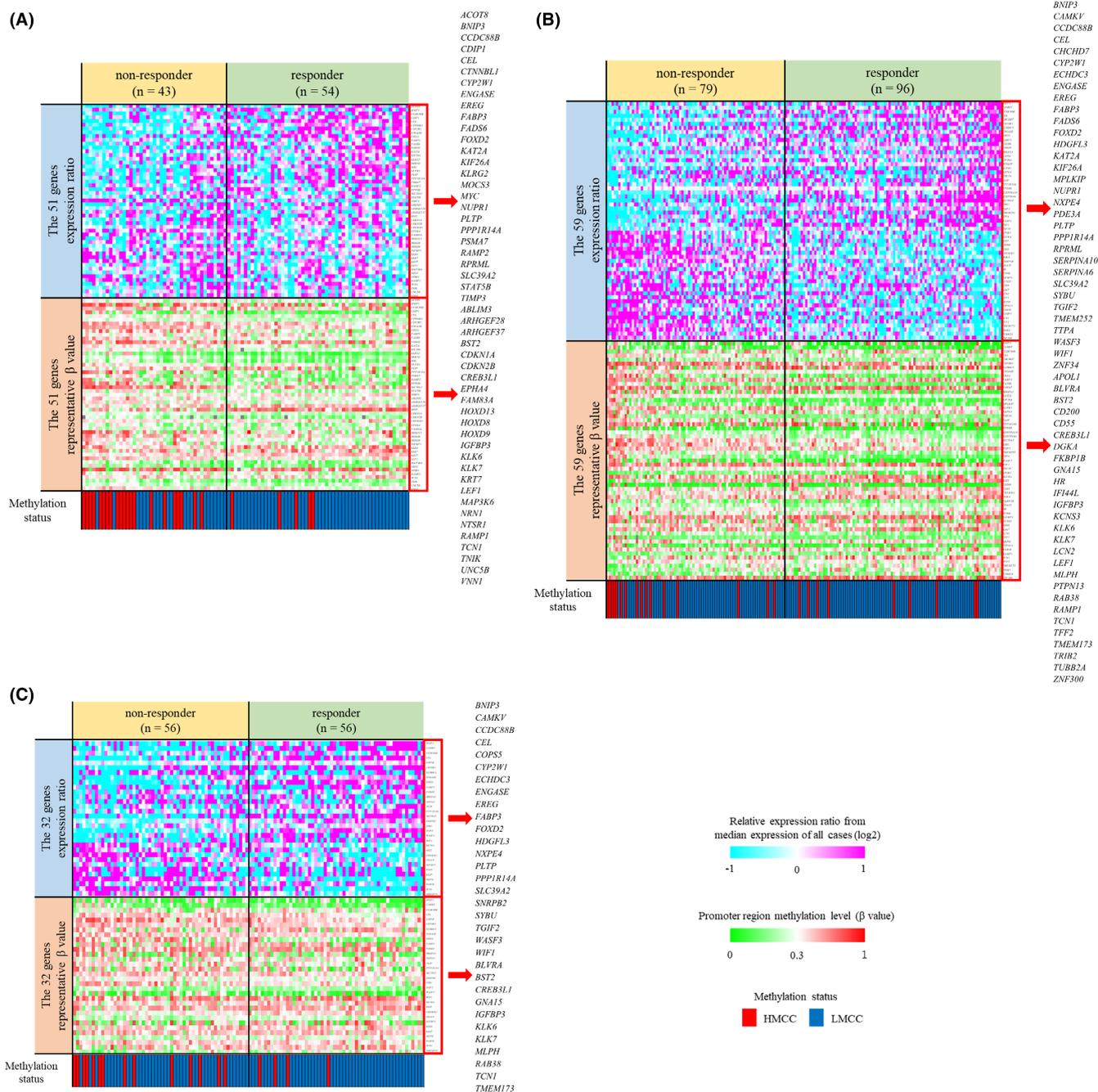

FIGURE 2 Genes with a significant inverse correlation between promoter region β values and expression levels in each cohort. Heatmaps of genes with a significantly inverse correlation between promoter region β values and expression levels in (A) TUH cohort, 51 genes; (B) TCGA cohort, 59 genes; and (C) RBWH cohort, 32 genes. Top: relative expression ratio (log2) from median expression of all patients for each gene. Middle: representative β values of promoter region for each gene (one CpG site per gene). Bottom: classification according to DNA methylation status. Right: the magnified figures of the extracted gene rows. HMCC, highly methylated colorectal cancer; LMCC, low-methylated colorectal cancer; RBWH, Royal Brisbane and Women's Hospital; TCGA, The Cancer Genome Atlas; TUH, Tohoku University Hospital

3.5 | Correlation between changes in promoter region methylation level (Δβ value) and gene expression (expression ratio) during carcinogenesis in paired samples

To extract genes with expression levels altered by aberrant methylation during carcinogenesis, we analyzed the correlation between changes in promoter region methylation level (Δβ value) and changes in gene expression (expression ratio) in the NCM-TT paired cohorts.

An inverse correlation was observed between promoter region Δβ value and expression ratio in the following cohorts: OCH paired (69 genes), TCGA paired (69 genes), and RBWH paired (47 genes). The heatmaps of representative promoter region Δβ values per gene and expression ratios are shown for each cohort in Figure 3A–C.

In the NCM-TT cohorts, a total of 109 genes had an inverse correlation between promoter region Δβ value and expression ratio in at least one of three cohorts (Table S9).

3.6 | Identification of CpG sites that regulate gene expression by aberrant methylation

The results in Sections 3.4 and 3.5 reveal 57 genes in common between the two analyses (Table 1). For each of the six cohorts, a correlation coefficient was determined for each of the CpG sites in the promoter region of the 57 genes (Table S10). CpG sites that are important for the regulation of gene expression must show a consistent inverse correlation between DNA methylation status and expression levels in the independent analysis of each of the six cohorts. Therefore, CpG sites with ≥75.0% concordance in correlation coefficients among the multiple cohorts were considered important candidate CpG sites for gene regulation. Of the total 447 CpG sites of 57 genes, 285 (63.8%) showed a consistent inverse correlation between promoter region methylation status and expression level (Figure 4).

3.7 | Comparison of promoter DNA methylation levels in the 57 genes in NCM

To determine whether aberrant DNA methylation is already present in NCM or arises during carcinogenesis, the β values for all promoter CpG sites in the 57 genes were compared between the two groups in NCMs based on the cetuximab signature or DNA methylation status classification of matched TTs. No CpG sites differed significantly in β values in either comparison (Figure S2).

3.8 | Comparison of patient backgrounds according to cetuximab signature

Patient backgrounds were compared between non-responders and responders in the TUH cohorts (Table S11). The percentage of right-sided colon cases was significantly higher in non-responders than in responders ($p < 0.01$). No other patient background factors differed significantly between the groups.

3.9 | Comparison of patient backgrounds according to DNA methylation status

Patient backgrounds were compared between the HMCC and LMCC groups in the TUH cohorts (Table S12). The percentage of right-sided colon cases was significantly higher in the HMCC than in the LMCC group ($p < 0.01$). No other patient background factors differed significantly between the groups.

3.10 | Predictive value of cetuximab signature and DNA methylation status for the efficacy of anti-EGFR antibody therapy

To investigate the efficacy of anti-EGFR antibody therapy according to cetuximab signature and DNA methylation status, we compared the PFS after anti-EGFR antibody therapy between each group (Figure 5). According to the cetuximab signature, the PFS of the non-responder group was significantly lower than that of the responders ($p < 0.0001$). According to DNA methylation status, the PFS was significantly lower among patients with HMCC than LMCC ($p < 0.0001$).

3.11 | Factors contributing to the PFS of anti-EGFR antibody therapy

To identify factors contributing to PFS of anti-EGFR antibody therapy in the TUH cohorts, we performed univariate and multivariate analyses of patient background factors, cetuximab signature, and methylation status as independent variables (Table 2). Univariate analysis showed a lower risk of disease progression in the responder group [hazard ratio (HR) = 0.41; $p < 0.0001$], LMCC group (HR = 0.27; $p < 0.0001$), and irinotecan combination (HR = 0.54; $p = 0.01$). Multivariate analysis of significant predictors in univariate analysis showed a lower risk of disease progression in the responder group (HR = 0.45; $p = 0.002$), LMCC group (HR = 0.37; $p = 0.0004$), and irinotecan combination (HR = 0.45; $p = 0.001$).

4 | DISCUSSION

Our results show that the expression of specific genes in the cetuximab signature, a predictor of the efficacy of anti-EGFR antibody therapy, is regulated by promoter region methylation.

The relationship between promoter DNA methylation and gene expression is commonly investigated by comparative analysis of DNA promoter methylation and gene expression levels before and
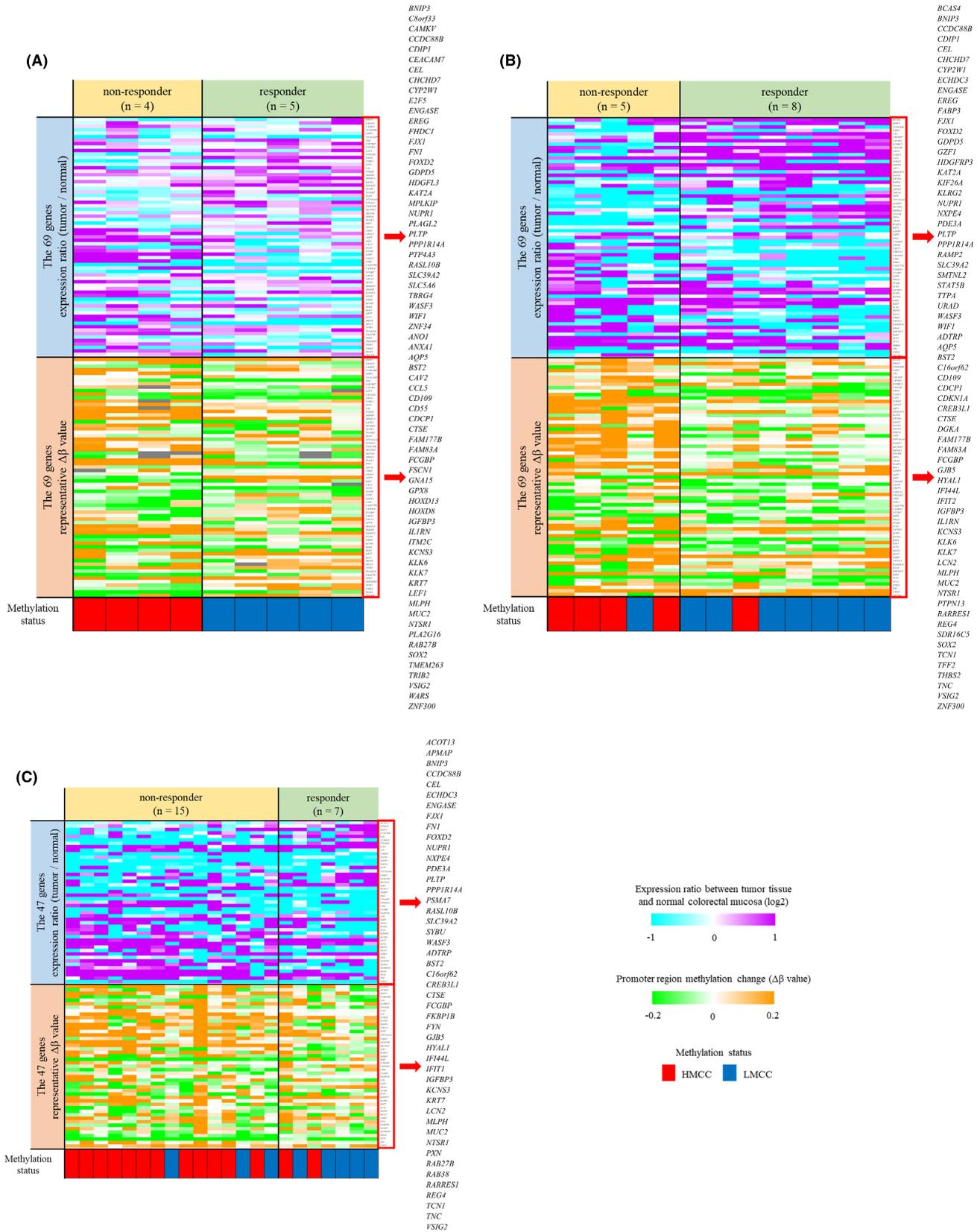
FIGURE 3 Genes with an inverse correlation between promoter region Δβ values and expression ratio in each cohort. Heatmaps of genes with an inverse correlation between promoter region Δβ values and the expression ratio in (A) OCH paired cohort, 69 genes; TCGA paired cohort, 69 genes; and (C) RBWH paired cohort, 47 genes. Top: expression ratio (log2) between tumor and normal colorectal mucosa for each gene. Middle: representative promoter region Δβ value for each gene (one CpG site per gene). Bottom: classification results based on DNA methylation status. Right: the magnified figures of the extracted gene rows. HMCC, highly methylated colorectal cancer; LMCC, low-methylated colorectal cancer; OCH, Osaki Citizen Hospital; RBWH, Royal Brisbane and Women's Hospital; TCGA, The Cancer Genome Atlas

TABLE 1 The 57 genes commonly extracted in the two analyses in Results, Sections 3.4 and 3.5

Gene symbol	Gene name	Response annotation[a]
BNIP3	BCL2 interacting protein 3	responder
CAMKV	CaM kinase like vesicle associated	responder
CCDC88B	coiled-coil domain containing 88B	responder
CDIP1	cell death inducing p53 target 1	responder
CEL	carboxyl ester lipase	responder
CHCHD7	coiled-coil-helix-coiled-coil-helix domain containing 7	responder
CYP2W1	cytochrome P450 family 2 subfamily W member 1	responder
ECHDC3	enoyl-CoA hydratase domain containing 3	responder
ENGASE	endo-beta-N-acetylglucosaminidase	responder
EREG	epiregulin	responder
FABP3	fatty acid binding protein 3	responder
FOXD2	forkhead box D2	responder
HDGFL3	HDGF like 3	responder
KAT2A	lysine acetyltransferase 2A	responder
KIF26A	kinesin family member 26A	responder
KLRG2	killer cell lectin like receptor G2	responder
MPLKIP	M-phase specific PLK1 interacting protein	responder
NUPR1	nuclear protein 1, transcriptional regulator	responder
NXPE4	neurexophilin and PC-esterase domain family member 4	responder
PDE3A	phosphodiesterase 3A	responder
PLTP	phospholipid transfer protein	responder
PPP1R14A	protein phosphatase 1 regulatory inhibitor subunit 14A	responder
PSMA7	proteasome subunit alpha 7	responder
RAMP2	receptor activity modifying protein 2	responder
SLC39A2	solute carrier family 39 member 2	responder
STAT5B	signal transducer and activator of transcription 5B	responder
SYBU	syntabulin	responder
TTPA	alpha tocopherol transfer protein	responder
WASF3	WAS protein family member 3	responder
WIF1	WNT inhibitory factor 1	responder
ZNF34	zinc finger protein 34	responder
BST2	bone marrow stromal cell antigen 2	non-responder
CD55	CD55 molecule	non-responder

TABLE 1 (Continued)

Gene symbol	Gene name	Response annotation[a]
CDKN1A	cyclin dependent kinase inhibitor 1A	non-responder
CREB3L1	cAMP responsive element binding protein 3 like 1	non-responder
DGKA	diacylglycerol kinase alpha	non-responder
FAM83A	family with sequence similarity 83 member A	non-responder
FKBP1B	FK506 binding protein 1B	non-responder
GNA15	G protein subunit alpha 15	non-responder
HOXD13	homeobox D13	non-responder
HOXD8	homeobox D8	non-responder
IFI44L	interferon induced protein 44 like	non-responder
IGFBP3	insulin like growth factor binding protein 3	non-responder
KCNS3	potassium voltage-gated channel modifier subfamily S member 3	non-responder
KLK6	kallikrein related peptidase 6	non-responder
KLK7	kallikrein related peptidase 7	non-responder
KRT7	keratin 7	non-responder
LCN2	lipocalin 2	non-responder
LEF1	lymphoid enhancer binding factor 1	non-responder
MLPH	melanophilin	non-responder
NTSR1	neurotensin receptor 1	non-responder
PTPN13	protein tyrosine phosphatase, non-receptor type 13	non-responder
RAB38	RAB38, member RAS oncogene family	non-responder
TCN1	transcobalamin 1	non-responder
TFF2	trefoil factor 2	non-responder
TRIB2	tribbles pseudokinase 2	non-responder
ZNF300	zinc finger protein 300	non-responder

[a] Response annotation: Upregulated genes in the responder or non-responder groups.

after the administration of demethylating agents to cultured cells.[19] However, demethylating agents cause non-specific demethylation of a wide range of CpG sites,[20] leading to demethylation not only in promoter regions but also in the gene body. The DNA methylation of gene body regions is reported to be associated with increased gene expression[21–23]; thus, accurate assessment of the relationship between promoter DNA methylation and gene expression may be difficult. In this study, we analyzed both the correlation between promoter methylation and gene expression levels in TTs and the correlation between changes in promoter methylation and changes in gene expression levels between paired NCMs and TTs. This analysis allowed us to identify critical changes in gene expression caused by
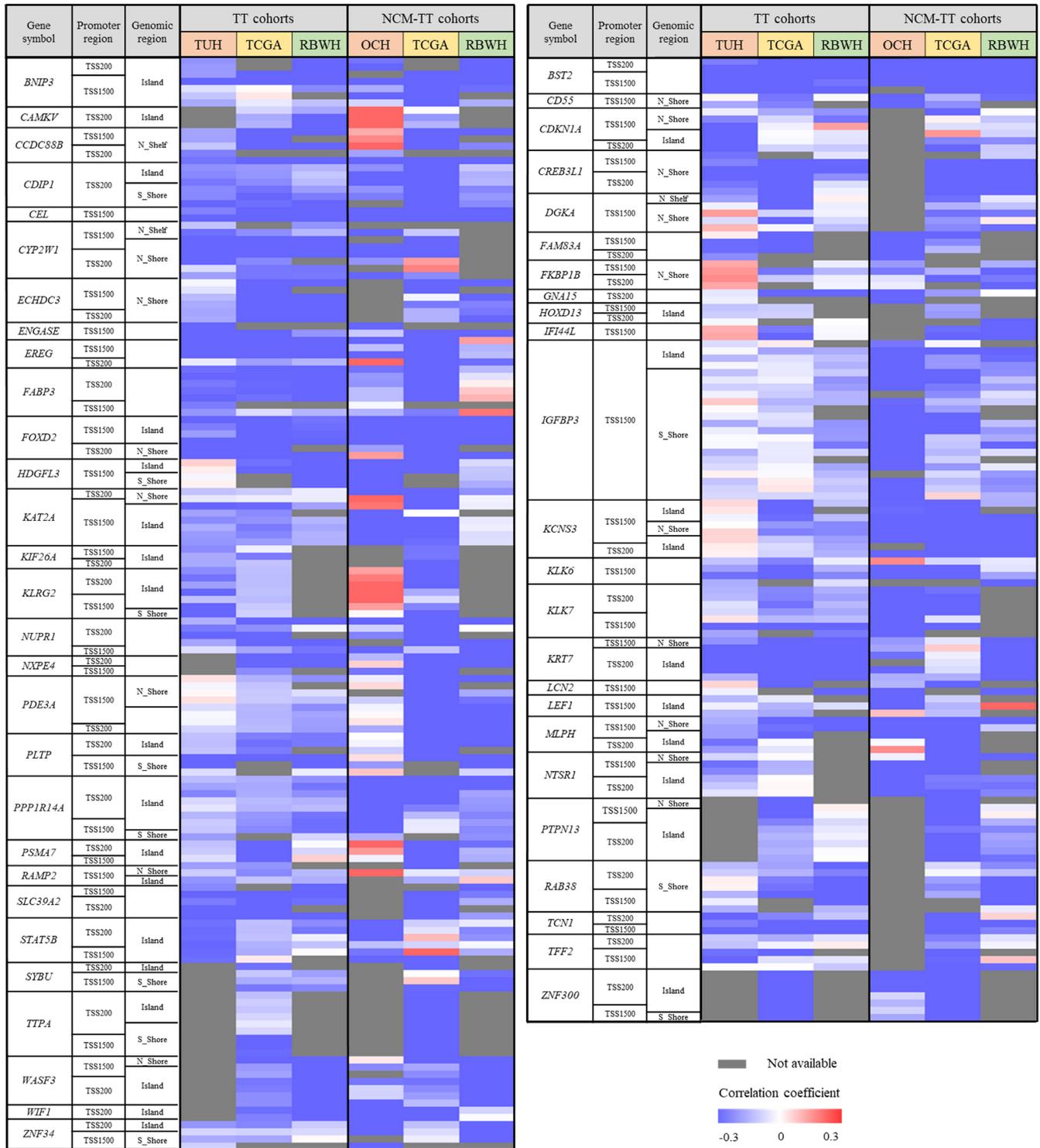
FIGURE 4 Heatmap of the 57 genes with an inverse correlation between promoter region methylation and expression level in multiple cohorts. Heatmap of the 57 genes with an inverse correlation between promoter region methylation and expression levels in ≥75% of the total cohort. Gray-background CpG sites: not available; TSS1500: 200-1500 bases upstream of the transcription start site; TSS200: 0-200 bases upstream of the transcription start site; Island: CpG island; N_Shore: 0-2000 bases upstream of the CpG island; S_Shore: 0-2000 bases downstream of the CpG island; N_Shelf: 2000-4000 bases upstream from the CpG island. NCM, normal colorectal mucosa; OCH, Osaki Citizen Hospital; RBWH, Royal Brisbane and Women's Hospital; TCGA, The Cancer Genome Atlas; TT, tumor tissue; TUH, Tohoku University Hospital

aberrant DNA methylation that occurs during carcinogenesis and determine the significance of DNA methylation status in regulating gene expression. These two analyses identified 57 genes for which

promoter methylation inversely correlated with expression level. This result indicates that about a quarter of the cetuximab signature genes are regulated by DNA methylation. Studies of mCRC
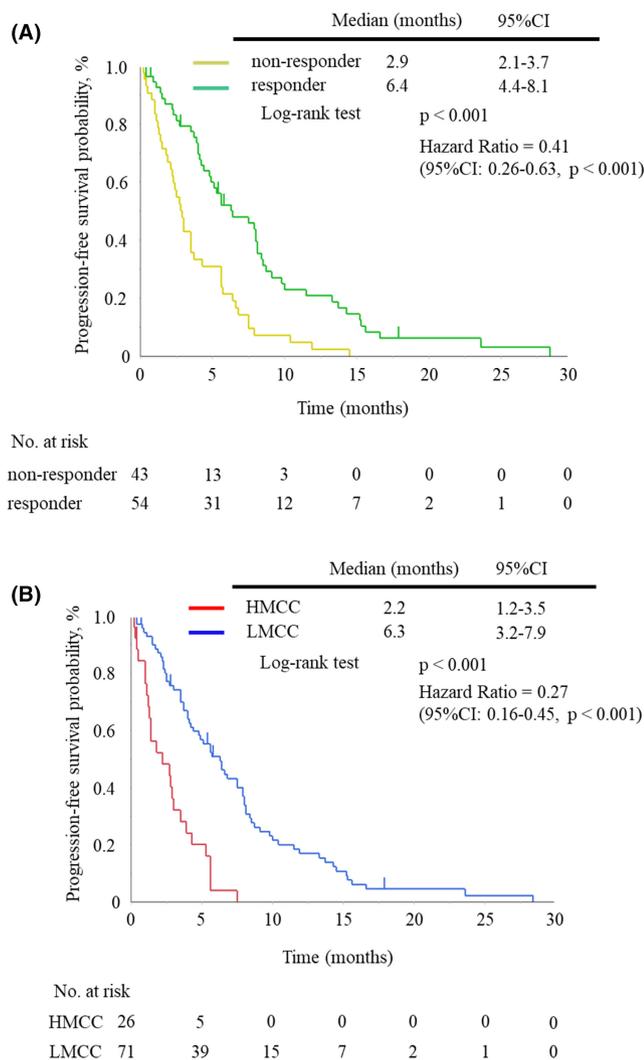
FIGURE 5 Kaplan-Meier curve of progression-free survival after anti-EGFR antibody therapy according to cetuximab signature or DNA methylation status. Kaplan-Meier curves for progression-free survival after anti-EGFR antibody therapy according to (A) cetuximab signature or (B) DNA methylation status. Non-responder group: yellow line ($n = 43$); responder group: green line ($n = 54$); HMCC group: red line ($n = 26$); LMCC group: blue line ($n = 71$). CI, confidence interval; EGFR, epidermal growth factor receptor; HMCC, highly methylated colorectal cancer; LMCC, low-methylated colorectal cancer

have reported an association between changes in gene expression due to DNA methylation abnormalities and the therapeutic effect of anti-EGFR antibody for a single gene,[19,24,25] but no report has shown such an association for multiple genes. Thus, our analytical method is unique and useful.

We also compared the correlation coefficients between gene expression level and DNA methylation at each CpG site in the promoter of the 57 genes in six independent cohorts. About two-thirds of all promoter CpG sites that exhibited an inverse relationship between DNA methylation status and expression level in multiple cohorts were identified as important CpG sites for regulating gene expression via DNA methylation. To determine whether aberrant DNA

methylation also occurs in NCMs, we compared promoter methylation levels in 57 genes between the two groups in NCMs based on the cetuximab signature or DNA methylation status classification of matched TTs. In both comparisons, we observed no significant difference in promoter methylation levels between the groups for any of the 57 genes. Our results suggest that aberrant methylation of DNA associated with the regulation of gene expression occurs during the process of carcinogenesis.

Among the 57 genes suggested to be regulated by DNA methylation, *EREG* is reported to be regulated by promoter methylation, and high gene expression is associated with prolonged PFS after anti-EGFR antibody therapy.[19,24] In our analysis, *EREG* promoter methylation significantly correlated inversely with *EREG* expression levels, consistent with previous reports. *BNIP3*, an apoptosis-promoting gene of the BCL-2 family, is overexpressed in hypoxic regions of tumors and induces autophagy and cell death.[26] Blockade of growth factor signaling such as epidermal growth factor (EGF) and insulin like growth factor (IGF) is reported to induce BNIP3-mediated apoptosis.[27] *BNIP3* expression was observed to be downregulated by promoter methylation, suggesting that anti-EGFR antibody may not induce apoptosis in tumor cells and cause resistance. Furthermore, high expression of *BST2* promotes signaling in the downstream Ras-Raf-MEK-ERK and JAK-STAT pathways through activation of EGFR,[28] which can induce resistance to anti-EGFR antibody therapies. *KLK6* and *MLPH* have also been listed in gene expression signatures associated with panitumumab resistance.[29]

Finally, we analyzed the factors contributing to PFS of anti-EGFR antibody therapy in the TUH cohort. Regarding patient backgrounds, right-sided colon cases were significantly more common among non-responders and those with HMCC. Among patients with right-sided CRC, a higher percentage are CIMP positive and have a lower response to anti-EGFR antibody therapy than among those with left-sided CRC.[30,31] Our results are consistent with these previous reports. Non-responders and patients with HMCC had a significantly lower PFS than did responders and those with LMCC. The classification based on the cetuximab signature had not been validated in a clinical cohort, except by Schutte et al. This study shows that the cetuximab signature is clinically useful in a Japanese cohort. In univariate and multivariate analyses of PFS for anti-EGFR antibody therapy, classification based on DNA methylation status and classification based on cetuximab signature were both independent predictors of PFS. DNA methylation status was most significantly associated with PFS and more strongly predictive of therapeutic efficacy of anti-EGFR antibody therapy than was the cetuximab signature or irinotecan combination. If all genes in the cetuximab signature were regulated by DNA methylation, we would expect that DNA methylation status and cetuximab signature would be strongly correlated. However, multivariate analysis of the PFS for anti-EGFR antibody therapy showed that DNA methylation status and cetuximab signature were independent of each other. In this analysis, only about a quarter of the cetuximab signature genes were suggested to be regulated by DNA methylation in the promoter region, suggesting

TABLE 2 Cox regression analysis for progression-free survival of anti-EGFR antibodies in the Tohoku University Hospital (TUH) cohorts

	Progression-free survival						
	Univariate				Multivariate		
	n	HR	95%CI	*P*-value[a]	HR	95%CI	*P*-value[a]
Cetuximab signature							
Non-responder	43	1.00					
Responder	54	0.41	0.26–0.63	< 0.0001	0.45	0.28–0.74	0.002
DNA methylation status							
HMCC	26	1.00					
LMCC	71	0.27	0.16–0.45	< 0.0001	0.37	0.21–0.64	0.0004
Age							
<65	69						
≥65	28	0.88	0.57–1.38	0.59			
Sex							
Male	66						
Female	31	1.22	0.78–1.92	0.38			
Primary site							
Right side	27						
Left side or rectum	70	0.83	0.52–1.32	0.42			
Stage at diagnosis							
≤III	35						
IV	64	0.89	0.58–1.38	0.61			
No. of organs with metastasis							
≤1	46						
≥2	51	1.16	0.77–1.78	0.47			
BRAF mutation status							
Wild or NA	91						
Mutant	6	1.02	0.40–2.61	0.96			
Minor *RAS* mutation status[b]							
Wild	86						
Mutant	11	1.84	0.94–3.57	0.07			
No. of previous regimens							
≤1	14						
≥2	83	1.37	0.74–2.51	0.32			
Type of anti-EGFR treatment							
Monotherapy	26	1.00					
Combination with irinotecan	71	0.54	0.34–0.87	0.01	0.45	0.27–0.73	0.001

Abbreviations: CI, confidence interval; EGFR, epidermal growth factor receptor; HMCC, highly methylated colorectal cancer; HR, hazard ratio; LMCC, low-methylated colorectal cancer; NA, not available.

[a] χ^2 test.

[b] Minor *RAS* mutation status: *RAS* mutation except for *KRAS* codons 12 and 13.

that cetuximab signature and DNA methylation status have no confounding effects on each other. The remaining cetuximab signature genes not regulated by DNA methylation may be regulated by other means, including noncoding RNAs, microRNAs,[32,33] and histone modifications.[34] Another possibility is that the expression of

some cetuximab signature genes is indirectly affected by regulatory methylation of genes located upstream or downstream in signaling pathways. Furthermore, methylation of gene body regions may regulate the expression of some cetuximab signature genes, as gene body methylation is associated with increased gene expression.[21–23]

Because most of the probes for analyzing DNA methylation arrays are designed to target CpG sites in promoters, detailed methylation analysis of gene body regions is difficult. Therefore, further analysis of gene body methylation is needed.

The limitation of this study is contamination of the clinical samples by stromal cells, which may affect the apparent methylation and expression levels. Stromal cell contamination might result in the incomplete extraction of genes regulated by promoter region methylation. In the future, we would like to conduct in vitro analyses in the absence of stromal cells to further analyze gene expression regulation by DNA methylation. Furthermore, the RBWH cohort was an in silico analysis using ArrayExpress data without clinical samples available; thus, we could not perform additional mutation analysis.

While this study focused on cetuximab signature genes, we expect that many genes outside of this group affect the sensitivity to anti-EGFR antibodies and are regulated by DNA methylation. We will continue to extract more genes regulated by DNA methylation that are associated with the efficacy of anti-EGFR therapy to clarify the molecular mechanisms underlying resistance to such therapy.

In conclusion, we found that aberrant DNA methylation regulates specific gene in cetuximab signature during carcinogenesis, suggesting that it is one of the important factors that define sensitivity to anti-EGFR antibody agents.

DISCLOSURE

C. Ishioka received research funding from Ono, MSD; research grants from Takeda, Daiichi-Sankyo, Ono, Asahi-Kasei Pharma, Taiho, Chugai, and Bayer; and honoraria from Chugai. S. Takahashi received research funding from Merck Biopharma. All remaining authors have declared no conflict of interest.

ORCID

Yasufumi Otsuki https://orcid.org/0000-0002-5794-5115
Kota Ouchi https://orcid.org/0000-0002-5572-1337
Shin Takahashi https://orcid.org/0000-0001-6695-3099
Chikashi Ishioka https://orcid.org/0000-0002-3023-1227

REFERENCES

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2021;71:209-249.
2. Douillard JY, Oliner KS, Siena S, et al. Panitumumab-FOLFOX4 treatment and RAS mutations in colorectal cancer. *N Engl J Med.* 2013;369:1023-1034.
3. Bokemeyer C, Köhne C-H, Ciardiello F, et al. FOLFOX4 plus cetuximab treatment and RAS mutations in colorectal cancer. *Eur J Cancer.* 2015;51:1243-1252.
4. Cutsem EV, Lentz HJ, Kohne CH, et al. Fluorouracil, leucovorin, and irinotecan plus cetuximab treatment and RAS mutations in colorectal cancer. *J Clin Oncol.* 2015;33:692-700.
5. Heinemann V, Weikersthal LFV, Decker T, et al. FOLFIRI plus cetuximab versus FOLFIRI plus bevacizumab as first-line treatment for patients with metastatic colorectal cancer (FIRE-3): a randomised, open-label, phase 3 trial. *Lancet Oncol.* 2014;15:1065-1075.
6. Lin JC, Jeong S, Liang G, et al. Role of nucleosomal occupancy in the epigenetic silencing of the MLH1 CpG island. *Cancer Cell.* 2007;12:432-444.
7. Shenker N, Flanagan JM. Intragenic DNA methylation: implications of this epigenetic mechanism for cancer research. *Br J Cancer.* 2012;106:248-253.
8. Juo YY, Johnston FM, Zhang DY, et al. Prognostic value of CpG island methylator phenotype among colorectal cancer patients: a systematic review and meta-analysis. *Ann Oncol.* 2014;25:2314-2327.
9. Cha Y, Kim KJ, Han SW, et al. Adverse prognostic impact of the CpG island methylator phenotype in metastatic colorectal cancer. *Br J Cancer.* 2016;115:164-1671.
10. Ouchi K, Takahashi S, Yamada Y, et al. DNA methylation status as a biomarker of anti-epidermal growth factor receptor treatment for metastatic colorectal cancer. *Cancer Sci.* 2015;106:1722-1729.
11. Schütte M, Risch T, Abdavi-Azar N, et al. Molecular dissection of colorectal cancer in pre-clinical models identifies biomarkers predicting sensitivity to EGFR inhibitors. *Nat Commun.* 2017;8:14262.
12. Fennell L, Dumenil T, Wockner L, et al. Integrative genome-scale DNA methylation analysis of a large and unselected cohort reveals 5 distinct subtypes of colorectal adenocarcinomas. *Cell Mol Gastroenterol Hepatol.* 2019;8:269-290.
13. Inoue M, Takahashi S, Soeda H, et al. Gene-expression profiles correlate with the efficacy of anti-EGFR therapy and chemotherapy for colorectal cancer. *Int J Clin Oncol.* 2015;20:1147-1155.
14. Irizarry RA, Boistad BM, Collin F, et al. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.* 2003;31:e15.
15. Affymetrix Inc. Microarray normalization using Signal Space Transformation with probe Guanine Cytosine Count Correction. Affymetrix White Paper. 2015;P/N;EMI05133-2.
16. Zhao Y, Li MC, Konate MM, et al. TPM, FPKM, or normalized counts? a comparative study of quantification measures for the analysis of rna-seq data from the NCI patient-derived models repository. *J Transl Med.* 2021;19:269.
17. Soeda H, Shimodaira H, Gamoh M, et al. Phase II trial of cetuximab plus irinotecan for oxaliplatin- and irinotecan-based chemotherapy-refractory patients with advanced and/or metastatic colorectal cancer: Evaluation of efficacy and safety based on *KRAS* mutation status (T-CORE0801). *Oncology.* 2014;87:7-20.
18. Ouchi K, Takahashi S, Okita A, et al. A modified MethyLight assay predicts the clinical outcomes of anti-epidermal growth factor receptor treatment in metastatic colorectal cancer. *Cancer Sci.* 2022;113:1057-1068.
19. Lee MS, McGuffey EJ, Morris JS, et al. Association of CpG island methylator phenotype and EREG/AREG methylation and expression in colorectal cancer. *Br J Cancer.* 2016;114:1352-1361.
20. Kelly TK, deCarvalho DD, Jones PA. Epigenetic modifications as therapeutic targets. *Nat Biotechnol.* 2010;28:1069-1078.
21. Maunakea AK, Nagarajan RP, Bilenky M, et al. Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature.* 2010;466:253-257.
22. Varley KE, Gertz J, Bowling KM, et al. Dynamic DNA methylation across diverse human cell lines and tissues. *Genome Res.* 2013;23:555-567.
23. Yang X, Han H, deCarvalho DD, et al. Gene body methylation can alter gene expression and is a therapeutic target in cancer. *Cancer Cell.* 2014;26:577-590.

24. Qu X, Sandmann T, Frierson HJr, et al. Integrated genomic analysis of colorectal cancer progression reveals activation of EGFR through demethylation of the EREG promoter. *Oncogene*. 2016;35:6403-6415.

25. Scartozzi M, Bearzi I, Mandolesi A, et al. Epidermal growth factor receptor (EGFR) gene promoter methylation and cetuximab treatment in colorectal cancer patients. *Br J Cancer*. 2011;104:1786-1790.

26. Zhang J, Ney PA. Role of BNIP3 and NIX in cell death, autophagy, and mitophagy. *Cell Death Differ*. 2009;16:939-946.

27. Kothari S, Cizeau J, McMillan-Ward E, et al. BNIP3 plays a role in hypoxic cell death in human epithelial cells that is inhibited by growth factors EGF and IGF. *Oncogene*. 2003;22:4734-4744.

28. Zhang G, Li X, Chen Q, et al. CD317 activates EGFR by regulating its association with lipid rafts. *Cancer Res*. 2019;79:2220-2231.

29. Barry GS, Cheang MC, Chang HL, et al. Genomic markers of panitumumab resistance including ERBB2/HER2 in a phase II study of KRAS wild-type (wt) metastatic colorectal cancer (mCRC). *Oncotarget*. 2016;7:18953-18964.

30. Dekker E, Tanis PJ, Vleugels JLA, et al. Colorectal cancer. *Lancet*. 2019;394:1467-1480.

31. Lee MS, Menter DG, Kopetz S. Right versus left colon cancer biology: integrating the consensus molecular subtypes. *J Natl Compr Canc Netw*. 2017;15:411-419.

32. Huntzinger E, Izaurralde E. Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nat Rev Genet*. 2011;12:99-110.

33. Holoch D, Moazed D. RNA-mediated epigenetic regulation of gene expression. *Nat Rev Genet*. 2015;16:71-84.

34. Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. *Cell Res*. 2011;21:381-395.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.