Association of CpG island methylator phenotype and *EREG/AREG* methylation and expression in colorectal cancer

Michael S Lee[1,2], Elizabeth J McGuffey[3], Jeffrey S Morris[4], Ganiraju Manyam[4], Veerabahdran Baladandayuthapani[4], Wei Wei[4], Van K Morris[5], Michael J Overman[5], Dipen M Maru[6], Zhi-Qin Jiang[5], Stanley R Hamilton[6] and Scott Kopetz*,[5]

[1]Division of Hematology/Oncology, Department of Medicine, University of North Carolina at Chapel Hill, 170 Manning Drive, CB 7305, Chapel Hill, NC 27599, USA; [2]Lineberger Comprehensive Cancer Center, 170 Manning Drive, CB 7295, Chapel Hill, NC 27599, USA; [3]United States Naval Academy, Department of Mathematics, 572C Holloway Road, Annapolis, MD 21402, USA; [4]Department of Biostatistics, University of Texas MD Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX 77030, USA; [5]Department of Gastrointestinal Medical Oncology, University of Texas MD Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX 77030, USA and [6]Department of Pathology, University of Texas MD Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX 77030, USA

Background: High *EREG* and *AREG* expression, and left-sided primary tumours are associated with superior efficacy of anti-epidermal growth factor receptor (EGFR) therapy in metastatic colorectal cancer (CRC), but a unifying explanation of these findings is lacking.

Methods: RNA-seq, gene expression arrays, and DNA methylation profiling were completed on 179 CRC tumours. Results were validated using independent The Cancer Genome Atlas data sets. An independent cohort of 198 *KRAS* wild-type metastatic CRC tumours was tested for CpG island methylator phenotype (CIMP) status, and progression-free survival (PFS) with the first anti-EGFR regimen was retrospectively determined.

Results: *EREG* and *AREG* expression was highly inversely correlated with methylation and was inversely associated with right-sided primary tumour, *BRAF* mutation, and CIMP-high status. Treatment of CRC cell lines with hypomethylating agents decreased methylation and increased expression of *EREG*. Inferior PFS with anti-EGFR therapy was associated with CIMP-high status, *BRAF* mutation, *NRAS* mutation, and right-sided primary tumour on univariate analysis. Among known *BRAF/NRAS* wild-type tumours, inferior PFS remained associated with CIMP-high status (median PFS 5.6 *vs* 9.0 mo, $P = 0.023$).

Conclusions: *EREG* and *AREG* are strongly regulated by methylation, and their expression is associated with CIMP status and primary tumour site, which may explain the association of primary tumour site and *EREG/AREG* expression with anti-EGFR therapy efficacy.

With the development of therapies for metastatic colorectal cancer (CRC) targeting the epidermal growth factor receptor (EGFR), such as cetuximab and panitumumab, there has been significant effort to identify rational biomarkers of resistance or susceptibility.

Initially, mutations in *KRAS* codons 12 and 13 were found to be predictive biomarkers of lack of response to anti-EGFR therapies (Amado *et al*, 2008; Karapetis *et al*, 2008; Van Cutsem *et al*, 2009; Bokemeyer *et al*, 2011), and more recently mutations in exons 3

and 4 of *KRAS* and in *NRAS* have also been identified as predictive of resistance to anti-EGFR therapies (Douillard *et al*, 2013). However, resistance to anti-EGFR therapy exists even among patients whose tumours are wild-type in *KRAS* and *NRAS*, demonstrating the need to find and characterise additional biomarkers of resistance.

The predictive impact of other mutations in signalling pathway components downstream of EGFR has been unclear. Among stage IV metastatic CRC patients treated with chemotherapy and cetuximab, *BRAF* mutation is associated with poor prognosis (De Roock *et al*, 2010; Tol *et al*, 2010), but is not necessarily predictive of resistance to anti-EGFR therapy. A pooled analysis of two studies that randomised *KRAS* wild-type metastatic CRC patients to chemotherapy or chemotherapy plus cetuximab found that *BRAF* mutant patients had inferior overall survival (OS) than *BRAF* wild-type patients, but the addition of cetuximab was associated with non-significant trend towards improved survival in the *BRAF* mutant patients (HR 0.62, 95% CI 0.36–1.06, $P = 0.076$) (Bokemeyer *et al*, 2012). There are also conflicting results regarding the prognostic effect of *PIK3CA* mutations or PTEN expression loss on patients treated with anti-EGFR therapy (Laurent-Puig *et al*, 2009; Loupakis *et al*, 2009; De Roock *et al*, 2010; Tol *et al*, 2010; Karapetis *et al*, 2014). The role of these mutations continues to undergo investigation.

A well-described biomarker associated with sensitivity to anti-EGFR therapy is increased tumour gene expression of *EREG* and *AREG*, which encode the EGFR ligands epiregulin and amphiregulin (Khambata-Ford *et al*, 2007; Jacobs *et al*, 2009). *EREG* and *AREG* expression is inhibited by blockade of EGFR signalling and is stimulated by treatment with other EGFR ligands, and it has been hypothesised that higher expression of *EREG* and *AREG* may indicate tumour cell dependence on an autocrine EGFR-activating loop, and thus predict for increased susceptibility to anti-EGFR therapy (Khambata-Ford *et al*, 2007). However, mechanisms dictating varying levels of *EREG* or *AREG* expression remain unclear.

The site of the primary tumour also appears to be predictive of efficacy of anti-EGFR therapy. Clinical trials in patients with *KRAS* exon 2 wild-type metastatic CRC reveal that patients with left-sided colorectal primary tumours have superior progression-free survival (PFS) on treatment with cetuximab-based regimens compared with patients with right-sided primary tumours (Von Einem *et al*, 2014; Brule *et al*, 2015). Although the mechanism of this distinction is not defined, right-sided primary CRCs are known to have distinct pathobiology and characteristics, including higher rates of *BRAF* mutation, microsatellite instability (MSI-high), and high CpG island methylator phenotype (CIMP-high) (Yamauchi *et al*, 2012). CIMP-high tumours are marked by widespread DNA hypermethylation, which can epigenetically silence genes when occurring within promoter loci (Weisenberger *et al*, 2006). We sought to determine whether there was an association between global methylation status as assessed by CIMP status, the methylation status of CpG islands within the *AREG* and *EREG* promoters, the mRNA expression of *AREG* and *EREG*, and the site of the primary colorectal tumour in multiple independent data sets collected within the University of Texas MD Anderson Cancer Center (MDACC) and through The Cancer Genome Atlas (TCGA).

MATERIALS AND METHODS

Patient cohorts and data analysis. A cohort of 179 patients at MDACC with stages I–IV CRC provided informed consent for biomarker analyses on tumour tissue and retrospective analysis of patient records for research purposes, and the study was approved by the institutional review board. Eligible tumour specimens required at least 30% tumour cells on central review by a certified pathologist, and macrodissection of fresh frozen primary tumour tissue was performed. Gene expression analysis was performed using Agilent microarrays (Agilent Technologies, Santa Clara, CA, USA). RNA was isolated and assessed for quality, and RNA of adequate quality was amplified, labelled, and hybridised to the microarray. Data are expressed as *z*-scores. Methylation profiling was determined using the Illumina Infinium HumanMethylation450 BeadChip Kit (Illumina, San Diego, CA, USA), using bisulphite sequencing of methylated sites to determine methylation status at over 480 000 CpG sites covering 99% of Ref-seq genes.

Two additional cohorts of patients with stages II–IV colon or rectal adenocarcinoma were analysed from data generated by TCGA Research Network (http://cancergenome.nih.gov/): one cohort, 'TCGA27' ($n = 218$), had gene expression analysed by the Custom Agilent 244 K gene expression microarray (Agilent Technologies) and expressed as *z*-scores, and had methylation profiling performed using the Illumina Infinium HumanMethylation27 BeadChip as described (The Cancer Genome Atlas Network, 2012); and the second cohort, 'TCGA450' ($n = 356$), had gene expression analysed by RNA sequencing (The Cancer Genome Atlas Network, 2012) and expressed as $\log_2$-transformed values, and also had methylation profiling performed using the Illumina Infinium HumanMethylation450 BeadChip. In order to evaluate the correlation of CpG island methylation and gene expression, the TCGA pan-cancer data set was utilised (The Cancer Genome Atlas Research Network *et al*, 2013).

A final independent cohort of 440 stage IV CRC MDACC patients who had been heavily pretreated was enrolled on the Assessment of Targeted Therapies Against Colorectal Cancer (ATTACC) protocol between August 2010 and October 2013 for screening and assignment to 10 individual phase I or II companion clinical trials based on testing of banked formalin-fixed, paraffin-embedded tumour tissue with gene sequencing, CIMP testing, and immunohistochemical staining. The patients in the ATTACC protocol provided informed consent for biomarker analyses on archived tumour tissue and retrospective analysis of patient records for research purposes, and the study was approved by the institutional review board. Of the 440 patients enrolled in ATTACC, 198 were *KRAS* exon 2 wild type by standard of care testing and were successfully tested for CIMP status. From the ATTACC specimens, bisulphite pyrosequencing of six well-defined, traditionally utilised CpG islands (Toyota *et al*, 1999) was performed, and specimens with ⩾ 40% methylation of CpG islands were deemed CIMP-high. The PCR primers used at MD Anderson for bisulphite pyrosequencing CpG islands in p14, p16, MLH1, MINT1, MINT2, and MINT31 are listed in Supplementary Table 1A.

Definition of CIMP status. CIMP status was determined from methylation arrays by using the following two methodologies: the first method was based on assessment of the methylation status of the six CpG islands used to determine CIMP status on clinical specimens in the ATTACC cohort; and the second method was based on clustering of methylation profiles using methodology paralleling that in TCGA (The Cancer Genome Atlas Network, 2012).

To extrapolate the CIMP status based on the six-locus clinical panel from the Illumina Infinium HumanMethylation450 Bead-Chip results, the PCR primers for the bisulphite-treated sequences for p14, p16, MLH1, MINT1, MINT2, and MINT31 were obtained. A search for the corresponding sequence within the human genome was undertaken using BiSearch (http://bisearch.enzim.hu/) (Tusnady *et al*, 2005; Aranyi *et al*, 2006), and the identity of the sequence was verified by ensuring that the sequencing primer was located within the identified region. To identify the corresponding nucleotide location within GRCh37, NCBI Blast was performed

(Supplementary Table 1B). Subsequently, the CpG islands from the Illumina Infinium HumanMethylation450 array that were located within the primer regions were determined, and 1–2 CpG islands within the PCR-amplified regions were found for the MLH1, MINT2, and MINT31 primers. For the MINT1, p14, and p16 primers, no CpG islands on the panel were found within the amplified regions, although 1–2 CpG islands located <300 nucleotides away for each of the primer pairs were found (Supplementary Table 1C). The beta-value of each of these CpG islands was transformed to an M-value (Du *et al*, 2010), and the distribution of M-values for each data set was plotted and found to be bimodal. For each CpG island, a threshold was determined at the point between the two modes to dichotomise the methylation status in the MDACC data. This threshold was applied to the M-value transformations of the TCGA data and was found to also reflect the point between the two modes (Supplementary Table 2A and B). Then, a voting scheme mirroring the clinical CIMP panel was enacted to determine CIMP status (CIMP-high represents $>40\%$ or $\geqslant 3/6$ markers).

For the TCGA samples assayed with the Illumina Infinium HumanMethylation27 BeadChip instead of the larger Illumina Infinium HumanMethylation450 BeadChip, it was not possible to align the clinical panel with existing CpG island probes. Instead, a clustering method of methylation profiles was used, following the methods described in the TCGA manuscript (The Cancer Genome Atlas Network, 2012).

Cell line data sets. To determine the effect of hypomethylating agents azacitidine and decitabine on methylation of specific CpG island promoters, raw experimental data were obtained from ArrayExpress database (www.ebi.ac.uk/arrayexpress) using accession number E-MTAB-417. In this study, HCT116 colon cancer cells were treated with $1\,\mu$M azacitidine or decitabine for 24 h. Subsequently, the Illumina Infinium HumanMethylation27 BeadChip was used to determine methylation status (Hagemann *et al*, 2011).

To determine the effect of the hypomethylating agent azacitidine on expression of *EREG* and *AREG*, experimental data was obtained from the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) using accession number GSE57341. In this study, a panel of 14 CRC cell lines was treated with 500 nM of azacitidine for 72 h. Cells were harvested at 1, 3, 7, 10, 14, 21, or 28 days, and expression was assayed using the Agilent 44K Expression Array (Li *et al*, 2014).

Statistical analysis. Gene expression microarray data were provided as *z*-scores. RNA-seq data were provided as raw values that were base 2 logarithmically transformed for statistical analysis. To identify association between clinical characteristics and expression or methylation levels, the Mann–Whitney *U*-test was utilised if variables were dichotomous, or the Kruskal–Wallis test was utilised if variables had >2 possible values. To perform survival analysis with PFS, the Kaplan–Meier method was used and significance testing was performed using the log-rank test. Additional univariate and multivariate Cox regression analyses were performed using IBM SPSS Statistics version 21 (Armonk, NY, USA). In addition, integrated Bayesian analysis of high-dimensional multiplatform genomics (iBAG) (Wang *et al*, 2013) data was performed using the nonlinear approach (Jennings *et al*, 2013) to determine the percent of variability in expression that is explained by methylation. This method involved fitting an additive regression model in which expression was regressed on methylation and copy number as additive nonparametric predictors, with residual error representing variability in expression not explained by methylation or copy number, that is, explained by some other upstream regulator. After model fitting, percent variability explained by methylation, copy number, and other causes were estimated through the corresponding coefficients of determination.

For the MDACC ATTACC cohort, to determine associations with PFS and OS, multiple imputation methodology was used, with predictor variables including age, sex, CIMP-high *vs* low/none, right- *vs* left-sided primary tumours, duration of first EGFR regimen, progression status, prior bevacizumab, presence or absence of additional cytotoxic chemotherapy, PTEN status, *PIK3CA* mutation status, *BRAF* mutation status, *NRAS* mutation status, and number of prior chemotherapy regimens. Any missing values were addressed using an assumption of missing at random and multivariate normality using the MCMC method in SPSS to create 20 imputed data sets. Subsequently, Cox proportional hazards models were used to perform univariate and multivariate regression analyses for PFS with the first anti-EGFR regimen or OS.

RESULTS

***EREG* and *AREG* expression are strongly inversely associated with methylation of loci within the promoters of *EREG* and *AREG*.** A cohort of 179 MDACC patients with CRC had primary tumour specimens assayed for *EREG* and *AREG* methylation and expression. Characteristics of the patient cohort are described in Supplementary Table 3. Eight out of nine CpG islands located within the *EREG* gene promoter or body were significantly inversely associated with expression, as were four out of five CpG islands located within the *AREG* gene and promoter (Supplementary Table 4A and B). The strongest correlation with *EREG* expression was with the CpG island cg19308222 ($\rho = -0.726$, $P < 10^{-6}$; Figure 1A). The strongest correlation with *AREG* expression was with the CpG islands cg26611070 ($\rho = -0.520$, $P < 10^{-6}$) and cg022334660 ($\rho = -0.501$, $P < 10^{-6}$; Supplementary Figure 1a and b). These findings demonstrate that there is a significant inverse association between DNA methylation and expression of *EREG* and *AREG*, and identify promoter methylation as a significant regulatory mechanism of expression of *EREG* and *AREG*.

To confirm these findings, the association of *EREG* and *AREG* methylation and expression was determined in two independent cohorts of colon and rectal adenocarcinoma cancer specimens from the TCGA. In the TCGA450 cohort ($n = 356$; Supplementary Tables 4A and B and 5), the strongest correlation with *EREG* expression was again with the CpG island cg19308222 ($\rho = -0.671$, $P < 10^{-6}$; Figure 1B), and the strongest correlation with *AREG* expression was again with the CpG island cg02334660 ($\rho = -0.457$, $P < 10^{-6}$; Supplementary Figure 1c and d). In the TCGA27 cohort ($n = 218$; Supplementary Table 6), strong inverse correlation was again observed between *EREG* expression and cg19308222 ($\rho = -0.659$, $P < 10^{-6}$; Figure 1C). Because the panel of CpG islands tested for methylation was smaller in this array, there were no CpG islands tested within the *AREG* promoter.

To put this strong negative correlation in context, among colon adenocarcinoma specimens in the TCGA, the inverse correlation between methylation at cg19308222 and *EREG* expression was the 219th out of 18 945 most-negative correlation (top 1.2%) and between methylation at cg02334660 and *AREG* expression was the 1801th out of 18 945 (top 9.5%) among methylation/gene expression pairs (Broad Institute TCGA Genome Data Analysis Center, 2014b). Similarly, among rectal adenocarcinoma specimens in the TCGA, the inverse correlation between methylation at cg19308222 and *EREG* expression was the 155th out of 19 209 most-negative correlation (top 0.8%), and the inverse correlation between methylation at cg26611070 and *AREG* expression was the 1686th out of 19 209 most-negative correlation (top 8.8%) among methylation/gene expression pairs (Broad Institute TCGA Genome Data Analysis Center, 2014a).
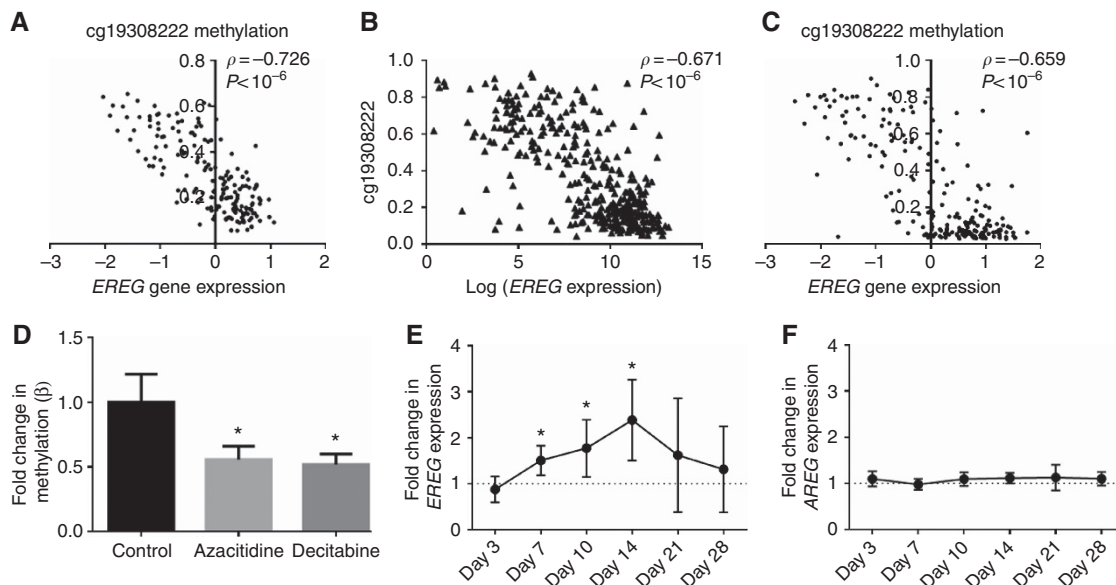
Figure 1. *EREG* methylation is inversely associated with expression, and is modulated by hypomethylating agents. (**A–C**) Scatter plots of methylation β-value at the *EREG* CpG island cg19308222 compared with *EREG* expression in the (**A**) MDACC cohort ($n = 179$), (**B**) TCGA450 cohort ($n = 356$), or (**C**) TCGA27 cohort ($n = 218$). (**D**) Methylation at cg19308222 after treatment with 1 μM azacitidine or decitabine for 24 h. (**E** and **F**) Expression of *EREG* (A_23_P41344), but not *AREG* (A_23_P249071), is increased after treatment with 500 nM azacitidine for 72 h and cell harvesting after the indicated duration. *$P < 0.05$ compared with control.

iBAG analysis. Additional analysis was completed to determine the extent to which variation in expression of *EREG* and *AREG* was attributable to methylation, compared with copy number variation or other unspecified upstream regulators, using iBAG (Wang *et al*, 2013). In the MDACC cohort, 64.5% of *EREG* expression variation was explained by methylation of cg19308222, compared with 2.9% by copy number variation and 32.7% from other causes. In *AREG*, if considering cg02334660 alone, 33.8% of expression was explainable by variation in methylation, compared with 0.5% by copy number variation and 65.7% by other causes. If considering cg2661070 alone, 34.8% of expression was explained by variation in methylation, compared with 0.6% by copy number variation and 64.6% by other causes. In the TCGA27 cohort, 57.5% of *EREG* expression variation was explained by methylation of the cg19308222 locus, compared with 0.9% from copy number variation and 41.7% from other causes. The variation for *AREG* could not be determined as there was no *AREG* methylation locus in the Illumina Infinium 27K panel.

By comparison, in a large set of 799 genes analysed by iBAG in two CRC cohorts, both the TCGA27 cohort and the MDACC cohort, we found the median percent of expression explained by variation in methylation was 5.78%. Only 5 of these 799 genes (0.63%) had >60% of expression variability explained by methylation, placing *EREG* within the top 1% of genes whose expression was predominantly modulated by methylation.

Demethylating agents decrease methylation levels and increase expression of *EREG*. In order to evaluate whether methylation modulates *EREG* expression, we utilised RNA arrays and Illumina Infinium HumanMethylation27 BeadChip data sets for CRC cell lines treated with azacitidine and decitabine. *EREG* methylation at cg19308222 was significantly reduced by treatment with either azacitidine or decitabine (Figure 1D). In contrast, the *EREG* associated CpG site cg04941721 was not altered with either agent, consistent with the limited correlation of this site with *EREG* expression. No CpG probes for *AREG* are present on the Illumina 27K array, limiting evaluation of this target. Expression of *EREG* was evaluated at various time points after 72 h of treatment with

azacitidine in a panel of 14 CRC cell lines, demonstrating a time-dependent increase in expression of *EREG*, which was not seen in *AREG* (Figure 1E and F).

Right-sided primary, CIMP-high, MSI-high, and BRAF-mutated cancers are associated with higher levels of gene methylation and lower levels of expression of *EREG* and *AREG*. Next, we examined whether additional clinical and pathological characteristics of CRC were associated with significant differences in methylation and expression of *AREG* and *EREG*. In the MDACC cohort, right-sided primary tumour, MSI-high status, *BRAF* V600E mutant status, and mucinous histology were all associated with significantly lower levels of *AREG* and *EREG* expression, and significantly higher levels of methylation of the *AREG* loci cg02334660 and cg26611070, and the *EREG* locus cg19308222 (Table 1 and Supplementary Figure 2). CIMP was defined by one of two different methods, and CIMP-high status by either method was associated with significantly lower *AREG* and *EREG* expression, and significantly higher methylation of *EREG* cg19308222 (Table 1 and Figure 2A and B). In the TCGA450 cohort, similar to the MDACC results, MSI-high status, *BRAF* mutant status, mucinous histology, and right-sided primary tumour were significantly associated with lower levels of *AREG* and *EREG* expression, and with higher levels of methylation at the *AREG* loci cg02334660 and cg26611070, and the *EREG* locus cg19308222 (Table 2 and Supplementary Figure 4). CIMP-high status, determined by clinical method, was again associated with lower *AREG* and *EREG* expression, and with higher methylation of the *AREG* and *EREG* loci (Figure 2C and D). In the TCGA27 cohort, MSI-high status, *BRAF* mutant status, mucinous histology, and right-sided primary tumour were all again associated with significantly lower levels of *AREG* and *EREG* expression, and significantly higher levels of methylation of the *EREG* locus cg19308222 (Supplementary Table 7). In addition, CIMP-high status, as assessed by clustering, was also significantly associated with lower levels of *AREG* and *EREG* expression, and higher levels of *EREG* methylation at cg19308222.

Table 1. Univariate analyses of clinical and pathological characteristics from the MDACC cohort (n = 179)

Variables	n	AREG expression (z-score) Median	P	EREG expression (z-score) Median	P	AREG methylation cg02334660 Median	P	AREG methylation cg26611070 Median	P	EREG methylation cg19308222 Median	P
CIMP status (clinical)			0.005		<0.001		0.243		0.182		<0.001
High	52	−0.144		−0.383		0.459		0.477		0.427	
Low/none	127	0.142		0.187		0.435		0.449		0.207	
CIMP status (cluster)			0.001		<0.001		0.034		0.033		<0.001
High	35	−0.255		−0.393		0.487		0.519		0.452	
Low/none	143	0.134		0.153		0.429		0.446		0.225	
Mucinous			0.005		<0.001		0.004		0.003		<0.001
Yes	29	−0.185		−0.698		0.517		0.547		0.423	
No	130	0.137		0.144		0.425		0.44		0.23	
Location of primary			<0.001		<0.001		<0.001		<0.001		<0.001
Right-sided	82	−0.12		−0.258		0.489		0.518		0.331	
Left-sided	77	0.193		0.252		0.395		0.413		0.2	
MSI status			<0.001		0.002		<0.001		<0.001		0.001
MSI-high	22	−0.293		−0.478		0.646		0.703		0.453	
MSS	93	0.157		0.172		0.408		0.433		0.229	
BRAF V600E			<0.001		<0.001		<0.001		<0.001		<0.001
Mutant	19	−0.275		−0.775		0.542		0.556		0.452	
Wild-type	133	0.154		0.172		0.42		0.442		0.229	
KRAS codon 12/13			0.002		0.006		0.87		0.998		0.051
Mutant	49	−0.061		−0.196		0.448		0.457		0.298	
Wild-type	103	0.163		0.189		0.423		0.453		0.232	
PIK3CA			0.46		0.584		0.056		0.089		0.235
Mutant	21	0.088		0.138		0.482		0.509		0.282	
Wild-type	131	0.064		0.079		0.423		0.444		0.241	
Stage			0.952		0.6		0.281		0.293		0.905
2	74	0.052		0.108		0.466		0.47		0.244	
3	78	0.095		0.012		0.426		0.443		0.268	
4	7	−0.016		−0.317		0.399		0.428		0.229	
Lymphovascular invasion			0.112		0.074		0.168		0.173		0.147
Yes	104	0.117		0.123		0.416		0.442		0.248	
No	55	−0.03		−0.196		0.472		0.492		0.313	

Abbreviations: CIMP = CpG island methylator phenotype; MDACC = MD Anderson Cancer Center; MSI = microsatellite instability. *P*-values obtained by Mann–Whitney *U*-test if variable is dichotomous or Kruskal–Wallis test if >2 variables.

Among the subgroup of 84 patients in the MDACC cohort known to be wild-type both in *BRAF* codon 600 and in *KRAS* codons 12 and 13, right-sided primary tumour location remained significantly associated with lower *AREG* and *EREG* expression, and higher methylation of the *AREG* loci cg02334660 and cg26611070, and the *EREG* locus cg19308222 (Supplementary Table 8 and Supplementary Figure 3). In this subgroup, CIMP-high status as determined by the clinical method was significantly associated with higher methylation levels of *EREG* cg19308222 and was non-significantly associated with lower *EREG* expression. Similarly, in the subgroup of 163 patients in the TCGA450 cohort who were wild-type both in *BRAF* and in *KRAS* codons 12 and 13, MSI-high status, mucinous histology, and right-sided primary tumour remained significantly associated with lower *AREG* and *EREG* expression, and higher levels of methylation at the *AREG* and *EREG* loci. CIMP-high status remained significantly associated with lower *EREG* expression and higher levels of methylation at the *AREG* and *EREG* loci.

Finally, among the subgroup of 134 patients in the TCGA450 cohort who were wild-type in *BRAF*, *KRAS* exons 2–4, and *NRAS*, MSI-high status, mucinous histology, right-sided primary tumour, and CIMP-high status remained significantly associated with lower *AREG* and *EREG* expression, and higher levels of methylation at the *AREG* and *EREG* loci (Supplementary Table 9 and Supplementary Figure 5). Similarly, among the 100 patients in the TCGA27 cohort who were wild-type in *BRAF*, *KRAS*, and *NRAS*, a significant association with lower *EREG* and *AREG*

expression, and higher *EREG* cg19308222 methylation remained with CIMP-high status, right-sided primary tumour, and MSI-high status (Supplementary Table 7).

CIMP-high status is associated with inferior PFS to anti-EGFR therapy in KRAS wild-type patients. An independent cohort of 198 patients with *KRAS* wild-type metastatic CRC enrolled in the ATTACC protocol at MDACC was successfully tested for CIMP status (Supplementary Table 10). Of this group, 173 patients had previously been treated with an anti-EGFR therapy, and PFS with the first anti-EGFR regimen was retrospectively determined in 167 patients. Of this group, 26.3% (44/167) patients were CIMP-high. Compared with the CIMP-low/none group, the CIMP-high group was significantly more likely to have right-sided primary tumour (45.5% *vs* 23.6%, *P* = 0.011), MSI-high status (18.5% *vs* 1.6%, *P* = 0.009), *BRAF* mutation (42.9% *vs* 6.9%, *P* < 0.0001), and male sex (77.3% *vs* 60.2%, *P* = 0.045).

Kaplan–Meier analysis found that among the entire cohort of 167 *KRAS* wild-type patients, inferior PFS with the first anti-EGFR therapy regimen was significantly associated with CIMP-high status (median PFS 4.0 *vs* 6.5 mo, *P* < 0.001), *BRAF* mutation (median PFS 2.8 *vs* 6.5 mo, *P* = 0.004), *NRAS* mutation (median PFS 4.4 *vs* 7.2 mo, *P* = 0.006), and right-sided primary tumour (median PFS 4.7 *vs* 6.5 mo, *P* = 0.040) (Figure 2E and F and Supplementary Figure 6a). These findings were recapitulated on univariate analysis by Cox proportional hazards regression analysis. On multivariate Cox regression
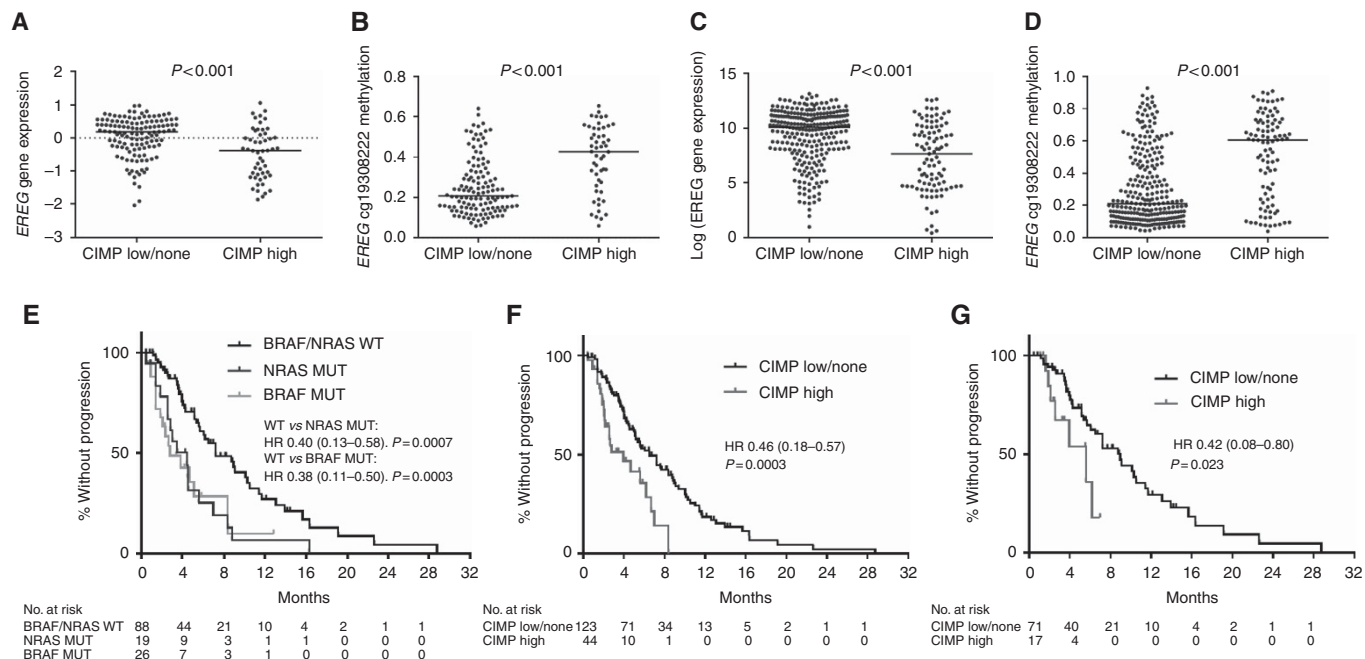
Figure 2. CIMP status is associated with *EREG* methylation and expression, and with the duration of progression-free survival with anti-EGFR therapy. (**A** and **B**) BLiP plots comparing (**A**) z-score of *EREG* expression and (**B**) methylation levels at cg19308222 in CIMP-high *vs* low/none as determined by clinical method in the MDACC cohort ($n = 179$). (**C** and **D**) BLiP plots comparing (**C**) log$_2$-transformed *EREG* expression and (**D**) methylation levels at cg19308222 in CIMP-high *vs* low/none as determined by clinical method in the TCGA450 cohort ($n = 356$). (**E**–**G**) Kaplan–Meier curves of PFS with first anti-EGFR regimen among patients in the MDACC ATTACC cohort (**E**) grouped by *BRAF* and *NRAS* mutation status (for *NRAS* mutant *vs BRAF/NRAS* WT, $P = 0.0007$; for *BRAF* mutant *vs BRAF/NRAS* WT, $P = 0.0003$); (**F**) grouped by CIMP-high *vs* CIMP-low/none status among all patients in the cohort ($n = 167$); or (**G**) grouped by CIMP-high *vs* CIMP-low/none status among the subgroup of patients known to be wild type in *BRAF* and *NRAS* ($n = 88$).

analysis by multiple imputations, *NRAS* mutation (HR 2.27, 95% CI 1.25–4.13, $P = 0.007$), *BRAF* mutation (HR 2.50, 95% CI 1.22–5.13, $P = 0.012$), and CIMP-high status (HR 2.00, 95% CI 1.11–3.64, $P = 0.022$) remained significant. Right-sided primary tumour was not significant (HR 1.43, 95% CI 0.86–2.36, $P = 0.167$) (Table 3). In the subgroup with *BRAF* and *NRAS* wild-type disease, inferior PFS with the first anti-EGFR regimen remained significantly associated with CIMP-high status (median PFS 5.6 *vs* 9.0 mo, $P = 0.023$) and trended with right-sided primary tumour (median PFS 5.6 *vs* 9.0 mo, $P = 0.053$) on Kaplan–Meier analysis (Figure 2G and Supplementary Figure 6b).

Notably, on univariate analysis of OS among the 198 patients, with 162 events, CIMP-high status (HR 1.53, 95% CI 1.08–2.16), right-sided primary tumour (HR 1.45, 95% CI 1.04–2.01), *BRAF* mutation (HR 2.46, 95% CI 1.61–3.75), and *NRAS* mutation (HR 1.70, 95% CI 1.03–2.81) were significantly associated with inferior survival. However, neither CIMP-high status nor right-sided primary tumour were significantly associated with OS on multivariate analysis among the entire cohort or on univariate analysis among the subgroup of 109 patients known to be wild-type in *BRAF* and *NRAS*.

DISCUSSION

Our study is one of the first to identify CIMP status, and more specifically methylation of loci within the *EREG* and *AREG* promoters, as a determinant of *EREG* and *AREG* expression levels and a prognostic biomarker with regards to PFS upon treatment with anti-EGFR therapy in patients with metastatic CRC. Given the association we and others have observed between *EREG* levels and the site of primary tumour, this provides additional explanation for

the clinical differences observed in right- *vs* left-sided primary tumours. Although our data are retrospective and are not derived from clinical trials, the use of multiple independent data sets available through the TCGA to validate our findings is a strength of our study.

Although methylation of CpG islands within gene promoters is generally known to be associated with downregulation of expression, it remains unclear which methylated loci are most important in driving differential expression. Although the TCGA Pan-Cancer project has identified for each gene the single CpG island most inversely correlated with expression, our results extended these results specifically to CRCs. For *EREG*, we found that the CpG island cg19308222, which was most strongly inversely correlated with expression in the Pan-Cancer project, was again the most inversely correlated CpG island in CRC. However, for *AREG*, we found that the CpG islands cg02334660 and cg26611070 had the strongest inverse correlation with expression in CRC, while in the Pan-Cancer project, cg03244277 had the strongest inverse correlation. Indeed, both cg02334660 and cg26611070 are located within the body of the *AREG* gene. Although gene-body methylation is thought to be associated with increased expression (Yang *et al*, 2014), body methylation may conversely function to repress activation of intragenic promoters and functions in a tissue-specific context (Maunakea *et al*, 2010; Jones, 2012). The tissue-specific effects of methylation at specific loci in specific genes requires further investigation, as this may explain the differences in which locus was selected as the most anti-correlated with expression.

The significant contribution of *EREG* and *AREG* methylation on gene expression has not been previously well described. A small *in vitro* study of gastric cancer cell lines revealed inverse correlation between *EREG* promoter methylation and expression (Yun *et al*, 2012). Indeed, iBAG analysis demonstrated that ~57–65% of

Table 2. Univariate analyses of clinical and pathological characteristics from the TCGA450 cohort (n = 356)

Variable	n	AREG expression (log$_2$)		EREG expression (log$_2$)		AREG methylation cg02334660		AREG methylation cg26611070		EREG methylation cg19308222	
		Median	P	Median	P	Median	P	Median	P	Median	P
CIMP status			<0.001		<0.001		<0.001		<0.001		<0.001
High	97	9.474		7.657		0.541		0.572		0.605	
Low/none	259	10.137		10.096		0.432		0.452		0.211	
Location of primary			<0.001		<0.001		<0.001		<0.001		<0.001
Right-sided	169	9.597		8.408		0.508		0.5339		0.4466	
Left-sided	179	10.299		10.346		0.4346		0.4481		0.1859	
MSI status			<0.001		<0.001		<0.001		<0.001		<0.001
MSI-high	52	8.281		5.135		0.7399		0.7882		0.6063	
MSS	302	10.139		10.059		0.4377		0.4579		0.2279	
BRAF V600E			<0.001		<0.001		<0.001		<0.001		<0.001
Mutant	26	8.175		4.696		0.6727		0.6946		0.6581	
Wild-type	262	10.147		10.104		0.4536		0.4689		0.2331	
KRAS codon 12/13			0.331		0.313		0.862		0.854		0.133
Mutant	99	9.714		9.174		0.475		0.492		0.338	
Wild-type	189	10.136		10.034		0.467		0.479		0.239	
KRAS exons 2–4			0.249		0.206		0.374		0.401		0.335
Mutant	114	9.728		9.005		0.441		0.47		0.324	
Wild-type	174	10.16		10.146		0.473		0.49		0.239	
NRAS			0.725		0.763		0.153		0.2		0.618
Mutant	15	10.066		10.349		0.4371		0.4556		0.2851	
Wild-type	273	10.063		9.661		0.4747		0.4923		0.2503	
PIK3CA			0.566		0.94		0.834		0.714		0.458
Mutant	65	9.805		9.621		0.4812		0.4923		0.3232	
Wild-type	223	10.074		9.757		0.4621		0.4844		0.2464	
Stage			0.575		0.423		0.532		0.497		0.372
1	53	9.874		8.685		0.475		0.506		0.504	
2	125	9.742		9.464		0.458		0.468		0.28	
3	112	9.921		9.449		0.481		0.476		0.284	
4	49	10.299		10.26		0.471		0.479		0.209	
Lymphovascular invasion			0.443		0.871		0.936		0.961		0.536
Yes	91	9.846		9.661		0.498		0.52		0.285	
No	217	9.903		9.464		0.48		0.497		0.282	
Mucinous			<0.001		<0.001		<0.001		<0.001		<0.001
Yes	39	8.77		6.125		0.544		0.579		0.561	
No	311	10.102		9.919		0.452		0.468		0.239	

Abbreviations: CIMP = CpG island methylator phenotype; MSI = microsatellite instability; TCGA = The Cancer Genome Atlas. *P*-values obtained by Mann–Whitney *U*-test if variable is dichotomous or Kruskal–Wallis test if >2 variables.

variation of expression in *EREG* was attributable to methylation of cg19308222, and 33–35% of variation of expression in *AREG* was attributable to methylation of the *AREG* loci cg02334660 or cg2661070. Furthermore, treatment of CRC cell lines with demethylating agents indeed resulted in decreased methylation at the cg19308222 locus and increased expression of *EREG*. This provides the strongest evidence to date that methylation of these important EGFR ligands is a significant likely mechanism of regulation of expression, and also suggests that this mechanism can be manipulated by treatment with hypomethylating agents.

Several studies have already established that high expression of *EREG* and *AREG* is associated with improved outcomes with anti-EGFR therapy in refractory metastatic CRC. A study of tumours from 110 metastatic CRC patients treated with cetuximab found that high *EREG* or *AREG* expression was associated with longer PFS (HR 0.47, *P* = 0.0002; and HR 0.44, *P* < 0.0001, respectively) (Khambata-Ford *et al*, 2007), although it was unclear whether there was an interaction of *EREG* and *AREG* expression level with *KRAS* or *BRAF* mutation status. Another study in 121 irinotecan-refractory *KRAS* wild-type metastatic CRC patients who received anti-EGFR therapy on clinical trials found that *EREG* and *AREG* expression was significantly associated with response rate, disease control, PFS, and OS with anti-EGFR therapy, with the predictive

value of *EREG* expression superior to that of *AREG* expression (Jacobs *et al*, 2009). Analysis from the CO.17 trial of cetuximab in patients with refractory metastatic CRC found cetuximab had a larger effect in improvement of OS and PFS in *KRAS* wild-type and *EREG* expression-high patients (Jonker *et al*, 2014). Similarly, results from the PICCOLO trial of second-line panitumumab and irinotecan in metastatic CRC found significant improvement in PFS in patients with either *EREG* or *AREG* expression in the highest tertile (Seligmann *et al*, 2016). Finally, a retrospective evaluation of expression of 110 candidate genes from 144 primary tumours of *KRAS* wild-type refractory metastatic CRC patients found that high *EREG* and *AREG* expression was strongly associated with improved PFS and disease control rate with cetuximab (Baker *et al*, 2011). In contrast, two first-line studies failed to demonstrate that *EREG* or *AREG* expression is predictive of benefit with cetuximab when combined with multi-agent chemotherapy backbones (Adams *et al*, 2012; Cushman *et al*, 2015). Moreover, these studies did not further describe additional clinical and pathological variables that we now know may have impacted *EREG* and *AREG* expression.

Additional data reveal that patients whose tumours have high *EREG* and *AREG* expression are more likely to have clinical and tumour biological characteristics similar to those we found. Indeed,

Table 3. Univariate Cox regression analysis and multivariate Cox regression analysis after multiple imputations of MDACC ATTACC cohort for PFS

Variable	Hazard ratio (95% CI)	P
Univariate analysis		
CIMP-high	2.38 (1.47–3.85)	0.00043
Right-sided	1.56 (1.01–2.41)	0.043
BRAF mutant	2.14 (1.26–3.65)	0.005
NRAS mutant	2.12 (1.23–3.65)	0.007
PIK3CA mutant	0.97 (0.53–1.77)	0.92
PTEN loss	1.07 (0.57–2.03)	0.828
MSI-high	1.27 (0.39–4.12)	0.691
Non-white	1.24 (0.78–1.96)	0.362
Age	0.98 (0.96–1.00)	0.055
Male sex	1.07 (0.72–1.60)	0.727
Prior bevacizumab before EGFR	1.26 (0.76–2.07)	0.368
EGFR monotherapy	1.49 (0.75–2.97)	0.26
No. of prior chemo regimens	1.18 (0.91–1.53)	0.203
Multivariate analysis[a]		
CIMP-high	2.00 (1.11–3.64)	0.022
Right-sided	1.43 (0.86–2.36)	0.167
BRAF mutant	2.50 (1.22–5.13)	0.012
NRAS mutant	2.27 (1.25–4.13)	0.007

Abbreviations: ATTACC = Assessment of Targeted Therapies Against Colorectal Cancer; CI-confidence interval; CIMP = CpG island methylator phenotype; MDACC = MD Anderson Cancer Center; PFS = progression-free survival.
[a]Adjusted for the number of prior regimens, age, prior bevacizumab, and use of anti-EGFR as monotherapy or in combination with chemotherapy.
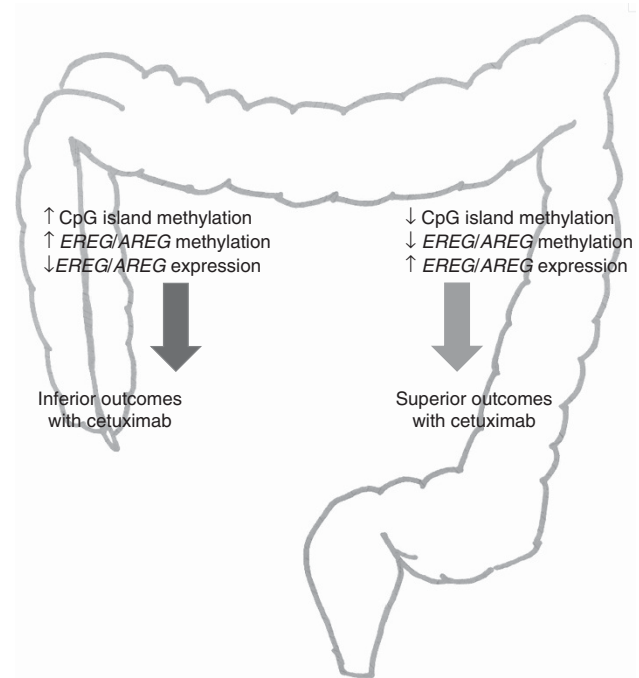

Figure 3. Left- and right-sided primary CRCs have distinct pathobiology, with different rates of CIMP, and contrasting levels of methylation and gene expression of *EREG* and *AREG*, providing a unifying explanation for differences in outcomes with anti-EGFR therapy like cetuximab.

data from 952/1630 samples from the randomised phase III COIN (addition of cetuximab to oxaliplatin-based first-line combination chemotherapy for treatment of advanced CRC) trial showed that high expression of *EREG* and *AREG* was significantly associated with wild-type *KRAS*, wild-type *BRAF*, left-sided primary colon tumour, and microsatellite stable disease (Adams *et al*, 2012). An independent study also found that left-sided primary colon carcinomas were more likely to have epiregulin overexpression (Missiaglia *et al*, 2014). Moreover, data from 331/696 samples from the randomised PICCOLO (Panitumumab, irinotecan, and ciclosporin in CRC) trial in second-line treatment of metastatic CRC similarly found that high expression of *EREG* and *AREG* was associated with wild-type *BRAF* and left-sided primary colon tumour (Seligmann *et al*, 2016).

Also, multiple trials showed that patients with *KRAS* wild-type left-sided CRC have improved outcomes upon treatment with cetuximab than those with right-sided CRC. In the AIO KRK-0104 trial, 146 patients were randomised to receive first-line CAPOX with cetuximab or CAPIRI (capecitabine/irinotecan) with cetuximab. In this trial, 95/146 patients had *KRAS* codon 12 and 13 wild-type tumour, with 68 left-sided and 27 right-sided primary tumours. Patients with left-sided *KRAS* exon 2 wild-type cancers had superior OS (HR 0.63, $P = 0.016$) and PFS (HR 0.67, $P = 0.02$) than patients with right-sided tumours. In patients with *KRAS/BRAF* wild-type tumours ($n = 79$), median PFS was 8.2 *vs* 5.9 mo (HR 0.81, $P = 0.47$) and median OS was 27.3 *vs* 16.2 mo (HR 0.60, $P = 0.11$; Von Einem *et al*, 2014). In the CO.17 study, 399 patients with metastatic CRC were randomised to receive cetuximab or best supportive care. In the cohort of patients with *KRAS* exon 2 wild-type status, there was a significant interaction between site of primary tumour and PFS benefit with cetuximab (for left-sided tumours, HR 0.28 and $P < 0.0001$; for right-sided tumours, HR 0.73 and $P = 0.26$), with predictive effect interaction term ($P = 0.002$) (Brule *et al*, 2013; Brule *et al*, 2015).

Notably, although CIMP-high status and right-sided primary CRC were both associated with significantly inferior OS on univariate analysis, neither remained significant after adjusting for *NRAS* and *BRAF* status. As CIMP-high status remained significantly associated with PFS with anti-EGFR therapy after adjusting

for these variables, this suggests that the inferior anti-EGFR PFS of CIMP-high patients is not solely due to an overall inferior prognosis, at least among the *KRAS/NRAS/BRAF* wild-type subpopulation. Indeed, several studies have noted inferior overall prognosis among patients with low *EREG/AREG* expression (Pentheroudakis *et al*, 2013; Stahler *et al*, 2016) or right-sided primary CRC (Loupakis *et al*, 2015). These studies are variable in whether they included only patients with *KRAS* wild-type tumours, whether they limited analysis to patients with anti-EGFR therapy only, and whether *EREG* and *AREG* were treated as a continuous or categorical variable and how any cut points were defined. Furthermore, none included CIMP status in their multivariable analyses. Furthermore, recent data from the PICCOLO study show that although high *EREG* and *AREG* levels were associated with improved OS among all patients and among the RAS wild-type subgroup, neither was significantly associated with OS among the RAS and *BRAF* wild-type subgroup. In that study, site of primary tumour was not associated with PFS or OS (Seligmann *et al*, 2016). Additional studies to confirm these findings in prospectively collected data will be important in determining the predictive *vs* prognostic role of these variables.

Previously, there had not been data for a plausible biological explanation for the differential outcomes of left- and right-sided primary CRCs with cetuximab-based treatment, but our findings provide evidence that variation in *EREG* and *AREG* methylation, and expression contribute to these differences (Figure 3). Although there is ample existing evidence of the association of right-sided primary cancers with CIMP-high status, our analysis provides a novel association between right-sided and CIMP-high CRCs with increased methylation specifically of the critical CpG islands within *EREG* and *AREG* that have the strongest inverse association with *EREG* and *AREG* expression. Accordingly, there is also an association with low *EREG* and *AREG* expression. Several studies have corroborated that low *EREG* and *AREG* expression are associated with inferior outcomes with cetuximab therapy, even if there is conflicting evidence whether *EREG* and *AREG* are

predictive or merely prognostic. Nevertheless, these data are consistent with our retrospective analysis showing that CIMP-high status is associated with inferior PFS with anti-EGFR-including therapy.

Although our findings unify disparate observations and associations between location of primary tumour, expression of *EREG* and *AREG*, and differential responses to cetuximab, they are based on retrospective analysis of several different data sets and thus cannot ascertain causality. Prospective incorporation of these variables into the analysis of adequately powered future randomised controlled trials of anti-EGFR therapies in first-line and refractory metastatic CRC will be necessary to determine whether CIMP status and methylation level of *EREG* and *AREG* are truly predictive of response. Further investigation is warranted into whether differential methylation and expression of additional genes in CIMP-high patients contribute to resistance to anti-EGFR therapy.

ACKNOWLEDGEMENTS

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Adams RA, Fisher D, Farragher S, Jasani B, Smith CG, James MD, Cheadle J, Nichols LL, Meade AM, Kaplan RS, Wilson RH, Wasan H, Maughan T (2012) Use of epiregulin (EREG) and amphiregulin (AREG) gene expression to predict response to cetuximab (cet) in combination with oxaliplatin (Ox) and 5FU in the first-line treatment of advanced colorectal cancer (aCRC). *J Clin Oncol* **30**: abstr 32.

Amado RG, Wolf M, Peeters M, Van Cutsem E, Siena S, Freeman DJ, Juan T, Sikorski R, Suggs S, Radinsky R, Patterson SD, Chang DD (2008) Wild-type KRAS is required for panitumumab efficacy in patients with metastatic colorectal cancer. *J Clin Oncol* **26**: 1626–1634.

Aranyi T, Varadi A, Simon I, Tusnady GE (2006) The BiSearch web server. *BMC Bioinformatics* **7**: 431.

Baker JB, Dutta D, Watson D, Maddala T, Munneke BM, Shak S, Rowinsky EK, Xu LA, Harbison CT, Clark EA, Mauro DJ, Khambata-Ford S (2011) Tumour gene expression predicts response to cetuximab in patients with KRAS wild-type metastatic colorectal cancer. *Br J Cancer* **104**: 488–495.

Bokemeyer C, Bondarenko I, Hartmann JT, De Braud F, Schuch G, Zubel A, Celik I, Schlichting M, Koralewski P (2011) Efficacy according to biomarker status of cetuximab plus FOLFOX-4 as first-line treatment for metastatic colorectal cancer: the OPUS study. *Ann Oncol* **22**: 1535–1546.

Bokemeyer C, Van Cutsem E, Rougier P, Ciardiello F, Heeger S, Schlichting M, Celik I, Kohne CH (2012) Addition of cetuximab to chemotherapy as first-line treatment for KRAS wild-type metastatic colorectal cancer: pooled analysis of the CRYSTAL and OPUS randomised clinical trials. *Eur J Cancer* **48**: 1466–1475.

Broad Institute TCGA Genome Data Analysis Center (2014a) Correlation between mRNA expression and DNA methylation. *Broad Institute of MIT and Harvard* doi:10.7908/C14M93GF.

Broad Institute TCGA Genome Data Analysis Center (2014b) Correlation between mRNA expression and DNA methylation. *Broad Institute of MIT and Harvard* doi:10.7908/C18051FP.

Brule SY, Jonker DJ, Karapetis CS, O'callaghan CJ, Malcolm J, Moore RW, Tebbutt NC, Underhill C, Yip D, Zalcberg JR, Tu D, Goodwin RA (2013) Location of colon cancer (right-sided [RC] versus left-sided [LC]) as a

predictor of benefit from cetuximab (CET): NCIC CTG CO.17. *J Clin Oncol* **31**: abstr 3528.

Brule SY, Jonker DJ, Karapetis CS, O'callaghan CJ, Moore MJ, Wong R, Tebbutt NC, Underhill C, Yip D, Zalcberg JR, Tu D, Goodwin RA (2015) Location of colon cancer (right-sided versus left-sided) as a prognostic factor and a predictor of benefit from cetuximab in NCIC CO.17. *Eur J Cancer* **51**: 1405–1414.

Cushman SM, Jiang C, Hatch AJ, Shterev I, Sibley AB, Niedzwiecki D, Venook AP, Owzar K, Hurwitz HI, Nixon AB (2015) Gene expression markers of efficacy and resistance to cetuximab treatment in metastatic colorectal cancer: results from CALGB 80203 (Alliance). *Clin Cancer Res* **21**: 1078–1086.

De Roock W, Claes B, Bernasconi D, De Schutter J, Biesmans B, Fountzilas G, Kalogeras KT, Kotoula V, Papamichael D, Laurent-Puig P, Penault-Llorca F, Rougier P, Vincenzi B, Santini D, Tonini G, Cappuzzo F, Frattini M, Molinari F, Saletti P, De Dosso S, Martini M, Bardelli A, Siena S, Sartore-Bianchi A, Tabernero J, Macarulla T, Di Fiore F, Gangloff AO, Ciardiello F, Pfeiffer P, Qvortrup C, Hansen TP, Van Cutsem E, Piessevaux H, Lambrechts D, Delorenzi M, Tejpar S (2010) Effects of KRAS, BRAF, NRAS, and PIK3CA mutations on the efficacy of cetuximab plus chemotherapy in chemotherapy-refractory metastatic colorectal cancer: a retrospective consortium analysis. *Lancet Oncol* **11**: 753–762.

Douillard JY, Oliner KS, Siena S, Tabernero J, Burkes R, Barugel M, Humblet Y, Bodoky G, Cunningham D, Jassem J, Rivera F, Kocakova I, Ruff P, Blasinska-Morawiec M, Smakal M, Canon JL, Rother M, Williams R, Rong A, Wiezorek J, Sidhu R, Patterson SD (2013) Panitumumab-FOLFOX4 treatment and RAS mutations in colorectal cancer. *N Engl J Med* **369**: 1023–1034.

Du P, Zhang X, Huang CC, Jafari N, Kibbe WA, Hou L, Lin SM (2010) Comparison of beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* **11**: 587.

Hagemann S, Heil O, Lyko F, Brueckner B (2011) Azacytidine and decitabine induce gene-specific and non-random DNA demethylation in human cancer cell lines. *PLoS One* **6**: e17388.

Jacobs B, De Roock W, Piessevaux H, Van Oirbeek R, Biesmans B, De Schutter J, Fieuws S, Vandesompele J, Peeters M, Van Laethem JL, Humblet Y, Penault-Llorca F, De Hertogh G, Laurent-Puig P, Van Cutsem E, Tejpar S (2009) Amphiregulin and epiregulin mRNA expression in primary tumors predicts outcome in metastatic colorectal cancer treated with cetuximab. *J Clin Oncol* **27**: 5068–5074.

Jennings EM, Morris JS, Carroll RJ, Manyam GC, Baladandayuthapani V (2013) Bayesian methods for expression-based integration of various types of genomics data. *EURASIP J Bioinform Syst Biol* **2013**: 13.

Jones PA (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet* **13**: 484–492.

Jonker DJ, Karapetis CS, Harbison C, O'callaghan CJ, Tu D, Simes RJ, Malone DP, Langer C, Tebbutt N, Price TJ, Shapiro J, Siu LL, Wong RP, Bjarnason G, Moore MJ, Zalcberg JR, Khambata-Ford S (2014) Epiregulin gene expression as a biomarker of benefit from cetuximab in the treatment of advanced colorectal cancer. *Br J Cancer* **110**: 648–655.

Karapetis CS, Jonker D, Daneshmand M, Hanson JE, O'callaghan CJ, Marginean C, Zalcberg JR, Simes J, Moore MJ, Tebbutt NC, Price TJ, Shapiro JD, Pavlakis N, Gibbs P, Van Hazel GA, Lee U, Haq R, Virk S, Tu D, Lorimer IA (2014) PIK3CA, BRAF, and PTEN status and benefit from cetuximab in the treatment of advanced colorectal cancer–results from NCIC CTG/AGITG CO.17. *Clin Cancer Res* **20**: 744–753.

Karapetis CS, Khambata-Ford S, Jonker DJ, O'callaghan CJ, Tu D, Tebbutt NC, Simes RJ, Chalchal H, Shapiro JD, Robitaille S, Price TJ, Shepherd L, Au HJ, Langer C, Moore MJ, Zalcberg JR (2008) K-ras mutations and benefit from cetuximab in advanced colorectal cancer. *N Engl J Med* **359**: 1757–1765.

Khambata-Ford S, Garrett CR, Meropol NJ, Basik M, Harbison CT, Wu S, Wong TW, Huang X, Takimoto CH, Godwin AK, Tan BR, Krishnamurthi SS, Burris 3rd HA, Poplin EA, Hidalgo M, Baselga J, Clark EA, Mauro DJ (2007) Expression of epiregulin and amphiregulin and K-ras mutation status predict disease control in metastatic colorectal cancer patients treated with cetuximab. *J Clin Oncol* **25**: 3230–3237.

Laurent-Puig P, Cayre A, Manceau G, Buc E, Bachet JB, Lecomte T, Rougier P, Lievre A, Landi B, Boige V, Ducreux M, Ychou M, Bibeau F, Bouche O, Reid J, Stone S, Penault-Llorca F (2009) Analysis of PTEN, BRAF, and

EGFR status in determining benefit from cetuximab therapy in wild-type KRAS metastatic colon cancer. *J Clin Oncol* **27**: 5924–5930.

Li H, Chiappinelli KB, Guzzetta AA, Easwaran H, Yen RW, Vatapalli R, Topper MJ, Luo J, Connolly RM, Azad NS, Stearns V, Pardoll DM, Davidson N, Jones PA, Slamon DJ, Baylin SB, Zahnow CA, Ahuja N (2014) Immune regulation by low doses of the DNA methyltransferase inhibitor 5-azacitidine in common human epithelial cancers. *Oncotarget* **5**: 587–598.

Loupakis F, Pollina L, Stasi I, Ruzzo A, Scartozzi M, Santini D, Masi G, Graziano F, Cremolini C, Rulli E, Canestrari E, Funel N, Schiavon G, Petrini I, Magnani M, Tonini G, Campani D, Floriani I, Cascinu S, Falcone A (2009) PTEN expression and KRAS mutations on primary tumors and metastases in the prediction of benefit from cetuximab plus irinotecan for patients with metastatic colorectal cancer. *J Clin Oncol* **27**: 2622–2629.

Loupakis F, Yang D, Yau L, Feng S, Cremolini C, Zhang W, Maus MK, Antoniotti C, Langer C, Scherer SJ, Muller T, Hurwitz HI, Saltz L, Falcone A, Lenz HJ (2015) Primary tumor location as a prognostic factor in metastatic colorectal cancer. *J Natl Cancer Inst* **107**: dju427.

Maunakea AK, Nagarajan RP, Bilenky M, Ballinger TJ, D'souza C, Fouse SD, Johnson BE, Hong C, Nielsen C, Zhao Y, Turecki G, Delaney A, Varhol R, Thiessen N, Shchors K, Heine VM, Rowitch DH, Xing X, Fiore C, Schillebeeckx M, Jones SJ, Haussler D, Marra MA, Hirst M, Wang T, Costello JF (2010) Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* **466**: 253–257.

Missiaglia E, Jacobs B, D'ario G, Di Narzo AF, Soneson C, Budinska E, Popovici V, Vecchione L, Gerster S, Yan P, Roth AD, Klingbiel D, Bosman FT, Delorenzi M, Tejpar S (2014) Distal and proximal colon cancers differ in terms of molecular, pathological, and clinical features. *Ann Oncol* **25**: 1995–2001.

Pentheroudakis G, Kotoula V, De Roock W, Kouvatseas G, Papakostas P, Makatsoris T, Papamichael D, Xanthakis I, Sgouros J, Televantou D, Kafiri G, Tsamandas AC, Razis E, Galani E, Bafaloukos D, Efstratiou I, Bompolaki I, Pectasides D, Pavlidis N, Tejpar S, Fountzilas G (2013) Biomarkers of benefit from cetuximab-based therapy in metastatic colorectal cancer: interaction of EGFR ligand expression with RAS/RAF, PIK3CA genotypes. *BMC Cancer* **13**: 49.

Seligmann JF, Elliott F, Richman SD, Jacobs B, Hemmings G, Brown S, Barrett JH, Tejpar S, Quirke P, Seymour MT (2016) Combined epiregulin and amphiregulin expression levels as a predictive biomarker for panitumumab therapy benefit or lack of benefit in patients with RAS wild-type advanced colorectal cancer. *JAMA Oncol*; e-pub ahead of print 11 February 2016; doi:10.1001/jamaoncol.2015.6065.

Stahler A, Heinemann V, Giessen-Jung C, Crispin A, Schalhorn A, Stintzing S, Fischer Von Weikersthal L, Vehling-Kaiser U, Stauch M, Quietzsch D, Held S, Von Einem JC, Holch J, Neumann J, Kirchner T, Jung A, Modest DP (2016) Influence of mRNA expression of epiregulin and amphiregulin on outcome of patients with metastatic colorectal cancer treated with 5-FU/LV plus irinotecan or irinotecan plus oxaliplatin as first-line treatment (FIRE 1-trial). *Int J Cancer* **138**: 739–746.

The Cancer Genome Atlas Network (2012) Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**: 330–337.

The Cancer Genome Atlas Research NetworkWeinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, Ellrott K, Shmulevich I, Sander C, Stuart JM (2013) The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet* **45**: 1113–1120.

Tol J, Dijkstra JR, Klomp M, Teerenstra S, Dommerholt M, Vink-Borger ME, Van Cleef PH, Van Krieken JH, Punt CJ, Nagtegaal ID (2010) Markers for EGFR pathway activation as predictor of outcome in metastatic colorectal cancer patients treated with or without cetuximab. *Eur J Cancer* **46**: 1997–2009.

Toyota M, Ahuja N, Ohe-Toyota M, Herman JG, Baylin SB, Issa JP (1999) CpG island methylator phenotype in colorectal cancer. *Proc Natl Acad Sci USA* **96**: 8681–8686.

Tusnady GE, Simon I, Varadi A, Aranyi T (2005) BiSearch: primer-design and search tool for PCR on bisulfite-treated genomes. *Nucleic Acids Res* **33**: e9.

Van Cutsem E, Kohne CH, Hitre E, Zaluski J, Chang Chien CR, Makhson A, D'haens G, Pinter T, Lim R, Bodoky G, Roh JK, Folprecht G, Ruff P, Stroh C, Tejpar S, Schlichting M, Nippgen J, Rougier P (2009) Cetuximab and chemotherapy as initial treatment for metastatic colorectal cancer. *N Engl J Med* **360**: 1408–1417.

Von Einem JC, Heinemann V, Von Weikersthal LF, Vehling-Kaiser U, Stauch M, Hass HG, Decker T, Klein S, Held S, Jung A, Kirchner T, Haas M, Holch J, Michl M, Aubele P, Boeck S, Schulz C, Giessen C, Stintzing S, Modest DP (2014) Left-sided primary tumors are associated with favorable prognosis in patients with KRAS codon 12/13 wild-type metastatic colorectal cancer treated with cetuximab plus chemotherapy: an analysis of the AIO KRK-0104 trial. *J Cancer Res Clin Oncol* **140**: 1607–1614.

Wang W, Baladandayuthapani V, Morris JS, Broom BM, Manyam G, Do KA (2013) iBAG: integrative Bayesian analysis of high-dimensional multiplatform genomics data. *Bioinformatics* **29**: 149–159.

Weisenberger DJ, Siegmund KD, Campan M, Young J, Long TI, Faasse MA, Kang GH, Widschwendter M, Weener D, Buchanan D, Koh H, Simms L, Barker M, Leggett B, Levine J, Kim M, French AJ, Thibodeau SN, Jass J, Haile R, Laird PW (2006) CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nat Genet* **38**: 787–793.

Yamauchi M, Lochhead P, Morikawa T, Huttenhower C, Chan AT, Giovannucci E, Fuchs C, Ogino S (2012) Colorectal cancer: a tale of two sides or a continuum? *Gut* **61**: 794–797.

Yang X, Han H, De Carvalho DD, Lay FD, Jones PA, Liang G (2014) Gene body methylation can alter gene expression and is a therapeutic target in cancer. *Cancer Cell* **26**: 577–590.

Yun J, Song SH, Park J, Kim HP, Yoon YK, Lee KH, Han SW, Oh DY, Im SA, Bang YJ, Kim TY (2012) Gene silencing of EREG mediated by DNA methylation and histone modification in human gastric cancers. *Lab Invest* **92**: 1033–1044.

Supplementary Information accompanies this paper on British Journal of Cancer website (http://www.nature.com/bjc)