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

Michael S Lee¹,², Elizabeth J McGuffey³, Jeffrey S Morris⁴, Ganiraju Manyam⁴, Veerabahdram Baladandayuthapani⁴, Wei Wei⁴, Van K Morris⁵, Michael J Overman⁵, Dipen M Maru⁶, Zhi-Qin Jiang⁵, Stanley R Hamilton⁶ and Scott Kopetz *,⁵

¹Division of Hematology/Oncology, Department of Medicine, University of North Carolina at Chapel Hill, 170 Manning Drive, CB 7305, Chapel Hill, NC 27599, USA; ²Lineberger Comprehensive Cancer Center, 170 Manning Drive, CB 7295, Chapel Hill, NC 27599, USA; ³United States Naval Academy, Department of Mathematics, 572C Holloway Road, Annapolis, MD 21402, USA; ⁴Department of Biostatistics, University of Texas MD Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX 77030, USA; ⁵Department of Gastrointestinal Medical Oncology, University of Texas MD Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX 77030, USA and ⁶Department 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 AREG and EREG, which encode the EGFR ligands epiregulin and amphiregulin (Kambata-Ford et al., 2007; Jacobs et al., 2009). AREG and EREG 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 AREG and EREG may indicate tumour cell dependence on an autocrine EGFR-activating loop, and thus predict for increased susceptibility to anti-EGFR therapy (Kambata-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-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 ofanked 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 \( \geq 40\% \) 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 profiling using methodology parallelizing 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 BeadChip 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.
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 (P = 0.726, P < 10\(^{-6}\); Figure 1A). The strongest correlation with AREG expression was with the CpG islands cg26611070 (P = 0.520, P < 10\(^{-6}\)) and cg022334660 (P = 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 (P = 0.671, P < 10\(^{-6}\); Figure 1B), and the strongest correlation with AREG expression was again with the CpG island cg26334660 (P = 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 (P = 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 cg022334660 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).
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 cg29334660 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).

**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.

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-high status 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 all associated with significantly 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.
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.

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

---

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

| Variables                              | AREG expression (z-score) | EREG expression (z-score) | AREG methylation cg02334660 | EREG methylation cg26611070 | EREG methylation cg19308222 |
|----------------------------------------|---------------------------|---------------------------|-----------------------------|----------------------------|----------------------------|
|                                        | n       | Median     | P    | Median     | P    | Median | P    | Median | P    | Median | P    |
| CIMP status (clinical)                 |         |            |      |            |      |        |      |        |      |        |      |
| High                                   | 52      | –0.144     | 0.005 | –0.383     | <0.001 | 0.459  | 0.243 | 0.477  | 0.182 | 0.427  | 0.207 |
| Low/none                               | 127     | 0.142      |      | 0.187      |      | 0.435  |      | 0.449  |      | 0.427  |      |
| CIMP status (cluster)                  |         |            |      |            |      |        |      |        |      |        |      |
| High                                   | 35      | –0.255     | 0.001 | –0.393     | <0.001 | 0.487  | 0.034 | 0.519  | 0.033 | 0.452  | <0.001 |
| Low/none                               | 143     | 0.134      |      | 0.153      |      | 0.429  |      | 0.446  |      | 0.225  |      |
| Mucinous                               |         |            |      |            |      |        |      |        |      |        |      |
| Yes                                    | 29      | –0.185     | 0.005 | –0.698     | <0.001 | 0.517  | 0.004 | 0.547  | 0.003 | 0.423  | <0.001 |
| No                                     | 130     | 0.137      |      | 0.144      |      | 0.425  |      | 0.44   |      | 0.23   |      |
| Location of primary tumour             |         |            |      |            |      |        |      |        |      |        |      |
| Right-sided                            | 82      | –0.12      | <0.001 | –0.258     | <0.001 | 0.489  | 0.001 | 0.518  | 0.001 | 0.331  | <0.001 |
| Left-sided                             | 77      | 0.193      |      | 0.252      |      | 0.395  |      | 0.413  |      | 0.2    |      |
| MSI status                             |         |            |      |            |      |        |      |        |      |        |      |
| MSI-high                               | 22      | –0.293     | <0.001 | –0.478     | <0.001 | 0.646  | 0.001 | 0.703  | 0.001 | 0.453  | 0.001 |
| MSS                                     | 93      | 0.157      |      | 0.172      |      | 0.408  |      | 0.433  |      | 0.229  |      |
| BRAF V600E                             |         |            |      |            |      |        |      |        |      |        |      |
| Mutant                                 | 19      | –0.275     | <0.001 | –0.775     | <0.001 | 0.542  | 0.001 | 0.556  | 0.001 | 0.452  | <0.001 |
| Wild-type                              | 133     | 0.154      |      | 0.172      |      | 0.42   |      | 0.442  |      | 0.229  |      |
| KRAS codon 12/13                       |         |            |      |            |      |        |      |        |      |        |      |
| Mutant                                 | 49      | –0.061     | 0.002 | –0.196     | 0.006 | 0.448  | 0.076 | 0.457  | 0.001 | 0.298  | 0.051 |
| Wild-type                              | 103     | 0.163      |      | 0.189      |      | 0.423  |      | 0.453  |      | 0.232  |      |
| PIK3CA                                 |         |            |      |            |      |        |      |        |      |        |      |
| Mutant                                 | 21      | 0.088      | 0.46  | 0.138      | 0.584 | 0.482  | 0.056 | 0.509  | 0.089 | 0.282  | 0.235 |
| Wild-type                              | 131     | 0.064      |      | 0.079      |      | 0.423  |      | 0.444  |      | 0.241  |      |
| Stage                                  |         |            |      |            |      |        |      |        |      |        |      |
| 2                                      | 74      | 0.052      | 0.952 | 0.108      | 0.6   | 0.466  | 0.281 | 0.47   | 0.293 | 0.244  | 0.905 |
| 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                |         |            |      |            |      |        |      |        |      |        |      |
| Yes                                    | 104     | 0.117      | 0.112 | 0.123      | 0.074 | 0.416  | 0.168 | 0.442  | 0.173 | 0.248  | 0.147 |
| 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.
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
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 cg26611070. Furthermore, treatment of CRC cell lines with demethylating agents indeed resulted in decreased methylation at the cg19308222 locus and increased expression of AREG. 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,
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 ceclopisin 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 wild-type tumour, with 68 left-sided and 27 right-sided primary tumours. Patients with left-sided CRC had 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 are 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

Kyoto Encyclopedia of Genes and Genomes: CIMP status and EREG level in colorectal cancer

| 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*          |                       |         |
| 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.

*Adjusted for the number of prior regimens, age, prior bevacizumab, and use of anti-EGFR as monotherapy or in combination with chemotherapy.
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 AREG and EREG, 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 AREG and EREG 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

This research was supported by American Society of Clinical Oncology Conquer Cancer Foundation Young Investigator Awards individually granted to MSL and VKM; NIH R01 CA160736 (to VB); NCI Cancer Center Support Grant (CCSG) P30 CA016672 (to VB, WW, and SK); NIH R01 CA172670 (to SK), and NIH R01 CA184843 (to SK).

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.

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

Baker JB, Dutta D, Watson 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.

Baker JB, Dutta D, Watson D, Maddala T, Munneke BM, Shak S, Bowinsky UK, 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 Bauder F, Schuch G, Zuel S, 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, Bokemeyer C, Bondarenko I, Hartmann JT, De Braud F, Schuch G, Zubel A, 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. *BM Bioinformatics* 11: 587.

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

Jacobi B, De Roeck W, Piessevaux H, Van Oirbeck R, Bisman S, De Schutter J, Fieuws S, Van den Oende P, Peeters M, Van Laethem JL, Humblet Y, Van Cutsem E, De Hertogh G, Dertogh H, Laurent-Puig P, Van Cutsem E, Teijar 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, genes and bodies and beyond. *Nat Rev Genet* 13: 484–496.

Joseph 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 DJ, 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, Harq 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 NZCIC CTG CO.17. *Ann Oncol* 25: 744–753.

Kharbati-Ford S, 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, Baselag J, Clark EA, Mauro DJ (2007) Expression of epiregulin and amphiregulin mRNA expression in primary tumors predicts outcome in metastatic colorectal cancer treated with cetuximab. *J Clin Oncol* 27: 5068–5074.

Laurent-Puig P, Cayre A, Manceau G, Buc E, Bachel J, Lecomte T, Imanoglu C, Rougier P, Lievre A, Landi B, Boigue V, Ducruex M, Ychou M, Bihlau F, Bouche O, Reid J, Stone S, Van Cutsem E, Lakka A (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, Scartozi M, Santini D, Masia G, Graziano F, Cremolini C, Rulli E, Canestri E, Funel N, Schiavon G, Petrini I, Magnani M, Tonini G, Campani D, Floriani I, Cescini S, Falcone A (2009) PTEN expression and KRAS mutations in 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, Seligmann JF, Elliott F, Richman SD, Jacobs B, Hemmings G, Brown S, Missiaglia E, Jacobs B, D’ario G, Di Narzo AF, Soneson C, Budinska E, Pentheroudakis G, Kotoula V, De Roock W, Kouvatseas G, Papakostas P, Stahler A, Heinemann V, Giessen-Jung C, Crispin A, Schalhorn A, Stintzing S, 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.

ToI, Dijkstra JR, Klomp M, Teerenstra S, Dommerholt M, Vink-Borger MG, 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.

Tusnad 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, Reuss F, Stroh C, Teijpar 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.

Van Eenom JC, Heinemann V, Van Weelkeshal LF, Vehling-Kaiser U, Stausch M, Hass HG, Deckter T, Klein S, Held S, Jung A, Kirchner T, Haas M, Holch J, Michl M, Aubele P, Bock W, 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 KRR-0104 trial. *J Clin Oncol* 40: 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, Kampan M, Young J, Long TI, Faaske 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, Huttonower 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 D, 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.

This work is published under the standard license to publish agreement. After 12 months the work will become freely available and the license terms will switch to a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 Unported License.