



## ORIGINAL ARTICLE

# A combined oncogenic pathway signature of *BRAF*, *KRAS* and *PIK3CA* mutation improves colorectal cancer classification and cetuximab treatment prediction

Sun Tian,<sup>1</sup> Iris Simon,<sup>1</sup> Victor Moreno,<sup>2,3</sup> Paul Roepman,<sup>1</sup> Josep Tabernero,<sup>4</sup> Mireille Snel,<sup>1</sup> Laura van't Veer,<sup>1</sup> Ramon Salazar,<sup>2</sup> Rene Bernards,<sup>1,5</sup> Gabriel Capella<sup>2</sup>

► Additional tables and figure are published online only. To view these files please visit the journal online (<http://dx.doi.org/10.1136/gutjnl-2012-302423>).

<sup>1</sup>Research and Development, Agendia BV, Amsterdam, The Netherlands

<sup>2</sup>IDIBELL, Institut Català d'Oncologia, L'Hospitalet de Llobregat, Spain

<sup>3</sup>Medical Oncology Department, Vall d'Hebron University Hospital, Universitat Autònoma de Barcelona, Barcelona, Spain

<sup>4</sup>Department of Clinical Sciences, Faculty of Medicine, University of Barcelona, Barcelona, Spain

<sup>5</sup>Division of Molecular Carcinogenesis, The Netherlands Cancer Institute, Amsterdam, The Netherlands

## Correspondence to

Gabriel Capella, Translational Research Laboratory, Catalan Institute of Oncology-IDIBELL, Barcelona 08908, Spain; [gcapella@iconcologia.net](mailto:gcapella@iconcologia.net)

ST and IS contributed equally to this work.

Revised 21 May 2012

Accepted 13 June 2012

Published Online First

14 July 2012

## ABSTRACT

**Objective** To develop gene expression profiles that characterise *KRAS*-, *BRAF*- or *PIK3CA*-activated-tumours, and to explore whether these profiles might be helpful in predicting the response to the epidermal growth factor receptor (EGFR) pathway inhibitors better than mutation status alone.

**Design** Fresh frozen tumour samples from 381 colorectal cancer (CRC) patients were collected and mutations in *KRAS*, *BRAF* and *PIK3CA* were assessed. Using microarray data, three individual oncogenic and a combined model were developed and validated in an independent set of 80 CRC patients, and in a dataset from metastatic CRC patients treated with cetuximab.

**Results** 175 tumours (45.9%) harboured oncogenic mutations in *KRAS* (30.2%), *BRAF* (11.0%) and *PIK3CA* (11.5%). Activating mutation signatures for *KRAS* (75 genes), for *BRAF* (58 genes,) and for *PIK3CA* (49 genes) were developed. The development of a combined oncogenic pathway signature-classified tumours as 'activated oncogenic', or as 'wildtype-like' with a sensitivity of 90.3% and a specificity of 61.7%. The identified signature revealed other mechanisms that can activate ERK/MAPK pathway in *KRAS*, *BRAF* and *PIK3CA* wildtype patients. The combined signature is associated with response to cetuximab treatment in patients with metastatic CRC (HR 2.51,  $p < 0.0009$ ).

**Conclusion** A combined oncogenic pathway signature allows the identification of patients with an active EGFR-signalling pathway that could benefit from downstream pathway inhibition.

## INTRODUCTION

The epidermal growth factor receptor (EGFR) is a member of the ERBB family of receptors that plays a key role in cell proliferation, adhesion and migration.<sup>1–3</sup> The EGFR downstream intracellular signal transduction pathways include components of the RAS/mitogen-activated protein kinase (MAPK), phosphatidylinositol 3-kinase (PI3K), signal transducer and activator of transcription, protein kinase C and phospholipase D pathways.<sup>4</sup>

Antibodies aimed directly at EGFR, such as cetuximab<sup>5</sup> and panitumumab,<sup>6</sup> have shown clinical benefit in metastatic colorectal cancer (CRC). However, only 20%–30% of these patients respond to EGFR inhibitors. The lack of clinical response to

## Significance of this study

### What is already known about this subject?

- The EGFR pathway became the focus of therapy development over the last few years, as the activated pathway drives cancer progression.
- Anti-EGFR inhibitors, like cetuximab and panitumumab, have shown clinical benefit in metastatic colorectal cancer.
- Activating *KRAS* mutations are established as negative predictor for response.
- Activating *KRAS* mutations, or mutations in other key molecules (*BRAF*, *PIK3CA*), cannot fully explain the extent of non-response in the patient population and, therefore, other causes for non-response need to be identified.

### What are the new findings?

- Combined transcriptional profiling of oncogenic mutations allows the identification of patients with an active EGFR-signalling pathway that could benefit from downstream pathway inhibition.
- Patients with a *BRAF* mutation display a distinct expression pattern.
- Mechanisms other than oncogenic mutations can cause a similar activation of the pathway and result in a similar transcriptional pattern.
- The development of activated pathway signature, as described here, allows the identification of all patients who have a similar phenotype as patients with oncogenic mutations. The signature is, therefore, more comprehensive and predictive than the mutation status alone.

the EGFR pathway inhibitors in many patients has been partially attributed to downstream or parallel activation of the pathway.<sup>7</sup> Notably, *KRAS* and *PIK3CA* pathways can also be activated by other upstream receptors.

Molecular alterations that constitutively activate EGFR signal transduction have been associated with lack of response to cetuximab or panitumumab. *KRAS* mutations account for only 30%–40% of non-responders to EGFR targeting in colorectal cancer.<sup>8–13</sup> Patients with activating mutations in



Open Access  
Scan to access more  
free content

## Significance of this study

## How might it impact on clinical practice in the foreseeable future?

- ▶ A better understanding of the underlying mechanism of response to anti-EGFR therapies will help to further personalise medicine and increase benefit.
- ▶ Our findings, and other published reports, demonstrate that expression signatures measuring pathway activation can identify patients who are sensitive to a pathway inhibition, and these signatures seem superior to measuring the mutation status alone. This observation should be confirmed in additional clinical studies.
- ▶ The development and use of such signatures might be of special interest when less well-characterised pathways are targeted, and knowledge about predictive markers is limited.

*BRAF*,<sup>14</sup> *PIK3CA*<sup>15 16</sup> and *NRAS*<sup>17</sup> might also not respond to these antibodies. Finally, loss of PTEN protein expression, a phosphatase that inactivates the PI3KCA downstream signalling, may associate with the lack of response to cetuximab.<sup>18</sup>

DNA microarray-based gene expression signatures can define cancer subtypes, recurrence of disease and response to specific therapies using.<sup>19 20</sup> Also, they have proved useful for the analysis of the activation of oncogenic pathways in cancer samples.<sup>21 22</sup> It has recently been demonstrated that an *RAS* pathway signature is superior to *KRAS* mutation status for the prediction of dependence on *RAS* signaling, and can predict response to PI3KCA and *RAS* pathway inhibitors.<sup>23</sup>

We hypothesised that analysing independent gene expression profiles of diverse oncogenic mutations in *BRAF*, *KRAS* or *PIK3CA* may uncover signatures of activated EGFR pathway signalling. In this study, we analysed the gene expression pattern of a large number of patients, and built a model for identifying patients with activated EGFR-signalling pathways. Since detection of signalling deregulation can be linked to sensitivity to targeted therapies,<sup>21</sup> we posit that such profiles may be helpful in predicting the response of individual patients to EGFR pathway inhibitors.

## MATERIAL AND METHODS

## Patients

For the training set, 381 fresh frozen tumour samples from patients with CRC were collected at four different hospitals (Institut Català d'Oncologia, Leiden University Medical Center, Netherlands Cancer Institute, Slotervaart General Hospital). Most patients had stage II (n=205) and stage III (n=116) CRC; 51 patients had stage I and 8 patients stage IV cancer. Main characteristics of the patients are depicted in table 1 and have also been described in reference<sup>24</sup> The validation study was performed on 80 tumour samples, 50 stage II and 30 stage III with similar patient characteristics as the training set (table 1). All tissue samples were collected from patients with appropriate informed consent. The study was carried out in accordance with the ethical standards of the Helsinki Declaration and was approved by the Medical Ethical Board of the participating medical centres and hospitals.

## Mutational analysis

Mutations in *BRAF* V600, *KRAS* codons 12, 13 and 61, and *PIK3CA* exons 9 and 20 were assessed in cDNA by Sanger

Table 1 Clinico-pathological information and mutation status

	Training set, all samples n = 381 n (%)	Validation set, all samples n = 80 n (%)
Gender		
M	209 (54.9)	40 (50.0)
F	172 (45.1)	40 (50.0)
Age		
Median	68 yrs	75 yrs
Localisation		
Left	202 (53.0)	31 (38.8)
Right	136 (35.7)	30 (37.5)
Rectum	37 (9.7)	8 (10.0)
Unknown	6 (1.6)	11 (13.8)
T-stage		
1	4 (1.0)	0 (0)
2	53 (13.9)	2 (2.5)
3	296 (77.7)	74 (92.5)
4	27 (7.1)	4 (5.0)
Unknown	1 (0.3)	0 (0)
N-status		
N0	259 (68.0)	50 (62.5)
N+	121 (31.8)	30 (37.5)
Unknown	1 (0.3)	0 (0)
Distal metastasis in follow-up		
No	299 (78.5)	68 (85.0)
Yes	81 (21.3)	12 (15.0)
Unknown	1 (0.3)	0 (0)
Stage		
1	51 (13.4)	0 (0.0)
2	205 (53.8)	50 (62.5)
3	116 (30.4)	30 (37.5)
4	8 (2.1)	0 (0.0)
Unknown	1 (0.3)	0 (0)
Grade		
Good	101 (26.5)	8 (10.0)
Moderate	233 (61.2)	52 (65.0)
Poor	45 (11.8)	16 (20.0)
Unknown	2 (0.5)	4 (5.0)
<i>KRAS</i> mutation		
No	266 (69.8)	51 (63.7)
Yes	115 (30.2)	29 (36.3)
<i>BRAF</i> mutation		
No	339 (89.0)	76 (95)
Yes	42 (11.0)	4 (5.0)
<i>PIK3CA</i> mutation		
No	337 (88.5)	64 (80.0)
Yes	44 (11.5)	16 (20.0)
Any mutation		
No	206 (54.1)	40 (50.0)
Yes	175 (45.9)	40 (50.0)

sequencing of PCR products using primers with M13 tails after RT-PCR. (ServiceXS BV). V600E *BRAF* mutation were analysed after amplification of exon 15, using primers 5'-TGATCAAACCTATAGATATTGCACGA (upstream) and 5'-TCATACAGAA-CAATTCCAAATGC (downstream). *KRAS* whole coding region was analysed using primers 5'-AGGCCTGCTGAAAATGACTG (upstream) and 5'-TGGTGAATATCTTCAAATGATTTAGT (downstream). For *PIK3CA* the primers used were 5'-CCACG-CAGGACTGAGTAACA (upstream) and 5'-GGCCAATCTTT-TACCCAAGCA (downstream) for exon 9, and 5'-TGAGCAAGA-GGCTTTGGAGT (upstream) and 5'-AGTGTGGAATCCA-GAGTGAGC (downstream) for exon 20. The Mutation Surveyor Software (SoftGenetics LLC) was used for sequence analysis.

### Gene expression profiling and signature development

RNA isolation, amplification, labelling, the hybridisation to Agilent full-genome microarrays and data processing was performed as previously described.<sup>24</sup> Since all samples contained at least 40% tumour cells, we assumed that they were informative for tumour cell status. The reference comprised a pool of 44 CRC specimens and was processed and labelled in the same manner as the test samples.

Normalised gene expression ratios for every hybridisation were combined to produce a single gene expression profile per patient, using Matlab software (MathWorks, Inc, Natick, Massachusetts, USA). Development of each of the three oncogenic signatures was performed according to a 10-fold cross-validation strategy (CV10).<sup>24</sup> For each of the CV10 iteration, 10% of samples with an activating mutation and 10% of the wildtype samples were randomly excluded from the gene selection procedure and used for testing of the identified gene signatures. Using the remaining 90% of samples, each gene was scored for its association with the mutation status in the training procedure (table 1). Only genes with a p value <0.05 (two-sided Student's t test) and a differential expression >1.25 were ranked-order based on their p value, and used for subsequent signature construction. For increasing gene set (2, 3...500) a nearest centroid classification model was constructed and tested on the initially excluded 10% of samples.<sup>25</sup> Signature performance and individual gene p values were recorded, and the complete CV10 procedure was repeated a thousand times. The final optimal set of genes was determined by combing all CV10 results and selection of the minimal number of genes with the higher overall accuracy (combined sensitivity and specificity).

The selected set of optimal gene probes was used to score all tumour samples for their correlation with the *KRAS*, the *BRAF* and the *PI3KCA* signature (Suppl Table S1). Depending on whether the signature score exceeded a predefined threshold, a sample was classified as activated oncogenic or as wildtype-like. The signature thresholds were based on an optimal sensitivity to correctly identify all known oncogenic mutation carriers and all other samples that harboured the activating oncogenic signatures. For a comprehensive identification of tumours that harbour activating oncogenic mutation(s) in the EGFR-signalling pathway, the three individual *KRAS*, *BRAF* and *PI3KCA* oncogenic signatures were combined. In this combined model, the gene expression pattern of each tumour is scored for its association with all the three oncogenic mutation signatures and placed into a 3D plot. All signature-related gene expression data are available on the Agendia Research web site (<https://research.agendia.com>).

### Signature validation in cetuximab-treated patients

Publicly available Affymetrix gene expression and clinical data from 80 samples from metastatic colon cancer patients who have been treated with cetuximab<sup>12</sup> were downloaded via the Gene Expression Omnibus with accession number GSE5851. Using all probes, the unsupervised K-means clustering with cityblock distance measurement can partition these 80 tumours into two groups: one group with 33% of samples and the other group with 67% of them. Clinical data regarding cetuximab treatment response were available for 68 patients. We could identify 49 of the 75 *KRAS* signature genes, 37 of the 59 *BRAF* signature genes and 34 of the 49 *PI3KCA* signature genes. The genes found on the Affymetrix arrays are highlighted in the Suppl Table S1. Readout of the signature scores on the Affymetrix data by the nearest centroid classifier methods was done in a similar fashion as used for the in-house Agilent microarray

data. Samples were ranked according to the signatures scores and classified as activating oncogenic (n=22) and wildtype-like (n=46) based on the 33-percentile.

### Additional analysis

Ingenuity Pathways Analysis (IPA) software (Ingenuity Systems, Inc, Redwood City, California, USA), was used for gene ontology analysis and to analyse the profile genes for their potential-enriched biological and/or molecular function and signalling pathway involvement. A comparison was made between tumours harbouring no oncogenic mutations in *BRAF*, *KRAS* and *PI3KCA*, but classified as *activating oncogenic* (n=79) and tumours harbouring no oncogenic mutations in *BRAF*, *KRAS* and *PI3KCA* that were classified as *wildtype-like* (n=127). The false discovery rates-adjusted p values by Benjamini and Hochberg linear step-up procedure were calculated using a Student's t test. Differentially expressed genes (Benjamini and Hochberg-adjusted p value <0.05) were input into IPA. To eliminate false protein-protein interactions in the pathway, only experimentally observed direct inhibitions and direct activations were considered. The area under Receiver Operating Characteristics (ROC) curve analysis was performed using MedCalc for Windows, version 9.6.4.0 (MedCalc Software, Mariakerke, Belgium).

## RESULTS

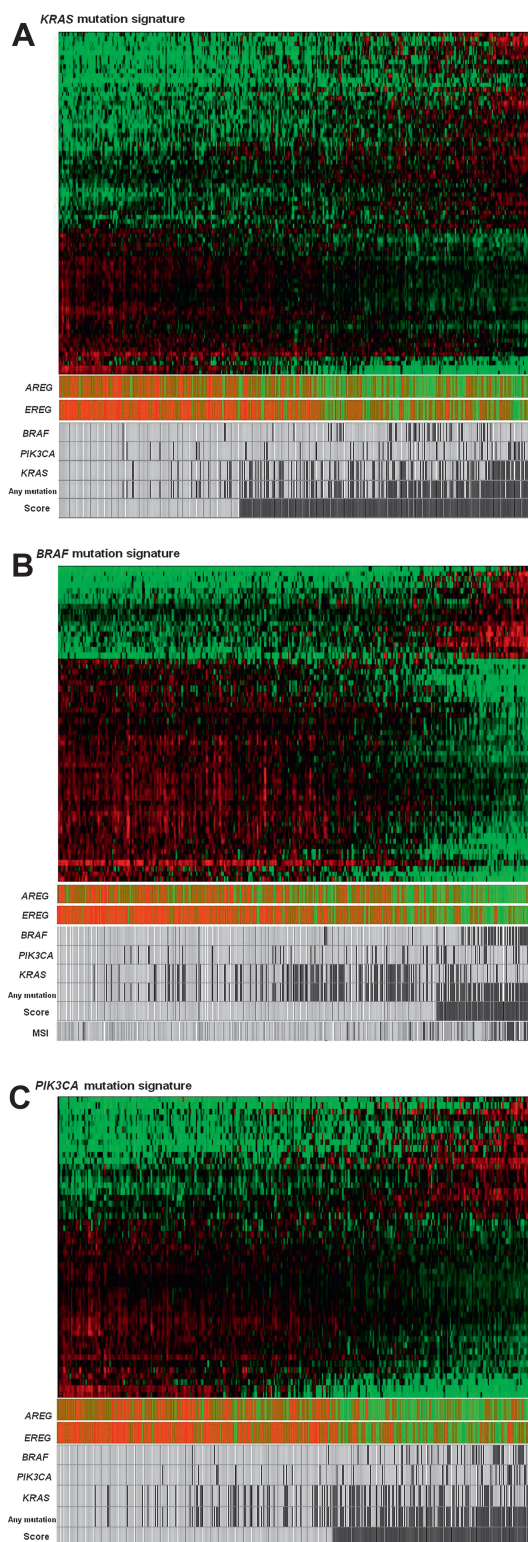
### Mutational analysis

Three hundred and eighty-one primary colorectal tumours for which we have recently reported full genome gene expression analysis were analysed.<sup>24</sup> *KRAS* mutations were detected in 115 out of 381 cases (30.2%), *PI3KCA* mutations in 44 cases (11.5%) and *BRAF* mutations in 42 cases (11%) (table 1). Altogether, 175 tumours (45.9%) harboured at least one oncogenic mutation. The majority of *KRAS* mutations was found at codon 12 (22%), while 6.8% had mutation at codon 13, and 1.9% of patients harboured a mutation at codon 61. For *PI3KCA*, mutations were analysed in two previously reported 'hotspot' regions in exons 9 and 20, corresponding to the accessory (helical) and catalytic domains of *PI3KCA*, respectively. Mutations E542K/L, E545K/G, Q546K and D549E in exon 9 and mutations M1043I, A1046A and H1047L/R/Y in exon 20 were considered activating.<sup>26</sup> The majority of patients had mutations in exon 9 (7.1%) and only 4.7% in exon 20. One sample showed mutations in both exons. A limited overlap of mutations in different genes was observed. *BRAF* and *PI3KCA* mutations coexisted in nine cases (2.4%) and *KRAS* and *PI3KCA* mutations coexisted in 16 (4.2%) cases.

### Development and validation of mutation signatures

*KRAS*, *BRAF* and *PI3KCA* mutation status was used to capture their gene-specific expression patterns. Using a 10-fold cross-validation procedure, three sets of genes were identified (Suppl Table S1) that were optimally suited for constructing an activating oncogenic signature for *KRAS* (75 genes, figure 1A), for *BRAF* (58 genes, figure 1B) and for *PI3KCA* (49 genes, figure 1C). None of the signatures correlated with disease relapse after a median follow-up time of 59.9 months (p values 0.572, 0.383, 0.441 for *KRAS*, *BRAF* and *PI3KCA*-signatures, respectively) indicating that the signatures are not prognostic. The *KRAS* oncogenic signature correctly classified 105 of the 115 tumours with a *KRAS* mutation (sensitivity 91.3%), and showed a specificity of 62% for an overall accuracy of 72.3% (table 2). No differences were evident between tumours harbouring codon 12 *KRAS* mutations or those harbouring codon 13 mutations (data not shown). The *BRAF* oncogenic signature correctly classified





**Figure 1** *KRAS*, *BRAF* and *PIK3CA*-activating oncogenic gene signatures. Gene expression signatures specific for activating oncogenic mutations in (A) *KRAS* (75-gene signature), (B) *BRAF* (58-gene signature) or (C) *PIK3CA* (49-gene signature) genes. The heatmap represents relative gene expression levels of the signature genes across 381 colon tumour samples. Tumours are sorted according to the signature outcome (score). High gene expression is coloured in red, low expression in green. *Score*: tumour classified as activating oncogenic by the profile score are displayed as black boxes, tumours that are classified as wildtype-like are displayed as white boxes. *KRAS*, *BRAF* and *PIK3CA*: presence of the respective oncogenic mutations as

38 of the 42 *BRAF*(+) tumour samples (sensitivity 90%), and classified 189 of the 206 tumours with no mutation as non-mutation-like (specificity 92%) resulting in an overall accuracy of 91% (table 2). *BRAF* mutations have been reported to be more frequent in patients with microsatellite instability (MSI-H).<sup>27</sup> Indeed, a correlation (Matthews correlation MCC=0.48) was observed between *BRAF* signature and the presence of MSI-H status. MSI status was known for 269 patients of whom 29 were classified as MSI-H (28). Fourteen (14/29=48.3%) of the MSI-H patients had a *BRAF* mutation and 20 (20/29=69%) were classified as *BRAF*-oncogenic by the signature (figure 1B). The *PIK3CA* signature showed a sensitivity of 75% with a specificity of 80% and an overall accuracy of 79% (table 2). Again, no evident differences were observed between exon 9 and exon 20 mutations (data not shown).

The three signatures were validated in an independent dataset of 80 tumour samples (table 2). In this validation set, the performance of the signatures was in very good agreement with results from the training set.

The heatmap of the three tumour classification signatures (figure 1) suggested that the oncogenic signatures of the different mutations might have some degree of overlap. The proportion of overlapping genes in the minimal optimal gene sets (Suppl Tables S1, S2 and S3) was higher than chance (p value<0.00001; hypergeometric statistics). Twenty-one overlapping genes for the *KRAS*+*PIK3CA* signatures represent 28% of the components of the 75 *KRAS* gene signature and 43% of *PIK3CA* 49-gene signature are shared. The 11-gene overlap between *BRAF* and *PIK3CA* represents 22% of *PIK3CA* 49-gene signature, and 19% of the *BRAF* 58-gene signature. The overlap for *KRAS* + *BRAF* is lower: six genes (8% of *KRAS* 75 gene, 10% of *BRAF* 58 gene) (Table Suppl S3).

Values of area under the ROC calculated from all 381 samples for all mutations displayed highly significant values when compared with all mutations (p=0.0001; binomial exact test) suggesting that any mutation signature is able, to partially discriminate other types of mutations (supplementary table S4). Concordance analysis using  $\kappa$  statistics point to a modest overlapping of *BRAF* signature with *PIK3CA* mutations (AUC *PIK3CA* 0.65,  $\kappa$ =0.141) and of *KRAS* signature with *BRAF* mutations (AUC *BRAF* 0.76,  $\kappa$ =0.124).

### Combined oncogenic pathway signature

When using a multidimensional classification model (or combined oncogenic pathway signature), tumours displaying a gene expression pattern similar to any of the three oncogenic signatures were classified as 'activating oncogenic', while tumours with a gene expression pattern that is negatively associated with all three oncogenic signatures are classified as wildtype-like and showed a *KRAS/BRAF/PIK3CA* wildtype profile (figure 2). The combined oncogenic pathway signature correctly predicted the presence of a mutation in *KRAS*, *BRAF* and/or *PIK3CA* for 158 of the 175 known mutation carriers (sensitivity of 90.3%) in the training set, and 37 of 40 known mutation carriers (sensitivity of 92.5%) in the validation set.

Interestingly, 79 of the 206 tumours that had no oncogenic mutations in *BRAF*, *KRAS* and *PIK3CA* were also classified as

[Continued] measured by sequence analysis is indicated by corresponding black boxes. Any mutation: tumours carrying any of *KRAS/BRAF/PIK3CA* mutation are indicated by black boxes. *AREG*, *EREG*: expression levels of *AREG* and *EREG*. *MSI*, tumours with known MSI-H status.



**Table 2** Performance of *KRAS*, *BRAF* and *PIK3CA* mutation signatures in training and validation set

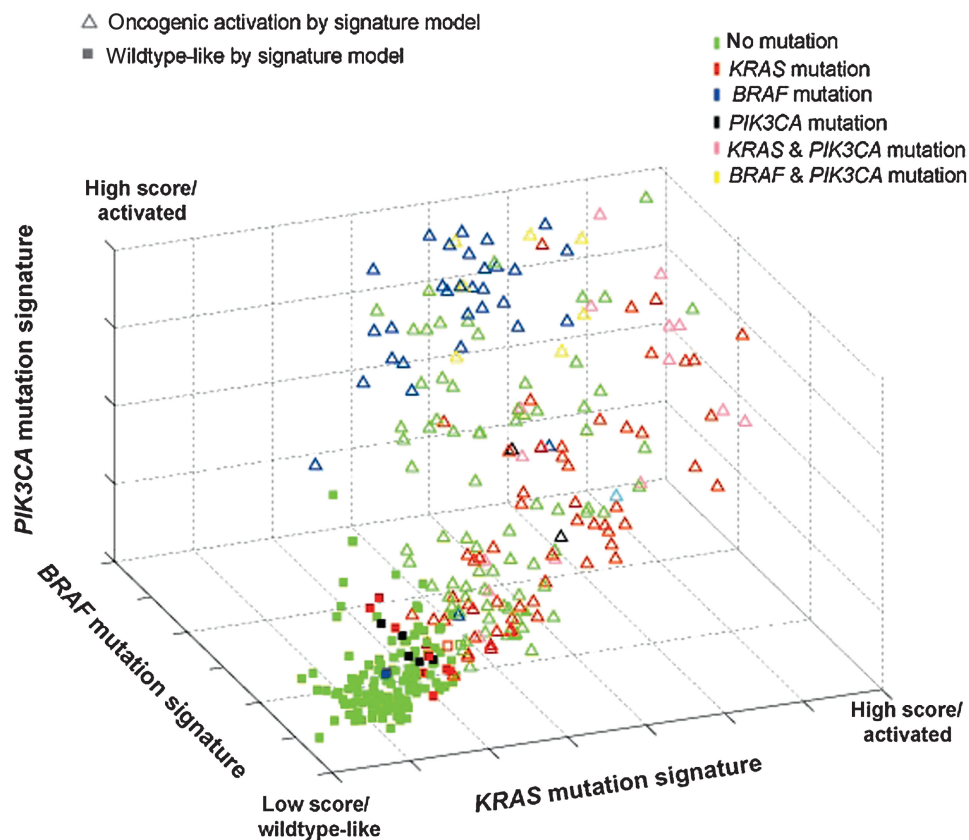
	<i>KRAS</i> signature	<i>BRAF</i> signature	<i>PIK3CA</i> signature
Training set (n=381)			
Sensitivity (%)	91	90	75
Specificity (%)	62	92	80
Accuracy (%)	72	91	79
Area under the ROC curve*	0.846	0.948	0.853
95% CI	(0.801 to 0.883)	(0.912 to 0.972)	(0.803 to 0.894)
z statistic	16.338	23.254	11.868
Significance level P (area=0.5)	<0.0001	<0.0001	<0.0001
Validation set (n=80)			
Sensitivity (%)	93	100	69
Specificity (%)	65	93	78
Accuracy (%)	77	93	75
Area under the ROC curve (AUC)	0.872	0.981	0.812
95% CI	(0.770 to 0.941)	(0.886 to 1.000)	(0.686 to 0.904)
z statistic	8.579	24.350	4.818
Significance level P (area=0.5)	<0.0001	<0.0001	<0.0001

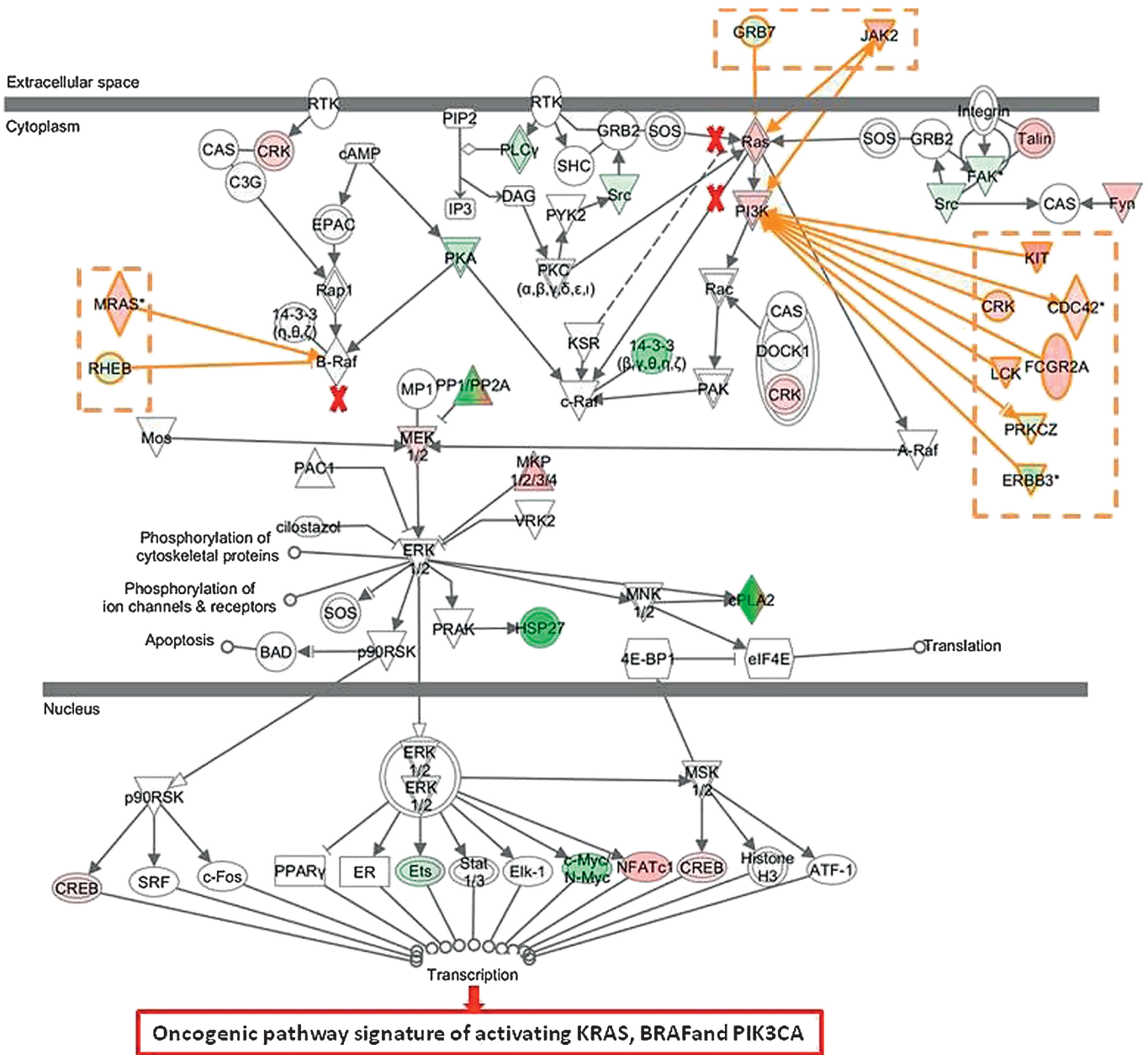
\*Calculated from training samples used for each signature (AUC).

oncogenic based on their gene expression signatures (specificity of 61.7%) indicating that many *KRAS*, *BRAF* and *PIK3CA* wildtype patients share the same phenotype of an activated EGFR pathway as the patients with at least one activating mutation. To better understand why patients with no activating mutation in *KRAS*, *BRAF* and *PIK3CA* (n=79) were classified as *activating oncogenic*, we compared their gene expression pattern with the pattern of those patients who also had no detected oncogenic mutations in *KRAS*, *BRAF* and *PIK3CA*, but were classified as *wildtype-like* (n=127). IPA revealed that genes belonging to the ERK/MAPK pathway were significantly enriched in the combined signature (p=0.027). A more detailed

analysis showed that there may be up to three alternate activating mechanisms in the set of wildtype patients having an oncogenic phenotype (figure 3). First, activation can occur directly through overexpression of key molecules in the pathway; indeed overexpression of *RAS*, *PI3K* and *MEK 1/2* is observed (see red nodes in figure 3). Second, known activators of the ERK/MAPK pathway, such as *KIT*, *CDC42*, *MRAS*, *CRK*, *FCGR2A*, *JAK2* and *LCK* are frequently overexpressed (activator genes coloured in red). Third, downregulation of proteins that are known to directly or indirectly inhibit molecules of the pathway (ie, RHEB) is observed as well (inhibitor genes coloured in green).

**Figure 2** Combined oncogenic pathway signature. Signature model for classification of 381 colon tumour samples based within a 3D space based on its association with the developed gene expression oncogenic signatures for *KRAS* (x-axis), *BRAF* (y-axis) and *PIK3CA* (z-axis). Each point represents a single tumour sample and is coloured according to its mutation status as measured by sequence analysis. The shape of the point represents that classification outcome based on the developed model in which triangles indicate tumours that are classified as mutation-like for the EGFR pathway by either *KRAS*, *BRAF* and/or *PIK3CA*; squares indicate tumours that are classified as non-mutation-like.





**Figure 3** Ingenuity Pathway Analysis of differentially expressed genes of patients with no activating mutation in KRAS, BRAF and PIK3CA classified as oncogenic by the combined model (n=79) versus patients classified as wildtype-like (n=127) The ERK/MEK pathway can be activated by mutations in KRAS, BRAF or PIK3CA (marked with red crosses). In tumours harbouring no activating mutation, the ERK/MEK pathway can be activated by (1) upregulation of RAS, PI3K and mitogen-activated protein kinase family genes (red nodes), (2) upregulation of activator genes (coloured in red) and (3) downregulation of inhibitor genes (coloured in green).

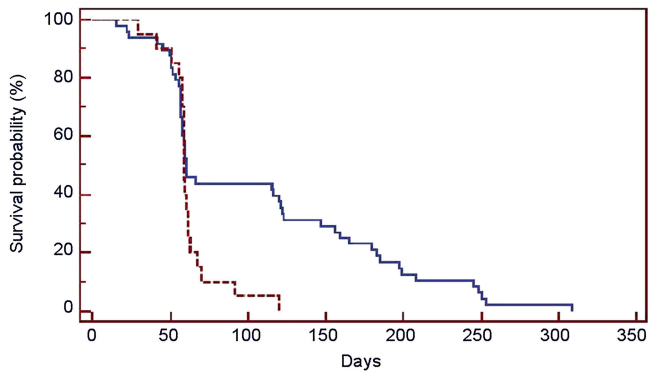
Interestingly, tumours with high expression levels of amphiregulin (AREG) and epiregulin (EREG), two ligands that can activate EGFR and have shown clinical relevance for the treatment of CRC patients,<sup>12</sup> clustered in the groups of patients with wildtype-like signatures (figure 1) showing an inverse correlation with the scores of the three KRAS, BRAF and PIK3CA oncogenic signatures (AREG: Pearson correlations -0.39, -0.42 and -0.47; EREG: -0.51, -0.53 and -0.58, respectively).

**Association of the combined oncogenic pathway signature with response to cetuximab**

The high proportion of mutation-negative cases that were displaying an oncogenic signature gave rise to the possibility that the signature could have some predictive value in the

response to cetuximab beyond the mutation status. We explored whether the signatures have predictive value in the response to cetuximab (Erbix) treatment in patients with metastatic CRC using a publicly available dataset of 80 metastatic colon cancer patients.<sup>12</sup> Of these, information available was useful in 68 cases, 20 of them KRAS mutation positive. No patient had a BRAF mutation, and the PIK3CA mutation status was unknown. Tumour samples were analysed in silico for the KRAS, BRAF and PIK3CA signatures, and the combined signature was used to assess the association with response. As shown in the original publication,<sup>17</sup> KRAS mutation status is an indicator of response with wildtype patients having a better outcome than patients with activating mutations (figure 4A). The 68 tumours were ranked according to their correlation with the gene

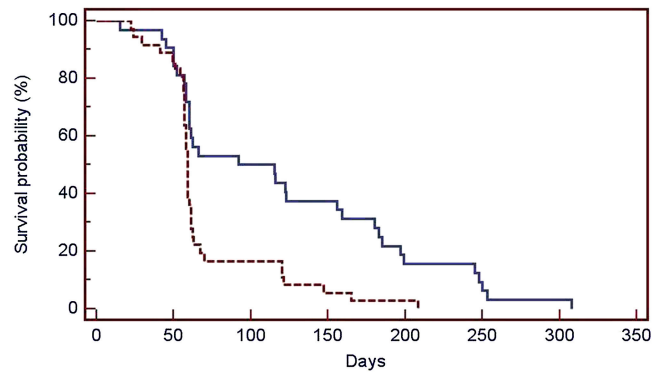
**A** *KRAS* mutations status  $p=0.0174$ , HR=2.23



Number at risk

Group: 0	48	40	21	14	6	2	1	0
Group: 1	20	18	1	0	0	0	0	0

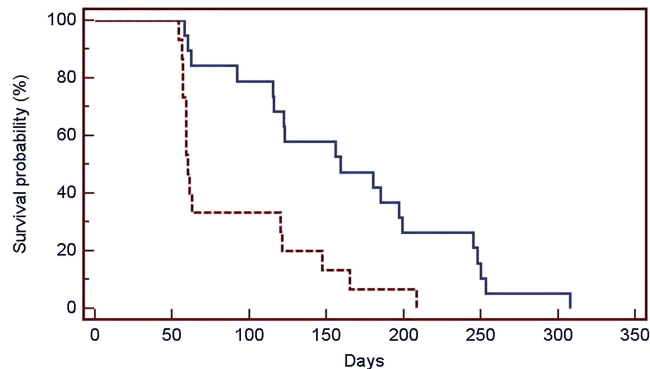
**B** Combined model  $p=0.0009$ , HR=2.51



Number at risk

Group: 0	32	27	16	12	5	2	1	1
Group: 1	36	31	6	2	1	0	0	0

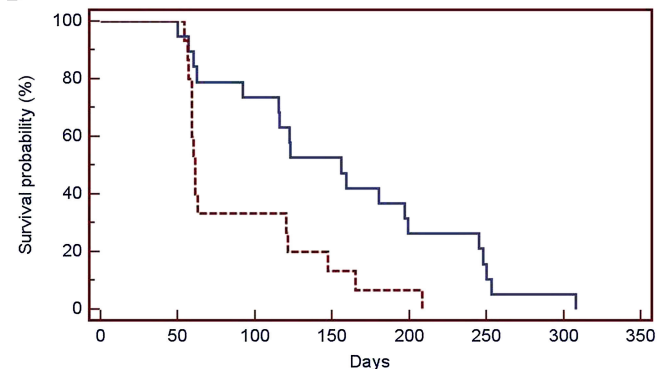
**C** Combined model on *AREG* high group  $p=0.001$ , HR=4.06



Number at risk

Group: 0	19	19	15	11	5	2	1	1
Group: 1	15	15	5	2	1	0	0	0

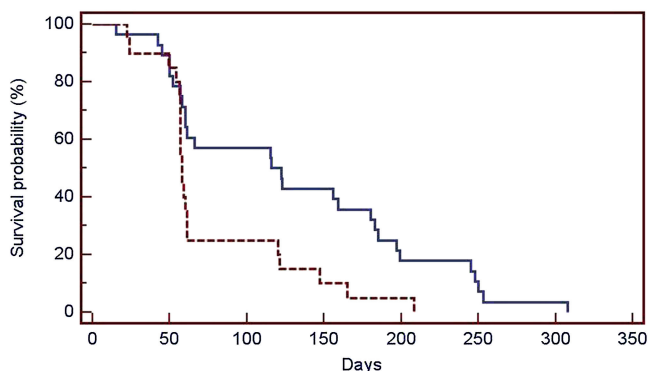
**D** Combined model on *EREG* high group  $p=0.005$ , HR=3.28



Number at risk

Group: 0	19	18	14	10	5	2	1	1
Group: 1	15	15	5	2	1	0	0	0

**E** Combined model on *KRAS* wild type group  $p=0.009$ , HR=2.49



Number at risk

Group: 0	28	23	16	12	5	2	1	1
Group: 1	20	17	5	2	1	0	0	0

**Figure 4** Classification by the combined signature is associated with response to cetuximab treatment. Kaplan–Meier survival analysis of 68 metastatic colorectal cancer patients who have received cetuximab treatment.<sup>12</sup> Tumour samples have been classified in silico as wildtype-like (Group 0, blue solid line) or as oncogenic (Group 1, red dashed line) by the (A) *KRAS* mutation status, by (B) the combined signature model. In addition, survival curves are shown for classification by the combined signature model of *AREG*-high (C) and *EREG*-high (D) samples groups, and for *KRAS* wildtype patients only (E).



signatures and classified as activating oncogenic or wildtype-like, based on the 33-percentile. Kaplan–Meier survival analysis resulted in a significant better survival of patients classified as wildtype (n=46) compared with patients classified as oncogenic (n=22) by the gene signatures (*KRAS* signature, HR=2.62 p=0.004, logrank test; *BRAF* signature, HR=2.08 p=0.018; *PIK3CA*, HR=2.67 p=0.003; Suppl figure 1). The combined oncogenic pathway signature also showed a strong predictive value (HR=2.51 p=0.0009, figure 4B).

We further evaluated whether the association of the combined signature with cetuximab response was independent of amphiregulin (*AREG*) and epiregulin (*EREG*) expression levels as these two ligands have been shown to have predictive value for cetuximab response.<sup>12</sup> For this purpose, the 68 tumours were split into an *AREG*-high (n=34) and an *AREG*-low (n=34) group based on the median *AREG* expression level. Within the *AREG*-high group, the oncogenic pathway signature remained significantly associated with survival of treated patients (HR=4.06 p=0.001, figure 4C). This analysis was repeated for *EREG* expression, and also within the *EREG*-high patient group the combined signature was able to identify patients who benefited more from cetuximab treatment (HR=3.28 p=0.005, figure 4D). In contrast, classification using the combined signature showed no association with survival within the *AREG*-low (p=0.99) and *EREG*-low groups (p=0.35). These results indicate that, although classification by the gene signature model is associated with *AREG/EREG* expression, it likely holds additional predictive value to further stratify patients that benefit most from anti-EGFR-targeted therapies.

We investigated whether the predictive power of the combined oncogenic pathway signature is independent of *KRAS* mutation status. The combined classification model showed a significant performance within *KRAS* wildtype patients (HR=2.49 p=0.009, figure 4E), indicating an association with survival beyond *KRAS* wildtype status. As expected, almost all

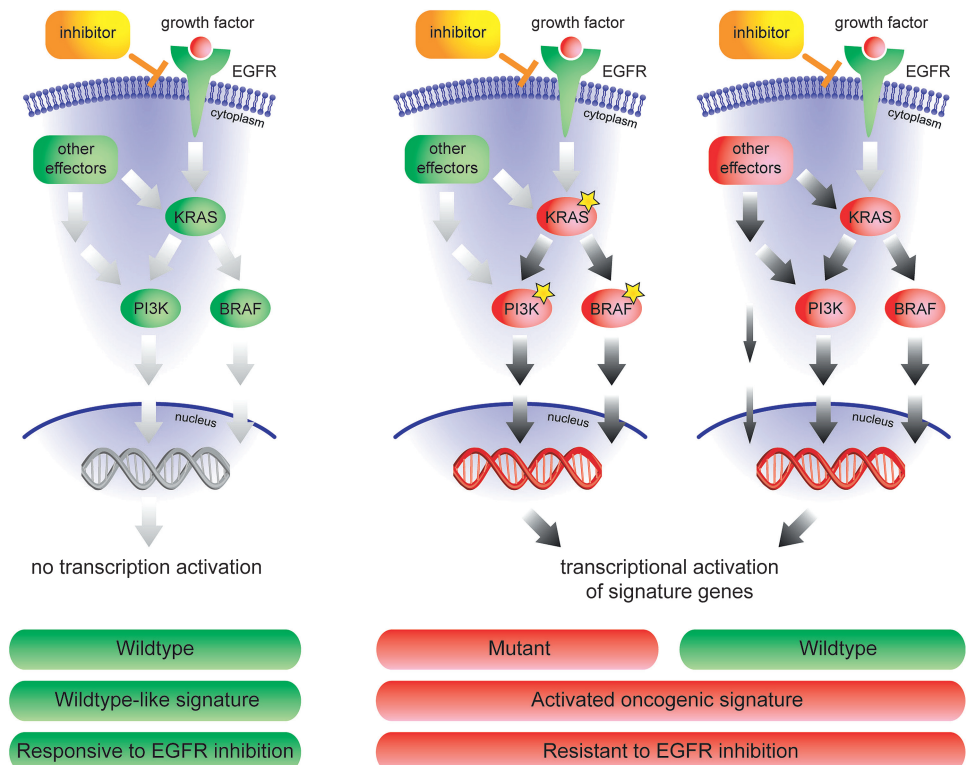
samples with a *KRAS* mutation were classified as mutation-like by the combined signature (16 of 20), and within this group the gene signature model showed no association with treatment response (p=0.63). Finally, *KRAS* status is not predictive in the 36 patients with activated oncogenic pathway (data not shown).

**DISCUSSION**

Three signalling pathway signatures were developed to characterise specific types of oncogenic mutations in colon cancer (*KRAS*, *BRAF* or *PIK3CA*). The results indicate that, in spite of widespread effects, tumours with these mutations display a partially shared gene expression pattern captured by our signature. The combined oncogenic pathway signature model has led to the classification of ‘activating oncogenic’ tumours with a significant proportion of mutation-negative cases.

The combined oncogenic pathway signature identifies tumours with one specific mutation, but also patients with an activated EGFR pathway caused by other aberrations or by unknown activation mechanisms (figure 5). A variety of mechanisms can activate the EGFR/ERK/MAPK pathway in wildtype patients: (1) directly by upregulation of RAS, PI3K and mitogen-activated protein kinase family genes; (2) indirectly by upregulation of activator genes, like KIT, CDC42, MRAS, CRK, FCGR2A, LCK and JAK2 or (3) by downregulation of inhibitor genes. These mechanisms of activation can only be detected by a signature that measures the effect of any downstream activation. Taken together, our data reveals novel underlying mechanism as to why a colorectal tumour with no known activating mutations in *KRAS*, *BRAF* or *PIK3CA* can display a shared gene expression pattern with mutation carriers, and explains resistance to the treatment of anti-EGFR drugs, such as cetuximab.

**Figure 5** Mechanism of response to anti-EGFR inhibitors. The combined oncogenic pathway signature can not only identify tumours with one specific mutation, but also patients with an activated EGFR pathway caused by other aberrations, or by unknown activation mechanisms.



In line with this observation, no particular enrichment for members of the canonical EGFR-signalling pathways is evident in the individual and combined signatures. However, a reduced number of these genes is shared, such as the members of the dual-specificity phosphatases (DUSPs) family *DUSP6* and *DUSP18*. Genes associated with the same response (ie, *DKK1*, *PHLDA1* and *SLC7A*) are also part of them. DUSP proteins shut down MAPK signalling and are involved in the late response to EGF. They play a relevant role in the negative feedback regulation of EGFR-signalling pathway attenuating the response to EGF.<sup>28</sup> *DUSP4* and *DUSP6* were originally identified as top resistance markers to cetuximab in unselected patients<sup>12, 29</sup> and have also been highlighted in in vitro systems.<sup>30</sup> The use of a four-gene model, including *AREG*, *EREG*, *DUSP6* and *SLC26A3*, has been shown to improve the identification of responders among preselected *KRAS* wildtype patients.<sup>29</sup> As there is a clear association between cetuximab response and mutation status, it is not surprising that we find *DUSP6* and members of the SLC family also in the mutation signatures. Interestingly, *DUSP6* and *SLC7A* have also been shown to correlate to response to MEK inhibitors<sup>31</sup> and PI3K pathway inhibitors.<sup>23</sup>

Among the individual signatures, the *BRAF* oncogenic-specific signatures showed the most distinct gene expression pattern, and had the best performance with an overall accuracy of 91%. The more explicit *BRAF* signature identified two large distinctive groups and showed no ambiguous region when compared with *KRAS* (figure 1A) and *PIK3CA* profiling (figure 1C). As expected, a strong correlation was observed with MSI-H status. *BRAF* mutations have been described to be more frequent (40%–50%) in MSI-H patients.<sup>27</sup> The better accuracy of the *BRAF* signature may reflect its more downstream position within the EGFR-signalling transduction pathway. *KRAS* and *PIK3CA* are located more upstream in the EGFR-signalling cascade and may, therefore, signal to a wider range of downstream signal transduction pathways and transcriptional effects. In contrast, the further downstream-located *BRAF* may signal to a more limited number of signal transduction pathways, resulting in a more discrete gene expression pattern. In spite of its high specificity, *BRAF* signature still captures a proportion of *KRAS* mutant cases in line with the recent description of a population of *BRAF* mutated-like *KRAS* mutants.<sup>32</sup>

This combined model is associated with response to cetuximab treatment in patients with metastatic CRC with better performance than each of the three single mutation signatures and using the *KRAS* mutation status alone. This combined signature has independent value from *KRAS* mutation and *AREG* and *EREG* expression status, which are currently the major determinants of response. It can be speculated that these patients might benefit more from downstream inhibitors like MEK, AKT or PI3K inhibitors. In fact, signatures measuring pathway activation can identify patients who are sensitive to the pathway inhibition, and the signatures are superior to measuring the mutation status alone.<sup>23, 31</sup>

The finding that the signatures and the combined model are in good agreement with known markers of response to cetuximab, but offer additional information beyond those markers, confirms our hypothesis that a genomic signature might be a more comprehensive solution to identify the patients who are most likely to benefit. This might be of special interest when less well-characterised pathways are targeted, and knowledge about predictive markers is limited.

The large numbers of well-annotated early stage tumours that have been profiled have provided us with adequate power to

develop and validate the signatures. The use of Sanger sequencing with RNA as a starting material is a robust and reasonably sensitive option when analysing tumour cell-enriched samples. While we are aware that some *KRAS*-positive cases may have been misclassified, we do not think that this affects the validity of our observations. The excellent correlation between the training and validation set and the strong biological plausibility of the signature genes has confirmed the potential of using simple predictors and simple reporter-selection strategies.<sup>33</sup> The utility of the profile still needs further validation in patients treated with EGFR-pathway inhibitors. Finally, it will be of interest to know whether these profiles will be useful in the prediction of response of the novel compounds that inhibit the EGFR pathway further downstream, that is, at the level of PI3K, BRAF, MEK or mTOR.

**Acknowledgements** We would like to thank Leiden University Medical Center, The Netherlands Cancer Institute, the Slotervaart General Hospital in The Netherlands, Institut Català d'Oncologia and Hospital Vall d'Hebron in Spain for their patient samples. We also thank Miguel Angel Pujana for deep insight and help in the early stages of this work.

**Contributors** ST, IS, VM, PR, JT, MS, LvV, RS, RB and GC have read and reviewed the final paper. ST, IS, JT, RB and GC have been involved in conception and planning of the project and analysis. ST, VM, PR, MS, LvV and RS have been involved in acquiring data and performing experiments. ST, IS, PR, RB and GC have written the manuscript.

**Funding** This study was partially supported by grants SAF 09-07319 SAF 06-6084 from the Spanish Ministry of Education and Science, from Fondo de Investigación Sanitaria (FIS 07/0475), F05-01 from the Fundació Gastroenterologia Dr. Francisco Vilardell, RCEP C03/09 and RTICCC C03/10, RTICCRD06/0020/050, Acción Transversal contra el Cáncer, the Instituto de Salud Carlos III (FIS P108/1635, FIS P108/1359, PS09/01037, CB07/02/2005 CIBERESP) and European Commission grant FP7-COOP-Health-2007-B 'HiperDart'. We are especially grateful to the Asociación Española contra el Cáncer.

**Competing interests** Employment or Leadership Position at Agendia: TS, IS, PR, MS, LvV and RB. Stock Ownership Agendia: LvV and RB. Honoraria: none. Research funding: none. Expert testimony: none. Other remuneration: TS, IS and PR are inventors on a patent relevant to the presented work.

**Patient consent** Obtained.

**Ethics approval** Ethics approval was provided by the Medical Ethical Board of the participating medical centres and hospitals.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data sharing statement** All signature-related gene expression data are available on the Agendia Research web site (<https://research.agendia.com>).

**Open Access** This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/3.0/>

## REFERENCES

- Mendelsohn J, Baselga J. Epidermal growth factor receptor targeting in cancer. *Semin Oncol* 2006;**33**:369–85.
- Hynes NE, Lane HA. ERBB receptors and cancer: the complexity of targeted inhibitors. *Nat Rev Cancer* 2005;**5**:341–54.
- Oda K, Matsuoka Y, Funahashi A, et al. A comprehensive pathway map of epidermal growth factor receptor signaling. *Mol Syst Biol* 2005;**1**:2005.0010.
- Marmor MD, Skaria KB, Yarden Y. Signal transduction and oncogenesis by ErbB/HER receptors. *Int J Radiat Oncol Biol Phys* 2004;**58**:903–13.
- Cunningham D, Humblet Y, Siena S, et al. Cetuximab monotherapy and cetuximab plus irinotecan in irinotecan-refractory metastatic colorectal cancer. *N Engl J Med* 2004;**351**:337–45.
- Van Cutsem E, Peeters M, Siena S, et al. Open-label phase III trial of panitumumab plus best supportive care compared with best supportive care alone in patients with chemotherapy-refractory metastatic colorectal cancer. *J Clin Oncol* 2007;**25**:1658–64.
- Hoadley KA, Weigman VJ, Fan C, et al. EGFR associated expression profiles vary with breast tumor subtype. *BMC Genomics* 2007;**8**:258.
- Amado RG, Wolf M, Peeters M, et al. Wild-type *KRAS* is required for panitumumab efficacy in patients with metastatic colorectal cancer. *J Clin Oncol* 2008;**26**:1626–34.

9. **Benvenuti S**, Sartore-Bianchi A, Di Nicolantonio F, *et al*. Oncogenic activation of the RAS/RAF signaling pathway impairs the response of metastatic colorectal cancers to anti-epidermal growth factor receptor antibody therapies. *Cancer Res* 2007;**67**:2643–8.
10. **De Roock W**, Piessevaux H, De Schutter J, *et al*. KRAS wild-type state predicts survival and is associated to early radiological response in metastatic colorectal cancer treated with cetuximab. *Ann Oncol* 2008;**19**:508–15.
11. **Di Fiore F**, Blanchard F, Charbonnier F, *et al*. Clinical relevance of KRAS mutation detection in metastatic colorectal cancer treated by Cetuximab plus chemotherapy. *Br J Cancer* 2007;**96**:1166–9.
12. **Khambata-Ford S**, Garrett CR, Meropol NJ, *et al*. Expression of epiregulin and amphiregulin and K-ras mutation status predict disease control in metastatic colorectal cancer patients treated with cetuximab. *J Clin Oncol* 2007;**25**:3230–7.
13. **Lievre A**, Bachet JB, Le Corre D, *et al*. KRAS mutation status is predictive of response to cetuximab therapy in colorectal cancer. *Cancer Res* 2006;**66**:3992–5.
14. **Di Nicolantonio F**, Martini M, Molinari F, *et al*. Wild-type BRAF is required for response to panitumumab or cetuximab in metastatic colorectal cancer. *J Clin Oncol* 2008;**26**:5705–12.
15. **Sartore-Bianchi A**, Martini M, Molinari F, *et al*. PIK3CA mutations in colorectal cancer are associated with clinical resistance to EGFR-targeted monoclonal antibodies. *Cancer Res* 2009;**69**:1851–7.
16. **Oda K**, Okada J, Timmerman L, *et al*. PIK3CA cooperates with other phosphatidylinositol 3'-kinase pathway mutations to effect oncogenic transformation. *Cancer Res* 2008;**68**:8127–36.
17. **De Roock W**, Claes B, Bernasconi D, *et al*. 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* 2010;**11**:753–62.
18. **Fratini M**, Saletti P, Romagnani E, *et al*. PTEN loss of expression predicts cetuximab efficacy in metastatic colorectal cancer patients. *Br J Cancer* 2007;**97**:1139–45.
19. **Sweet-Cordero A**, Mukherjee S, Subramanian A, *et al*. An oncogenic KRAS2 expression signature identified by cross-species gene-expression analysis. *Nat Genet* 2005;**37**:48–55.
20. **van't Veer LJ**, Bernards R. Enabling personalized cancer medicine through analysis of gene-expression patterns. *Nature* 2008;**452**:564–70.
21. **Bild AH**, Yao G, Chang JT, *et al*. Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* 2006;**439**:353–7.
22. **Sole X**, Bonifaci N, Lopez-Bigas N, *et al*. Biological convergence of cancer signatures. *PLoS One* 2009;**4**:e4544.
23. **Loboda A**, Nebozhyn M, Klinghoffer R, *et al*. A gene expression signature of RAS pathway dependence predicts response to PI3K and RAS pathway inhibitors and expands the population of RAS pathway activated tumors. *BMC Med Genomics* 2010;**3**:26.
24. **Salazar R**, Roepman P, Capella G, *et al*. Gene expression signature to improve prognosis prediction of stage II and III colorectal cancer. *J Clin Oncol* 2011;**29**:17–24.
25. **Glas AM**, Floore A, Delahaye LJ, *et al*. Converting a breast cancer microarray signature into a high-throughput diagnostic test. *BMC Genomics* 2006;**7**:278.
26. **Ikenoue T**, Kanai F, Hikiba Y, *et al*. Functional analysis of PIK3CA gene mutations in human colorectal cancer. *Cancer Res* 2005;**65**:4562–7.
27. **Ogino S**, Nosho K, Kirkner GJ, *et al*. CpG island methylator phenotype, microsatellite instability, BRAF mutation and clinical outcome in colon cancer. *Gut* 2009;**58**:90–6.
28. **Amit I**, Citri A, Shay T, *et al*. A module of negative feedback regulators defines growth factor signaling. *Nat Genet* 2007;**39**:503–12.
29. **Baker JB**, Dutta D, Watson D, *et al*. Tumour gene expression predicts response to cetuximab in patients with KRAS wild-type metastatic colorectal cancer. *Br J Cancer* 2011;**104**:488–95.
30. **Oliveras-Ferreras C**, Vazquez-Martin A, Cufi S, *et al*. Stem cell property epithelial-to-mesenchymal transition is a core transcriptional network for predicting cetuximab (Erbixit) efficacy in KRAS wild-type tumor cells. *J Cell Biochem* 2011;**112**:10–29.
31. **Dry JR**, Pavey S, Pratilas CA, *et al*. Transcriptional pathway signatures predict MEK addiction and response to selumetinib (AZD6244). *Cancer Res* 2010;**70**:2264–73.
32. **Popovici V**, Budinska E, Tejpar S, *et al*. Identification of a Poor-prognosis BRAF-Mutant-Like population of patients with colon cancer. *J Clin Oncol* 2012;**30**:1288–95.
33. **van 't Veer LJ**, Dai H, van de Vijver MJ, *et al*. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002;**415**:530–6.