Genome-wide DNA methylation profiling of CpG islands in breast cancer identifies novel genes associated with tumorigenicity

Victoria K Hill (1), Christopher Ricketts (1), Ivan Bieche (2), Sophie Vacher (2), Dean Gentle (1), Cheryl Lewis (3), Eamonn R Maher (1), Farida Latif (1)

1. Medical and Molecular Genetics, School of Clinical and Experimental Medicine, College of Medical and Dental Sciences, University of Birmingham, Birmingham B15 2TT, UK.
2. Laboratory of Oncogenetic, Institut Curie - Hôpital René Huguenin, FNCLCC, F-92210, St-Cloud, France.
3. Department of Surgery, University of Texas Southwestern Medical Center, Dallas, Texas, USA.

Corresponding author: F LATIF, Medical and Molecular Genetics, School of Clinical and Experimental Medicine, College of Medical and Dental Sciences, University of Birmingham, Birmingham B15 2TT, UK.
Tel: (44)(0) 121 414 4452; Fax: (44)(0) 121 414 2538; e-mail: f.latif@bham.ac.uk
Abstract

Epigenetic profiling of tumor DNAs may reveal important new theranostic targets to improve prognosis and treatment of advanced cancer patients. In this study, we performed a genome-wide profile of DNA methylation patterns in sporadic breast tumors using the HumanMethylation27 BeadChips to assess relationships between DNA methylation changes and patient tumor characteristics. The arrays identified 264 hypermethylated loci/genes present in genomic CpG islands. Hierarchical clustering based on methylation levels divided the specimens into 3 distinct groups, within which certain clinical features also clustered. Statistically significant differences were determined between overall methylation levels of these clusters and estrogen and progesterone receptor status (p= 0.001), tumor relapse (p=0.035) and lymph node metastasis (p=0.042). We identified several individual methylated genes associated with clinical features, including 6 genes (RECK, SFRP2, UAP1L1, ACADL, ITR and UGT3A1) that showed statistical significance between methylation and relapse free survival. Notably, the RECK gene in this group has been associated in other cancers with poorest prognosis. Among the leading relapse-associated genes and the genes associated with estrogen/progesterone receptor (ER/PR) status, we sequenced an independent set of paired normal/tumor breast DNA samples to confirm tumor specificity of methylation. Further, we performed quantitative real-time RT-PCR to confirm reduced expression in methylated tumors. Our findings suggest utility for the DNA methylation patterns in these genes as clinically useful surrogate markers in breast cancer, as well as new molecular pathways for further investigation as therapeutic targets.
Introduction

Epigenetic silencing of tumour-related genes, due to hypermethylation has recently emerged as one of the pivotal alterations in cancer development (1-2). Hence, cancer is not only a polygenetic disease, but also a poly-epigenetic disease. There are three main clinical uses for the DNA methylation patterns associated with cancer development and progression. Firstly, DNA methylation patterns that are cancer cell specific and occur at an early stage of cancer development makes them ideal biomarkers for early cancer detection. Secondly, DNA methylation patterns can also be used for prognostic purposes such as survival and or relapse. Thirdly, unlike gene mutations, hypermethylation of gene promoters can be reversed by treatment with demethylating agents and are regarded as optimal targets for what is now known as epigenetic therapy. The clinical use of demethylating agents such as 5-aza-2’-deoxycytidine (decitabine) and 5-azacytidine (Vidaza), have had the greatest success in the field of haematological malignancies. In 2004, FDA approved the use of 5-azacytidine for the treatment of myelodysplastic syndromes (1).

It is estimated that more than one million women are diagnosed with breast cancer globally every year, and a substantial number of these women will die from the disease (3). The economic cost of treating breast cancer patients is substantial. Due to the heterogeneous nature of the disease, clinical progression and response to therapy is difficult to predict with the current prognostic factors and hence treatment regime are not as effective as they should be. In the present study we used the HumanMethylation27 BeadChips to explore associations between breast tumour methylation profiles and
important tumour characteristics such as tumour stage, metastasis, relapse, survival and hormone receptor status.

**Materials and Methods**

**Cell lines and patient DNA samples**

We used a total of nine breast cancer cell lines (HCC1806, HCC1419, HCC1937, HCC1143, HCC1395, MCF-7, T47-D, MDA-MB-231 and HTB19) during this study. All sporadic breast tumour samples were ductal carcinomas, we used one cohort of 39 tumours (tumour characteristics described in reference 4 and in figure 3) and a further independent cohort of 20 tumours with matched normal breast tissue DNA. None of the primary tumour samples were microdissected. All samples were acquired with ethical consent.

**Infinium array**

We used the Illumina Infinium HumanMethylation27 BeadChips (Illumina, San Diego, Ca, USA). The BeadChip contains 27,578 highly informative CpG sites covering over 14,000 human RefSeq genes. This allows researchers to interrogate all these sites per sample at a single nucleotide resolution. Bisulphite modification of DNA and chip processing and data analysis was carried out by following the manufacturer’s manual and was done at the Wellcome Trust Centre for Human Genetics Genomics Lab, Oxford, UK. Array results were run through BeadStudio software (Illumina, San Diego, CA, USA).

**List analysis**

Background normalised average β values were considered hypermethylated if no normal had a β-value >0.25. and if either (a) had a β-value >0.5 in more than 20% of samples, or.
(b) had a difference of at least 0.4 between sample $\beta$-value and mean normal $\beta$-value in more than 20% of samples. For subsequent list analysis using Fisher’s exact tests, samples were given a value of 0 or 1 representing unmethylated or methylated respectively as determined by the above criteria.

Clustering

Hierarchical clustering was performed in Cluster3. When clustering all samples for all probes, Euclidean distance and single linkage were used, when clustering the 291 hypermethylated loci for the 39 tumours, Euclidean distance and average linkage were used. Clustering results were viewed using the JavaTree program.

Statistics

Assessment of overall methylation levels with clinical features were analysed using ANOVA in MedCalc, significance was taken to be $p<0.05$. Association of single loci with clinical features were assessed using Fisher’s exact test, all p-values were corrected using false discovery rate (FDR) and taking significance to be $p<0.05$. Kaplan-Meier analysis was carried out in MedCalc taking into account Bonferroni correction when assessing significance, samples were considered significant if $p<0.05$. Student’s T-test was used to determine whether there was significance between quantitative expression of normals and tumours, again significance was taken as $p<0.05$.

DAVID analysis

The Database for Annotation, Visualization and Integrated Discovery (DAVID) was used to identify groups of genes within the selected list of 263 that were within groups based on related functions and KEGG pathways relevant to carcinogenesis using the functional annotation chart and pathways analysis functions within the functional annotation tool.
Methylation analysis of individual genes

Combined bisulphite restriction analysis (CoBRA) was used to assess the methylation status of genes in cell lines and a separate tumour/normal cohort. All DNA was modified using an EpiTect kit (Qiagen, Heidelberg, Germany) according to manufacturers’ instructions. All CoBRA PCR were semi-nested, except ACADL which was fully-nested due to unsatisfactory amplification using semi-nested primers. CoBRA primer sequences are shown in supplementary table S1. Primary PCRs were touchdown PCRs with a gene specific final annealing temperature. Secondary PCRs used 5μl primary product as starting material and were also touchdown PCRs with a longer cycle number (45 instead of 40). PCR products were digested with BstUI (Fermentas UK, York, UK) overnight at 37°C prior to visualisation on a 2% agarose gel.

Samples were cloned and sequenced to validate the array and confirm CoBRA results. PCR products were cloned into the pGEM-T easy vector (Promega, Madison, WI, USA) according to manufacturers’ instructions. Up to 12 colonies were picked per sample and sequenced following single-colony PCR using the following forward primer, 5’-TAATACGACTCACTATAGGG-3’ and reverse, 5’-ACACTATAGAATACTCAAGC-3’. Methylation indexes were calculated as a percentage of the number of methylated CpGs out of the total CpGs sequenced.

Cell line expression analysis

Breast cell line genomic DNA was demethylated by treatment with 5μM 5-azadc over 5 days. Daily media changes were made with the addition of fresh 5-azadc each time. All cell lines were maintained in DMEM (Sigma-Aldrich, St. Louis, MO, USA) supplemented with 2mM glutamine and 10% FCS at 37°C, 5% CO₂. RNA bee (AMS
Biotechnology, Abingdon, UK) was used to extract total RNA from cell lines according to manufacturers’ instructions. One microgram total RNA was converted to cDNA using Superscript III (Invitrogen, Carlsbad, CA, USA) and random hexamer primers (Fermentas UK, York, UK). RT-PCR was then carried out using a touchdown PCR program with a final annealing temperature of 56°C. All expression primer sets are shown in supplementary table S2.

**Real-time RT-PCR**

Quantitative real-time RT-PCR was performed as described previously (4). Briefly, cDNA was made from total RNA extracted from frozen normal and tumour breast tissues. We quantified transcripts of the *TBP* (TATA box-binding protein) gene as the endogenous RNA control. Results, expressed as N-fold differences in target gene expression relative to the *TBP* gene (termed N\textsubscript{target}), were determined by the following formula: N\textsubscript{target} = 2\textsuperscript{ΔC\textsubscript{tsample}}, where the ΔC\textsubscript{t} value of the sample was determined by subtracting the average Ct value of the target gene from the average Ct value of the *TBP* gene. The N\textsubscript{target} values of the samples were subsequently normalized such that the mean ratio of the normal breast samples would equal a value of 1. Primer sequences provided on request. PCR was performed using the SYBR Green PCR Core Reagents kit (Perkin-Elmer Applied Biosystems). The thermal cycling conditions comprised an initial denaturation step at 95°C for 10 min and 50 cycles at 95°C for 15 s and 65°C for 1 min. Experiments were performed with duplicates for each data point.

**Results**
HumanMethylation27 BeadChip data was obtained from 39 sporadic breast tumours and 4 matched breast tumour/normal pairs. Comparison of overall DNA methylation between normal breast DNA samples demonstrated little variance, with high correlation coefficient values ($R^2$) ranging from $R^2=0.974$ to $0.9302$ (Fig. S1). Additionally, hierarchical clustering (Euclidean, single linkage) of all the probes grouped all four matched normals together. Three out of four of the matched tumour/normal pairs demonstrated low correlations ($R^2$=0.724, 0.834 and 0.909), indicative of varying levels of aberrant methylation, while the fourth sample demonstrated a correlation similar to normals ($R^2=0.9381$). This suggested that altered methylation levels could be important in a large number of the sporadic breast tumour cohort analysed whilst for some tumours other genetic factors may be involved.

**Identification of hypermethylated CpG loci**

In order to identify the CpG loci showing the most significant tumour specific changes in methylation all probes with one or more normal sample showing a $\beta$ value $>$0.25 were excluded (n=11,211). Remaining probes were only considered hypermethylated if $>$20% of the tumour samples had either (a) $\beta$-values $\geq 0.5$ (n=261 probes, 242 loci), or, (b) a difference between the tumour sample $\beta$-value and the mean normal $\beta$-value of $\geq 0.4$ (n=202 probes, 182 loci). All imprinted and X chromosome genes were removed from the analysis and a final selection of 291 probes representing 264 genes was chosen (Suppl. Table S3, Fig. S2). The selected genes were analysed using PubMeth, identifying 34 genes out of the 264 (12.9%) that have been previously shown to be methylated in cancer, including 8 genes in breast cancer representing 10.3% of the genes in PubMeth methylated at a frequency of $>$20% in breast cancer (Suppl. Table S4).
Functional analysis of this selection using the DAVID bioinformatics resource (http://david.niaid.nih.gov) demonstrated the presence and enrichment of genes involved in functions related to tumorigenesis: cell adhesion (24 genes), regulation of cell proliferation (21 genes), negative regulation of cell death (15 genes), cell migration (12 genes), cell-cell adhesion (11 genes), regulation of cell cycle (10 genes) and tumour suppressor (6 genes) (Fig. S3). Similarly, DAVID KEGG pathway analysis demonstrated genes present from relevant pathways: pathways in cancer (12 genes), cell adhesion molecules (6 genes) and the Wnt signalling pathway (7 genes) (Fig. S3).

Further confirmation and analysis of the 291 selected probes was performed by bisulphite sequencing of (a) selected genes previously known to be associated with breast cancer, (b) analysis of most highly methylated genes, and (c) analysis to identify methylation specific to clinical features.

**Bisulphite sequencing confirmation of probe methylation**

For three selected hypermethylated genes previously known to be associated with breast cancer (SFRP5, CLDN6, and SIPA1) (5-8), bisulphite sequencing was performed on a region surrounding the CpG island probe for examples of high and low \( \beta \)-values. Sequencing demonstrated that \( \beta \)-values > 0.5 tended to be both highly methylated at the probe position and be indicative of wider CpG island methylation, whilst low \( \beta \)-values showed little methylation at the probe position, although in some cases, a greater level of methylation was observed in the surrounding region compared to the probe (Fig. 1). These results confirm the correlation of the \( \beta \)-value with methylation levels at the specific CpG being measured by the array.

**Analysis of breast tumour hypermethylated genes**
The selected genes were further filtered into two groups; the top ten most highly methylated genes with true CpG islands (amenable to CoBRA analysis) (Fig. S4a) and genes with more than a single selected probe (Fig. S5). Selected genes were then analysed for methylation in breast cancer cell lines, expression in breast cancer cell lines after treatment with 5’-aza-2’-deoxycytidine (5-azadc) and demonstration of tumour specific methylation in an additional independent set of 20 tumour/normal matched pairs, where appropriate.

Of the three most highly methylated genes with true CpG islands (ABCA3, COX7A1, and SST) COX7A1 and SST demonstrated increased expression after 5-azadc treatment in methylated breast tumour cell lines (Fig. S4b). In an independent cohort of breast tumour/normal pairs, COX7A1 demonstrated increased methylation in 70.6% (12/17) of tumours compared to corresponding normal tissue and a further 5 showed equal amounts of low level methylation in tumour and corresponding normal tissue. SST demonstrated tumour specific methylation in 31.6% (6/19) of pairs and increased methylation in a further 31.6% (6/19) of tumours compared to corresponding normal tissue DNA. One tumour showed equal amounts of methylation in the tumour and normal and no methylation in the remaining tumours (Fig S4b).

A total of 24 genes had more than a single selected probe, 4 of which (17%) had been previously demonstrated to be methylated in breast cancer (APC, CCND2, GSTP1 and PTPRO) (Fig. S5) (9-12). Further two genes (CDKL2 and ZNF154) from the above 24 genes were analysed in this study. Both CDKL2 and ZNF154 demonstrated increased expression after 5-azadc treatment in methylated breast tumour cell lines and also demonstrated altered methylation patterns in matched tumour/normal paired DNA (Fig
2). CDKL2 and ZNF154 demonstrated tumour specific methylation in 25% (4/16) and 56.3% (9/16) respectively, increased methylation in tumours compared to low level methylation of matched normal samples in 12.5% (2/16) and 25% (4/16) respectively. CDKL2 and ZNF154 also showed equal methylation in tumour and normal DNA in 18.8% (3/16) and 6.3% (1/16) of samples. No methylation was detected in the remainder of the tumours.

**Identification of gene hypermethylation specific to clinical features**

Initially, to investigate whether hypermethylation appeared to be associated with any clinical feature hierarchical clustering was performed on the 291 selected probes (Fig. 3). This split the sporadic tumour samples into 3 distinct methylation-based clusters; high methylation (18 tumours – mean β-value=0.433), intermediate methylation (7 tumours – mean β-value=0.317) and low methylation (12 tumours – mean β-value=0.176), with 2 outliers (Fig. 3). The high methylation cluster associated with samples that relapsed (high 12/18, intermediate 1/7 and low 3/12) and retained both the estrogen and progesterone receptor (high 14/18, intermediate 1/7 and low 2/12), whilst the low methylation cluster associated with the triple negative tumours (low 7/12, intermediate 2/7 and high 0/18) (Fig. 3).

ANOVA were performed using either the mean β-value (methylation) or the disease-free survival for each sample to compare between clinical features. The clusters demonstrated no statistically significant variation between disease-free survival (p=0.162) and no statistically significant variation was observed between methylation and tumour stage (p=0.214) or menopause (p=0.414) (Fig. S6). Statistically significant variation was
observed between methylation and tumour relapse (p=0.035), estrogen and progesterone receptor status (p=0.001) and lymph node metastasis (p=0.042) (Fig. 4).

In order to identify individual gene loci whose methylation status correlated with clinical features, Fisher’s exact tests with FDR (false discovery rate) correction were performed for all 291 selected probes between groups defined by a single clinical feature. This identified 9 methylated probes (9 genes) significantly associated with tumour relapse, 19 methylated probes (18 genes) significantly associated with presence of both estrogen and progesterone receptors and 2 probes specifically methylated in tumours with lymph node metastasis (Table S5). Probes specifically methylated in samples that relapsed should demonstrate strong correlations with disease-free survival and thus Kaplan-Meier analysis was performed. The 9 identified probes/genes (RECK, ACADL, SFRP2, ITR, UGT3A1, UAP1L1, SYDE1, UGT3A2, and PRKCB1) all demonstrated significant association with relapse-free survival and 6 remained significant after Bonferroni correction (underlined genes) (Fig. 5 and Table S6).

Selected genes associated with tumour relapse and the presence of both hormone receptors were further analysed for breast cancer cell line methylation, expression in breast tumour cell lines after treatment with 5-azadC, demonstration of tumour specific methylation in an additional 20 tumour/matched normal pairs and decreased expression in the original sporadic primary tumours (shown by HumanMethylation27 array to be methylated) by quantitative real-time RT-PCR, where appropriate.

For tumour relapse association, the three genes with the highest association with disease-free survival were selected; ACADL (p<0.0009), RECK (p=0.0009), and SFRP2 (p=0.0009) (Fig. 6). All three genes firstly showed increased expression or were re-
expressed (after 5-azadc treatment) in methylated breast tumour cell lines and secondly, demonstrated methylation in tumour/normal paired samples at a frequency of 30% (6/20), 22.2% (4/18) and 57.1% (8/14) of samples for \textit{ACADL}, \textit{RECK} and \textit{SFRP2} respectively (Fig. 6a and b (ii), (iii)) via COBRA analysis. COBRA analysis demonstrated no methylation in any of the corresponding normal breast tissues for \textit{RECK} and some methylation in adjacent normal tissues of methylated tumours in \textit{ACADL} (2/6) and \textit{SFRP2} (4/8), in all cases methylation observed in the associated normal tissue was lower than that observed in the tumour tissue. Bisulphite sequencing of \textit{ACADL} and \textit{RECK} demonstrated tumour specific methylation throughout the selected area of the CpG island (Fig. 6a and b (iv); Fig S7). Quantitative real-time RT-PCR demonstrated significant downregulation of \textit{ACADL} (p=0.043) and \textit{RECK} (p>0.0001) in methylated primary sporadic tumours compared to a panel of normal breast tissues (Fig. 6a and b (v)). In general, expression in unmethylated tumours was found to be higher than in methylated tumours for both \textit{ACADL} and \textit{RECK} (Fig. S8).

For association between methylation and the presence of both hormone receptors (ER and PR positive) three genes were selected (\textit{TNFRSF10D}, \textit{C1orf114} and \textit{COL1A2}) for analysis that were also present within the list of genes with multiple hypermethylated probes (Fig. S5). Again, all three genes firstly demonstrated increased expression in methylated breast tumour cell lines after 5-azadc treatment and secondly, demonstrated specific methylation in tumours of tumour/normal paired samples. \textit{COL1A2} demonstrated tumour specific methylation in 76.5% (13/17) of the paired samples (Fig. 7a). \textit{TNFRSF10D} demonstrated tumour specific methylation in 50% (8/16), increased methylation in the tumour compared to normal tissue in 18.8% (3/16) and 12.5% (2/16)
with similar levels in the tumour and normal tissue. C1orf114 demonstrated tumour specific methylation in 35.3% (6/17), increased methylation in the tumour compared to normal tissue in 42.2% (7/17) of tumours. To confirm the above COBRA methylation results we selected C1orf114 for bisulphite sequencing. Cloning and bisulphite sequencing of C1orf114 demonstrated tumour specific methylation throughout the selected area of the CpG island, hence confirming the COBRA results (Fig. 7b). In order to discern further biological relevance of the methylation found in the above three genes, we selected TNFRSF10D gene for expression analysis in breast tumour samples. Quantitative real-time RT-PCR analysis demonstrated significant downregulation of TNFRSF10D (p=0.016) in methylated primary sporadic tumours compared to a panel of normal breast controls (Fig. 7c). TNFRSF10D demonstrated higher expression levels in unmethylated tumours compared to methylated tumours in general, although not in every case, suggesting alternative mechanisms of downregulation (Fig. S8).

**Discussion**

The use of gene expression data from patient tumour samples to determine better treatment options is becoming increasingly common in clinical practice. The instability of mRNA makes expression profiling a challenging endeavour for utilization in routine clinical practice outside of major hospitals or commercial laboratories. Whilst DNA methylation patterns are more stable and can be detected using a number of high throughput and sensitive techniques which require little patient material. Using a high throughput quantitative methylation approach we established aberrant methylation patterns for a clinically well characterized cohort of breast tumours. The generated
methylation patterns demonstrated strong differentiation between tumour and normal DNA, with many tumour DNA samples showing a far greater number of hypermethylated CpG loci compared to normal samples, as well as differing patterns in relation to specific clinical features. The hypermethylated CpG loci identified in our study include both previously reported and novel methylated genes that fall into a wide range of functional categories. Tumours with lymph node metastasis, ERPR positive tumours and patients that had relapsed showed overall higher methylation versus patients with none of the above. There were 3 distinct methylation profiles generated within the tumour samples, ranging from the high methylation to intermediate to low methylation clusters. Although patient survival was not significantly different amongst the clusters (P> 0.05), statistically significant (p=<0.05) variation amongst the clusters was found for tumour relapse, ERPR receptor status and lymph node metastasis. Examination of individual loci/gene revealed that (1) methylation of 6 (RECK, SFRP2, ITR, UGT3A1, ACADL, UAP1L1) genes significantly associated with worst survival (2) Methylation of 18 (B3GAT1, TNFRSF10D, GRIA4, C1ORF114, DNAJC6, C6ORF145, CORIN, LAMA2, ITR, MCAM, C12ORF34, NPY, COL1A2, RSNL2, ZNF660, ADAMTS11, DPP4, MAT1A) genes/loci associated with ERPR positive receptor status (3) methylation of 2 (LDHD and ZNF660) genes associated with lymph node positive tumour status.

In order to show biological relevance for these methylated genes, we next examined relationship between gene expression and gene methylation for 2 genes most associated with poor survival (RECK and ACADL). In both cases expression analysis demonstrated that methylated breast tumours had lower gene expression versus normal breast tissue samples. We confirmed tumour specific and or increased methylation in
tumours compared to corresponding normal breast tissue for \textit{RECK} and \textit{ACADL} in an independent set of 20 breast normal/tumour paired DNA. \textit{RECK} is a membrane-anchored metalloprotease regulator, expression of \textit{RECK} is frequently reduced in various tumour types and reduced expression correlates with worst prognosis. Forced expression of \textit{RECK} in malignant cells and in mouse xenograft models leads to reduced angiogenesis, invasion and reduced metastasis (13-14). \textit{RECK} -/- mice die around embryonic day 10.5 with reduced tissue integrity and defects in various organs including blood vessels and central nervous system. \textit{RECK} has been shown to be methylated in NSCLC, gastric cancer, oral squamous cell carcinoma, colon cancer and melanoma. \textit{RECK} methylation is associated with poor prognosis and with lymph node metastasis in oral squamous cell carcinoma and NSCLC cancer respectively (15-16). But to our knowledge this is first report of \textit{RECK} methylation in breast cancer and that methylated \textit{RECK} is significantly associated with relapse and worst survival.

Effect of nutrition and lifestyle choices have been recognized to influence breast cancer risk and survival but there is a lack of studies investigating these breast cancer risk factors and tumour DNA methylation. A very recent study demonstrated that not only breast tumour size but also alcohol and folate intake were associated with DNA methylation profiles (17).

Hence it is interesting that we found \textit{ACADL} (acyl-CoA dehydrogenase, long chain) methylation is associated with poor survival in our cohort of breast cancer patients. The protein encoded by the \textit{ACADL} gene belongs to a family of mitochondrial flavoenzymes that are involved in fatty acid and branched chain amino acid metabolism. This protein is one of the four enzymes that catalyze the initial step of mitochondrial
beta-oxidation of straight-chain fatty acid. Yanagitani et al., demonstrated that transgenic mice expressing retinoic acid receptor alpha dominant negative form in hepatocytes exhibited severe fatty acid deposition (steatohepatitis) and liver tumours (18). The expression of mitochondrial enzymes involved in fatty acid beta-oxidation, including CAPT II, VLCAD, ACADL and HCD was greatly reduced in the transgenic mice compared to the wild type mice, thus hepatic steatosis was induced in transgenic mice due to impaired beta-oxidation of fatty acids in mitochondria. In addition Wnt signaling was activated in transgenic mice. Histological and biochemical abnormalities were reversed and occurrence of liver tumours was inhibited by feeding the transgenic mice with a high retinoic acid diet. The authors suggest that retinoic acid plays an important role in preventing hepatocellular carcinoma in association with fatty acid metabolism and Wnt signaling (18).

We also showed that SFRP2 methylation is significantly associated with poor survival. Secreted frizzled proteins (SFRPs) act as soluble modulators of Wnt signaling. There are 5 SFRP genes, the gene family has been shown to be methylated in various cancers including breast cancer. Methylation of SFRP1 and SFRP5 in breast cancer is associated with unfavorable prognosis (19, 5). SFRP2 methylation has previously been reported in breast, bladder, colon, liver, lung, oesophagus stomach cancer, leukemia etc. Whilst, detection of SFRP2 methylation in fecal DNA has been shown to be a marker for colorectal cancer and for detection of colorectal polyps (20-22). In breast cancer SFRP2 methylation was not associated with clinical patient outcome or other clinicopathological factors, but loss of SFPR2 protein expression was weakly linked to poor patient overall survival (23). Overexpression of SFRP2 in mammary cells lead to inhibition in cell
proliferation. A recent article linked methylation of sFRP2 and sFRP5 with adverse clinical outcome in patients with normal karyotype acute myeloid leukemia (24). In this report we have shown that methylation of SFRP2 in breast cancer is associated with poor patient survival.

The remaining three genes \{UGT3A1 (UDP glycosyltransferase 3 family, polypeptide A1); UAPIL1 (UDP-N-acetylglucosamine pyrophosphorylase 1-like 1); IRT/GPR180 (G protein-coupled receptor 180)\} associated with survival in our cohort of breast cancer patients are novel genes and to our knowledge epigenetic inactivation of these genes have not been described previously in cancer, further work will be required to understand their role in breast cancer development.

In a recent genome-wide DNA methylation study in Korean breast cancer patients, hyper/hypo methylation of five genes was linked to ER+/PR+ positive receptor status (25). We have identified 18 hypermethylated genes associated with ER+/PR+ receptor status in Caucasian breast cancer patients. Li et al identified FAM124B and ST6GALNAC1 hypermethylation associated with ER+/PR+ status in Korean breast cancer patients. These 2 genes were not linked to ERPR status in our cohort of breast tumour cases from Caucasian individuals. This could be due to the ethnical differences between the 2 sets of breast cancer patients in addition we used a high stringency criteria to identify differentially methylated genes in tumour versus normal DNA. Hence according to our selection criteria the above two genes were not differentially methylated in tumour versus normal DNA.

We identified TNFRSF10D (DCR2) gene amongst the 18 genes/loci significantly hypermethylated in ER/PR+ breast tumours compared to receptor negative breast
tumours. TNFRSF10D (DCR2), the protein encoded by this gene is a member of the TNF-receptor superfamily. Tumor necrosis factor (TNF)-related apoptosis-inducing ligand TRAIL can induce apoptosis in cancer cells but has little effect on normal cells, hence it is a promising anticancer agent. However, cancer cells are often resistant to TRAIL-induced cell death. It was recently demonstrated that depletion of adenine nucleotide translocase-2 in breast cancer cells sensitized these cells to TRAIL induced apoptosis by activating JNK and modulating the expression of TRAIL receptors, including downregulation of DCR2 expression both in in vitro and in vivo studies (26). DCR2 methylation has been previously reported in several cancers including, NSCLC, breast, prostate etc (27). COL1A2 methylation was also associated with ERPR receptor status in our breast cancer cohort, was cancer specific and could be reversed in methylated breast tumour cell lines following treatment with 5-azade. Using computational protocols Loss et al recently demonstrated COL1A2 methylation in breast tumour cell lines (28). In medulloblastoma COL1A2 methylation was found to be subgroup-specific and in bladder cancer COL1A2 gene inactivation due to methylation was shown to contribute to cell proliferation and migration of bladder cancer cells (29-30).

In summary we have identified a set of genes hypermethylated in breast cancer that are linked to patient clinical outcome, confirmation of our results in a larger breast cancer cohort and subsequent biological analysis of the proteins encoded by these genes may lead to a better understanding of breast cancer pathogenesis and development of clinical applications.
Conflict of interest

Authors declare no conflict of interest

Grant Support

F Latif laboratory funded by Breast Cancer Campaign and Cancer Research UK.

References


**Figure legends**

**Figure 1. Validation of the HumanMethylation27 BeadChip with clone sequencing.**

Clone sequencing results are shown for three samples each, for each of the three genes,
SFRP5, CLDN6 and SIPA1. In each case the Illumina ID (ILMNID) is given for the probe being validated and the location of the CpG being assessed in the array shown on a gene schematic and highlighted on the clone sequencing results. For each sample sequenced the β-value is given and the methylation index (MI) for the overall region as a percentage of methylated CpGs. For each sample, each line represents a single allele, black circles represent a methylated CpG and each white circle represents an unmethylated CpG.

**Figure 2. CoBRA and expression results for CDKL2 and ZNF154**

For CDKL2 and ZNF154, CoBRA results are shown for selected cell lines (top) and tumour normal pairs (bottom). In each case, undigested product (Un) is shown next to digested product (Di). T = tumour, N= normal, Meth= CoBRA result for cell lines. Expression results pre (-ve) and post (5-aza) treatment with 5-azadc are also shown next to the corresponding cell line methylation results. In each case, results are shown for the target gene and GAPDH. Expression of the target gene in normal breast tissue is also shown.

**Figure 3. Hierarchical clustering splits tumour samples into three main groups**

The dendogram can be seen aligned with sample information showing clinical status of: age at diagnosis (years), tumour size (mm), SBR stage (1,2 or 3), lymph node status (LyN = lymph node positive; N = lymph node negative), menopause status (MP = menopausal; N = not menopausal), relapse status (R= relapse; NR = no relapse), estrogen and progesterone receptor status (ERPR = estrogen and progesterone positive; ERnoPR =
estrogen positive only; PRnoER = progesterone positive only; noR = estrogen and progesterone negative), ERBB2 status (ERBB2 = ERBB2 overexpression, N=normal expression), triple negative and survival (disease free survival in days). The three cluster groups, showing high, medium and low methylation are separated with thick black lines and labelled underneath. Two outliers can be seen on the extreme left of the clustering image. Black and white shading refers to β-values >0.5 and <0.5 respectively.

**Figure 4. ANOVA analysis shows overall methylation levels are significantly different in relapsing, ERPR expressing and lymph node positive patients**

ANOVA analysis shows significant differences in overall methylation levels (as determined by the mean β-value) for (a) relapse (p=0.035), (b) ER and PR receptor status (p=0.001) and (c) lymph node metastasis (p=0.042).

**Figure 5. Survival and individual loci methylation**

The methylation status is shown for each tumour sample for each probe showing methylation significantly associating with relapse (black circles represent methylation, white circles represent no methylation). The numbers of methylated samples for relapsing tumours and non-relapsing tumours are shown for each probe along with FDR corrected p-values (a). Kaplan-Meier figures are shown for all six loci that showed significance (p<0.05). In each case dotted lines (M) represents methylation and the solid line (U) represents no methylation. P values are also shown (b).

**Figure 6. RECK and ACADL tumour specificity and primary tumour expression**
For both RECK (6a) and ACADL (6b), a gene schematic showing the location of the CoBRA primers is shown (i). CoBRA results are shown for selected cell lines (Meth.) along with expression (Exp.) with (5-aza) and without (-ve) treatment with 5-azaDC for both the target gene and GAPDH (ii). CoBRA results are shown for selected tumour normal pairs (iii) along with clone sequencing (iv). Quantitative real time expression results are shown for methylated primary tumours (v). All CoBRA results are shown with undigested product (U) run next to digested product (D). Each line in the clone sequencing results represents a single allele, each black circle a methylated CpG and each white circle an unmethylated CpG. Methylation indexes (MIs) were calculated as a percentage of methylated CpGs out of the total number of CpGs analysed. T=tumour N=normal.

**Figure 7. ERPR receptor status and individual loci methylation**

(a) CoBRA results are shown for cell lines and tumour normal pairs along with expression analysis +/- 5-azadc treatment for TNFRSF10D, COL1A2 and C1orf114. Un = undigested, Di = digested, 5aza= treatment with 5-azadc, -ve = without treatment with 5-azadc. (b) C1orf114 clone sequencing results are shown for tumour normal pairs, black circles representing methylated CpGs and white circles representing unmethylated CpGs, methylation indexes (MIs) were calculated as a percentage of methylated CpGs out of the total number of CpGs analysed. (c) Quantitative real-time PCR results are shown for TNFRSF10D. Significance between the mean normal expression and mean methylated tumour expression is given (Student’s t-test).
Figure 1

SFRP5
NM_003015 cg05937453

β value = 0.54
MI = 76.0 %

CLDN6
NM_021195 cg07384961

β value = 0.80
MI = 95.7 %

SIPA1
NM_153253 cg25361844

β value = 0.53
MI = 100 %

β value = 0.70
MI = 83.8 %

β value = 0.54
MI = 33.9 %

β value = 0.13
MI = 20 %
Figure 3

Cluster 1 – High Methylation

Cluster 2

Cluster 3 – Low Methylation
ANOVA p-value = 0.035
Sample size = 39 (Relapse 17 + No Relapse 22)
Levene’s Test for Equality of Variances p-value = 0.798

ANOVA p-value = 0.001
Sample size = 30 (ERPR exp. 17 + No ERPR 13)
Levene’s Test for Equality of Variances p-value = 0.358

ANOVA p-value = 0.042
Sample size = 39 (LYN 31 + N 8)
Levene’s Test for Equality of Variances p-value = 0.688
Figure 5

(a) Table showing methylation status for RELAPSE and NO RELAPSE groups.

(b) Kaplan-Meier survival curves for different genes:
- **RECK**: 
  - Survival probability at 6000 days:
  - Time (days) range: 0 to 6000
- **SFRP2**: 
  - Survival probability at 6000 days:
  - Time (days) range: 0 to 6000
- **UAP1L1**: 
  - Survival probability at 6000 days:
  - Time (days) range: 0 to 6000
- **UGT3A1**: 
  - Survival probability at 6000 days:
  - Time (days) range: 0 to 6000
- **ACADL**: 
  - Survival probability at 6000 days:
  - Time (days) range: 0 to 6000
- **ITR**: 
  - Survival probability at 6000 days:
  - Time (days) range: 0 to 6000

Survival probability values are given for specific time points, with p-values noted for each gene.

**RECK**: p = 0.0001
**SFRP2**: p = 0.0001
**UAP1L1**: p = 0.0005
**UGT3A1**: p = 0.0018
**ACADL**: p < 0.0001
**ITR**: p = 0.0041
Figure 6

(a) RECK

(i) Tumour vs Normal

(ii) HTB19, T47D, MDA-MB-231

(iii) T/N Pair 1206, T/N Pair 1018, T/N Pair 281

(iv) Analysis of RECK expression

(v) Statistical analysis

(b) ACADL

(i) Tumour vs Normal

(ii) HCC1806, MCF-7, T47D

(iii) T/N Pair 683, T/N Pair 1512, T/N Pair 498

(iv) Analysis of ACADL expression

(v) Statistical analysis
Figure 7

(a) TNFRSF10D

<table>
<thead>
<tr>
<th></th>
<th>HCC1395</th>
<th>T47D</th>
<th>MCF-7</th>
<th>T/N Pair 325</th>
<th>T/N Pair 1206</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meth.</td>
<td><img src="image1" alt="Image" /></td>
<td><img src="image2" alt="Image" /></td>
<td><img src="image3" alt="Image" /></td>
<td><img src="image4" alt="Image" /></td>
<td><img src="image5" alt="Image" /></td>
</tr>
<tr>
<td>Exp.</td>
<td><img src="image6" alt="Image" /></td>
<td><img src="image7" alt="Image" /></td>
<td><img src="image8" alt="Image" /></td>
<td><img src="image9" alt="Image" /></td>
<td><img src="image10" alt="Image" /></td>
</tr>
<tr>
<td>5-aza</td>
<td><img src="image11" alt="Image" /></td>
<td><img src="image12" alt="Image" /></td>
<td><img src="image13" alt="Image" /></td>
<td><img src="image14" alt="Image" /></td>
<td><img src="image15" alt="Image" /></td>
</tr>
</tbody>
</table>

(b) COL1A2

<table>
<thead>
<tr>
<th></th>
<th>HCC1395</th>
<th>HTB19</th>
<th>MCF-7</th>
<th>T/N Pair 139</th>
<th>T/N Pair 2006</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meth.</td>
<td><img src="image16" alt="Image" /></td>
<td><img src="image17" alt="Image" /></td>
<td><img src="image18" alt="Image" /></td>
<td><img src="image19" alt="Image" /></td>
<td><img src="image20" alt="Image" /></td>
</tr>
<tr>
<td>Exp.</td>
<td><img src="image21" alt="Image" /></td>
<td><img src="image22" alt="Image" /></td>
<td><img src="image23" alt="Image" /></td>
<td><img src="image24" alt="Image" /></td>
<td><img src="image25" alt="Image" /></td>
</tr>
<tr>
<td>5-aza</td>
<td><img src="image26" alt="Image" /></td>
<td><img src="image27" alt="Image" /></td>
<td><img src="image28" alt="Image" /></td>
<td><img src="image29" alt="Image" /></td>
<td><img src="image30" alt="Image" /></td>
</tr>
</tbody>
</table>

(c) C1orf114

<table>
<thead>
<tr>
<th></th>
<th>HCC1395</th>
<th>MCF-7</th>
<th>T47D</th>
<th>T/N Pair 1693</th>
<th>T/N Pair 2234</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meth.</td>
<td><img src="image31" alt="Image" /></td>
<td><img src="image32" alt="Image" /></td>
<td><img src="image33" alt="Image" /></td>
<td><img src="image34" alt="Image" /></td>
<td><img src="image35" alt="Image" /></td>
</tr>
<tr>
<td>Exp.</td>
<td><img src="image36" alt="Image" /></td>
<td><img src="image37" alt="Image" /></td>
<td><img src="image38" alt="Image" /></td>
<td><img src="image39" alt="Image" /></td>
<td><img src="image40" alt="Image" /></td>
</tr>
<tr>
<td>5-aza</td>
<td><img src="image41" alt="Image" /></td>
<td><img src="image42" alt="Image" /></td>
<td><img src="image43" alt="Image" /></td>
<td><img src="image44" alt="Image" /></td>
<td><img src="image45" alt="Image" /></td>
</tr>
</tbody>
</table>

(b) (c) MI = 87.9% MI = 0%

MI = 76%

MI = 0%

p = 0.016
Genome-wide DNA methylation profiling of CpG islands in breast cancer identifies novel genes associated with tumorigenicity

Victoria K Hill, Christopher Ricketts, Ivan Bieche, et al.

Cancer Res  Published OnlineFirst March 1, 2011.

Updated version  Access the most recent version of this article at: doi:10.1158/0008-5472.CAN-10-4026

Supplementary Material  Access the most recent supplemental material at: http://cancerres.aacrjournals.org/content/suppl/2011/03/01/0008-5472.CAN-10-4026.DC1

Author Manuscript  Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

E-mail alerts  Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions  To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions  To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.