Carcinogenesis vol.34 no.1 pp.102–108, 2013
doi:10.1093/carcin/bgs321
Advance Access publication October 8, 2012

DNA methylation profiling in breast cancer discordant identical twins identifies *DOK7* as novel epigenetic biomarker

Holger Heyn[1], F. Javier Carmona[1], Antonio Gomez[1], Humberto J.Ferreira[1,2], Jordana T.Bell[3], Sergi Sayols[1], Kirsten Ward[3], Olafur A.Stefansson[1], Sebastian Moran[1], Juan Sandoval[1], Jorunn E.Eyfjord[4,5], Tim D.Spector[3] and Manel Esteller[1,6,7,*]

[1]Cancer Epigenetics and Biology Program (PEBC), Bellvitge Biomedical Research Institute (IDIBELL), L'Hospitalet, Barcelona, Catalonia 08907, Spain,[2]Programme in Experimental Biology and Biomedicine, Centre for Neurosciences and Cell Biology, University of Coimbra, 3004-517 Coimbra, Portugal, [3]Department of Twin Research and Genetic Epidemiology, Kings College London, St Thomas' Hospital Campus, London SE1 7EH, UK, [4]The Icelandic Cancer Society, Molecular and Cell Biology Research Laboratory, 101 Reykjavik, Iceland, [5]Department of Medicine, University of Iceland, 101 Reykjavik, Iceland, [6]Department of Physiological Sciences II, School of Medicine, University of Barcelona, 08007 Barcelona, Catalonia, Spain and [7]Institucio Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Catalonia, Spain

*To whom correspondence should be addressed. Tel: 0034-93-260-7140; Fax: 0034-93-260-7219; Email: mesteller@idibell.cat
Correspondence can also be addressed to T.D.Spector. Tel: +44 207 188 6765; Fax: +44 207 188 6718;
Email: tim.spector@kcl.ac.uk

Using whole blood from 15 twin pairs discordant for breast cancer and high-resolution (450K) DNA methylation analysis, we identified 403 differentially methylated CpG sites including known and novel potential breast cancer genes. Confirming the results in an independent validation cohort of 21 twin pairs determined the docking protein *DOK7* as a candidate for blood-based cancer diagnosis. DNA hypermethylation of the promoter region was also seen in primary breast cancer tissues and cancer cell lines. Hypermethylation of *DOK7* occurs years before tumor diagnosis, suggesting a role as a powerful epigenetic blood-based biomarker as well as providing insights into breast cancer pathogenesis.

Introduction

Breast cancer is the most common female neoplasm affecting around one in nine women. There is a genetic susceptibility, which accounts for up to 30% of the heritability of breast cancers. Familial causes of breast cancer are almost exclusively related to *BRCA1* and *BRCA2*, genes involved in the homologous recombination-mediated DNA repair. In sporadic cancers, mutations are rarely found; however, epigenetic gene silencing by DNA hypermethylation of *BRCA1* is observed frequently (1–3). Identifying miss-regulated breast cancer genes enabled the development of therapies specifically targeting aberrant pathways, such as poly (ADP ribose) polymerase inhibitors, impairing an independent DNA repair mechanism selectively targeting *BRCA1/2*-mutated cells (4). In sporadic cases, hypermethylation of *BRCA1* was shown to sensitize tumor cells to poly (ADP ribose) polymerase inhibitors (5) and also conventional DNA damaging agents such as cisplatin (6). Aberrantly regulated *BRCA1* illustrates the potential of tumor-specific markers as diagnostic and novel treatment strategies. As mutations are observed at low frequencies, epigenetic profiling represents a promising approach to discover novel disease-specific markers.

Epigenetic changes are now known to play a key role in most kinds of cancer—both in the early and late stages of disease (7). High-resolution technologies, such as whole-genome bisulfite sequencing, unraveled hypomethylated blocks, covering large parts of the cancer methylome (8). However, distinct loci mainly related to CpG-rich islands and promoters are sites of hypermethylation, previously related to tumor-suppressor gene silencing (9). DNA methylation is not entirely independent from the genetic background as methylation quantitative trait loci represent single nucleotide polymorphisms highly associated to methylation events at CpG sites (10,11). Thus, using identical twins for epigenetic studies is the most efficient design available as it controls for genetic factors, age, cohort effects and many environmental influences that otherwise add variability and noise (12). Particularly, sample types presenting small changes in methylation benefit from a setup depleted of genetic variation. In this respect, the identification of epigenetic cancer biomarkers using biological fluids takes advantage of methylation profiling free from disturbing genetic influences.

In this study, we aimed to identify, novel breast cancer-specific epigenetic biomarker in blood. Consequently, we performed DNA methylation profiling using DNA extracted from whole blood of monozygotic (MZ) twins discordant for breast cancer and the Infinium DNA methylation BeadChip technology covering more than 450 000 CpG sites genome wide (13,14). This setup enabled us to unravel alteration in DNA methylation at high resolution independent of genetic variation. Similar studies screening for epigenetic differences of MZ twins discordant for type 1 diabetes (15) or systemic lupus erythematosus (16) previously established the potential of the study design, however using profiling platforms with much lower resolution.

Materials and methods

Sample preparation

DNA samples from blood were extracted from thawed frozen whole blood collected in ethylenediaminetetraacetic acid using the Nucleon Genomic DNA Extraction Kit BACC3. DNA samples from triple-negative breast tumors and adjacent normal tissue were obtained from freshly frozen samples. The samples were macroscopically examined by pathologist prior to DNA isolation and portions of normal and tumor tissue selected. DNA isolation was then carried out using a standard phenol–chloroform plus proteinase K protocol. The use of these samples was in accordance with permits from the Icelandic Data Protection Commission (2006050307) and Bioethics Committee (VSNb2006050001/03-16). Informed consent was obtained from all patients. All participants in this study were of Caucasian ethnicity.

Pyrosequencing

Specific sets of primers for PCR amplification and sequencing were designed using a specific software pack (PyroMark assay design version 2.0.01.15). Primer sequences were designed, when possible, to hybridize with CpG-free sites to ensure methylation-independent amplification. PCR was performed under standard conditions with biotinylated primers, and the PyroMark Vacuum Prep Tool (Biotage, Uppsala, Sweden) was used to prepare single-stranded PCR products according to manufacturer's instructions. Pyrosequencing reactions and methylation quantification were performed in a PyroMark Q96 System version 2.0.6 (Qiagen) using appropriate reagents and recommended protocols.

Infinium HumanMethylation450 BeadChip

All DNA samples were assessed for integrity, quantity and purity by electrophoresis in a 1.3% agarose gel, picogreen quantification and nanodrop measurements. All samples were randomly distributed into 96-well plates. Bisulfite conversion of 500 ng of genomic DNA was performed using EZ DNA methylation kit (Zymo Research) following manufacturer's instructions. About 200 ng of bisulfite-converted DNA was used for hybridization on the HumanMethylation450 BeadChip (Illumina). Briefly, samples were whole genome amplified followed by an enzymatic end-point fragmentation, precipitation and resuspension. The resuspended samples were hybridized onto the BeadChip for 16 h at 48°C and washed. A single nucleotide extension with

Abbreviations: bcDMP, differentially methylated CpG positions in breast cancer; DOK7, docking protein 7; MZ, monozygotic; UTR, untranslated region.

labeled dideoxy-nucleotides was performed, and repeated rounds of staining were applied with a combination of labeled antibodies differentiating between biotin and 2,4-dinitrophenol (DNP). Color balance adjustment and quantile normalization were performed in order to normalize the samples between the two color channels. DNA methylation level is displayed as beta-values ranging from 0–1. Beta-values with detection *P*-value >0.01 are considered to fall below the minimum intensity and threshold and were consequently removed from further analysis.

Statistical analysis

To identify consistently differentially methylated CpG sites Wilcoxon rank sum paired test was performed for normalized beta-values of paired twins. The *P*-values were adjusted using false discovery rate (17), and those CpGs with *P*-values <0.05 were selected. To cluster the twins, we used the 'complete' agglomeration method for hierarchical clustering using 'Euclidean' distances of a selected subset of CpG sites. The subset was calculated using a multivariate filter called correlation feature selection (18). This technique is based upon the hypothesis that good variable sets are those with variables highly correlated with the classification and uncorrelated to each other.

Network building and analysis

The genes associated to differentially methylated CpG sites selected by Wilcoxon rank sum test were mapped to known genetic interactions and co-expression data sets using GeneMANIA (19). To identify additional genes that interact to the input set genes, the resulting network of 279 nodes and 3095 edges was analyzed using the ClueGO plugin (20) in Cytoscape in order to identify enrichment in Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways.

Results

We used a group of discordant breast cancer twins from the UK-based EpiTwin study to see if we could uncover the key genes involved epigenetically in either the cancer process or the susceptibility to cancer using a whole epigenome approach. Consequently, we performed genome-wide DNA methylation profiling using DNA extracted from whole blood of MZ twins discordant for breast cancer. We obtained a comprehensive DNA methylation profile of 15 discordant twins (identification cohort), using the high-resolution Infinium HumanMethylation450 BeadChip platform (450K, Illumina), previously established to reliably detect methylation changes of >450 000 CpG sites (Supplementary Table 1, available at *Carcinogenesis* Online). To provide insight into the temporal and causal relationships

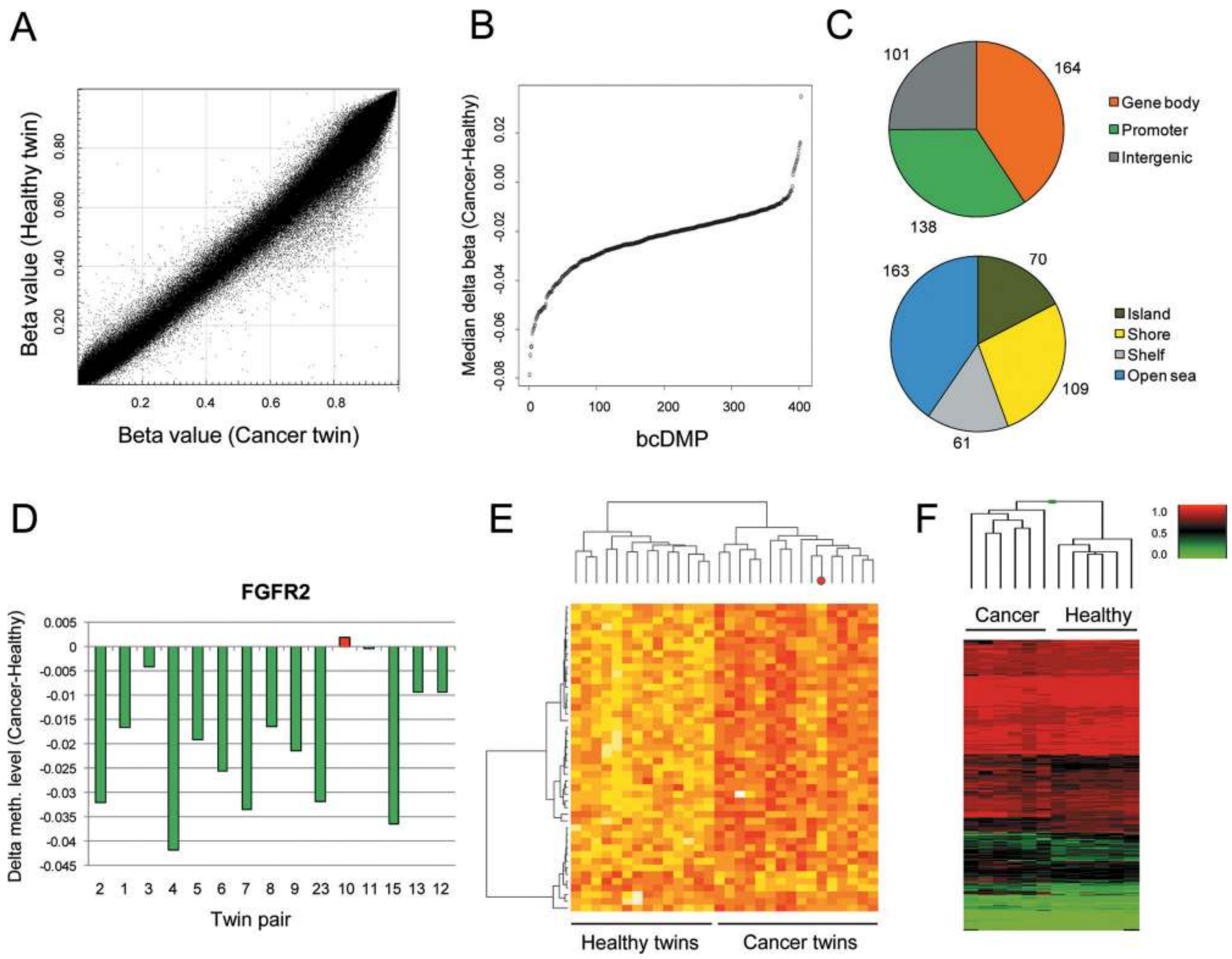

Fig. 1. Differentially methylated CpG sites within MZ twin pairs discordant for breast cancer. (**A**) DNA methylation level of CpG sites identified by the Infinium 450K DNA methylation assay. Displayed are normalized beta-values of one representative example of discordant twins (990836 and 989697). (**B**) 403 differentially methylated CpG sites (bcDMP) within twins discordant for breast cancer identified by Wilcoxon signed rank test ($P < 0.05$) and represented as median beta-value differences (cancer-healthy twin). bcDMP were ranked by median beta-value difference. (**C**) Genomic distribution of bcDMPs regarding their respective location to genes and CpG context. (**D**) Delta methylation level (cancer-healthy twin) of a differentially methylated promoter CpG site of *FGFR2* (cg12835048). (**E**) bcDMP varying epigenetically despite the identical genetic background of the twins identified by multivariate filter analysis. The red dot identifies a healthy sample within the cancer cluster. The bcDMP DNA methylation level is color coded (yellow: sample with lowest methylation level; red: sample with highest methylation level). (**F**) Hierarchical cluster of bcDMPs in six primary breast cancer pairs analyzed on the Infinium HumanMethylation450 BeadChip platform.

and predictive potential, samples from breast cancer patients before (7) and after diagnosis (8) were also analyzed.

Direct intra-twin pair comparisons of 15 pairs revealed high correlation between the pairs (Spearman's correlation; median $r^2 = 0.991$; range: 0.977–0.995). A representative example is presented in Figure 1A. To identify CpG sites altered within twin pairs, we applied a Wilcoxon pair rank test ($P < 0.001$), determining 403 consistently differentially methylated CpG positions in breast cancer samples (bcDMP) compared with the matched healthy twin (Figure 1B). All identified bcDMPs are listed in Supplementary Table 2, available at *Carcinogenesis* Online. Strikingly, 97% (390 of 403) of bcDMPs were hypomethylated in cancer patients with only 3% gaining methylation. Accordingly, a global loss of DNA methylation was previously reported in primary breast cancer specimens analyzed at base-pair resolution (8), supporting the sensitivity of our current blood-based approach. The identified sites were associated with 315 distinct genes and 138 gene promoters (Figure 1C). The majority (82.6%) of bcDMPs were located outside CpG-rich regions (CpG islands), with 27% located in CpG shores flanking the islands, recently described to be of high importance for gene regulation and tumorigenesis (21,22) (Figure 1C). Interestingly, the fibroblast growth factor receptor 2 (*FGFR2*) previously associated with breast cancer susceptibility in genome-wide association studies was among the hypomethylated genes (23) (Figure 1D). Gene ontology analysis revealed a functional enrichment of bcDMPs in biological processes presenting crucial checkpoints of cancer formation, such as cell cycle arrest (Fisher's exact test; $P < 3.2 \times 10^{-4}$) and regulation of programmed cell death (Fisher's exact test; $P < 1.7 \times 10^{-2}$). In this respect, network analysis of genes associated to bcDMPs revealed a significant enrichment (two-sided hypergeometric test with Bonferroni adjustment, $P < 0.01$) in cancer-specific pathways (KEGG) such as colorectal, pancreatic, small cell lung cancer and chronic myeloid leukemia (CML) (Figure 2). Drivers of pathway enrichment were prominent cancer candidates such as *MYC*, *SMAD3/4*, *MAPK10* and *PIK3CG* among others.

To extract CpG sites that have different methylation profile within twin pairs and also allow us to cluster cancer and healthy twin separately and hence to function as a breast cancer signature in blood, we applied a multivariate filter analysis called correlation feature selection (18). This technique is based upon the hypothesis that good variable sets are those with variables highly correlated with the classification and uncorrelated to each other (Figure 1E). Here, we identified 46 genes that varied epigenetically despite the identical genetic background of the twins and able to almost perfectly cluster the samples according to the presence of cancer. All identified sites are listed in Supplementary Table 3, available at *Carcinogenesis* Online. We then looked at the differentially methylated sites in a primary breast tumor setting using six samples and their matched normal controls on the 450K platform. We found these identified bcDMPs were able to separate the cancer and healthy samples using a hierarchical cluster approach (Figure 1F).

We then excluded case samples from the identification cohort, which were obtained before diagnosis. In this more stringently selected data set [eight twin pairs; average 2.1 years after diagnosis (range: 0–4 years)] reflecting more the consequences of disease, we found 5188 bcDMPs, with similar directional distribution (81.7% hypo- and 18.2% hypermethylated sites) as detected before (Wilcoxon signed rank test; $P < 0.05$). All identified sites are listed in Supplementary Table 4, available at *Carcinogenesis* Online. Among these identified

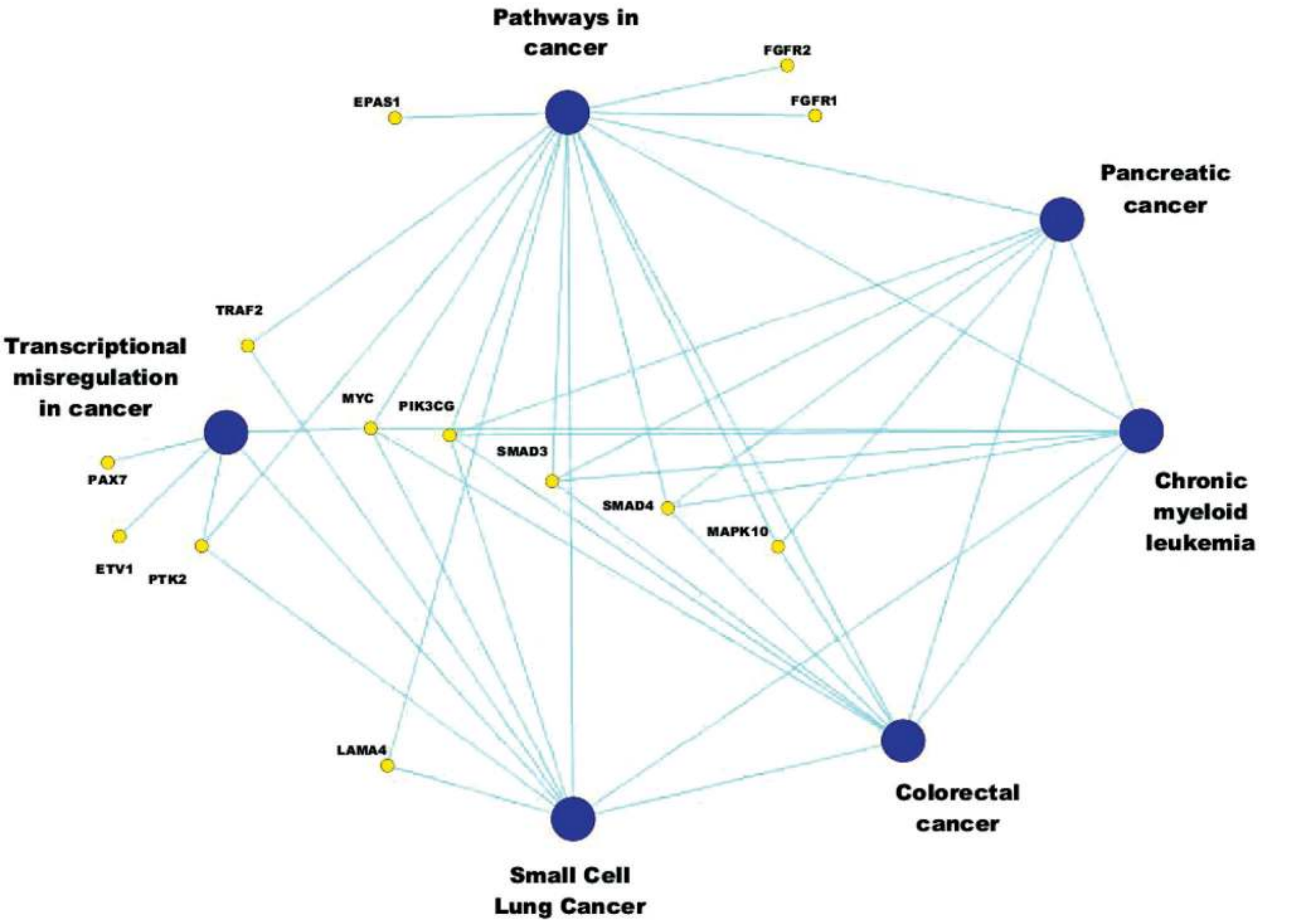

Fig. 2. bcDMP associated genes are enriched in cancer associated pathways. Network analysis using GeneMANIA and ClueGO identified cancer-related KEGG pathways enriched in bcDMP associated genes.

bcDMPs, we found further genes previously associated with breast cancer susceptibility and pathology, such as the lymphocyte-specific protein (*LSP1*; genome-wide association studies), the v-akt murine thymoma viral oncogene homolog 1 (*AKT1*) and cyclin D1 (*CCND1*; both Cancer Gene Census (24)). Broadening the search to genes previously associated to all cancer types identified in total 45 genes overlapping with CpG sites identified as bcDMPs (Table I).

To establish bcDMPs associated genes as novel epigenetic biomarkers for breast cancer, we aimed to validate differentially methylated CpG sites in an independent set of MZ twins discordant for breast cancer, as well as in matched primary breast cancer specimens. Therefore, we analyzed 21 (16 post- and 5 prediagnosis) additional twin pairs by locus-specific pyrosequencing for alterations in genes revealing high differences in the identification cohort or previously associated to breast cancer. In total, we profiled 14 CpG sites for differences in DNA methylation in the validation cohort. In detail, five Cancer Gene Census/genome-wide association studies (*LSP1*, *FGFR1*, *FGFR2*, *MYC*, *AKT1*), two imprinted genes (*PHLDA2*, *IGF2*) and seven genes showing high differences in the identification set (*HMGB3*, *FNIP2*, *TCRBV14S1*, *FAM196B*, *MAP9*, *THBS1*, *DOK7*) were pyrosequenced.

A CpG site (cg15652666; chr.4:3487436, HG19) in an alternative promoter of the docking protein 7 (*DOK7*) revealed clear differences between the paired samples analyzed (Figure 3A). Both the identification (Figure 3B) and validation of the postdiagnosis samples (Figure 3C) showed significant consistent DNA hypermethylation in the cancer twin compared with the paired healthy co-twin (Wilcoxon signed rank test; $P < 0.05$). Most strikingly, in addition to the original identified *DOK7* bcDMP, neighboring upstream CpG sites ($n = 4$) also revealed significant consistent methylation differences between the twin pairs (Wilcoxon signed rank test; $P < 0.05$; Figure 3D). Determining five consecutive CpG site to be affected, we suggest the whole loci as breast cancer differentially methylated region. When we excluded the paired information from the analysis and grouped the samples in a case-control analysis according to their healthy and cancer status, we also showed significant differences (Mann–Whitney test; $P < 0.05$; Figure 3E). Interestingly, the significant difference could be observed for the entire region (Mann–Whitney test; $P < 0.05$; Figure 3F). The bcDMP was located in a CpG island shore region 1.2 kb downstream of the alternative transcription start site in close proximity to the transcription factor binding sites of HMX1, PAX6 and CREB (Figure 3A). To verify that the region is of biological relevance, we determined the methylation profile overlapping the

Table I. bcDMPs overlapping genes previously associated to Cancer Gene Census

Gene	Cancer/disease type	Data set	Direction	Location	CpG context
MYC	Burkitt lymphoma, amplified in other cancers, B-CLL	Before/after diagnosis	Hypomethylation	Body	Shore
GNAS	Pituitary adenoma	Before/after diagnosis	Hypomethylation	Promoter	Shore
FGFR1	MPD, NHL	Before/after diagnosis	Hypomethylation	Promoter	Shore
POU5F1	Sarcoma	After diagnosis	Hypomethylation	Promoter	
JAZF1	Endometrial stromal tumors	After diagnosis	Hypomethylation	Body	
LMO1	T-ALL, neuroblastoma	Before/after and after diagnosis	Hypomethylation	Body	Shelf
NOTCH1	T-ALL	Before/after diagnosis	Hypomethylation	Body	Shore
SET	AML	After diagnosis	Hypomethylation	Body	Shelf
GNA11	Uveal melanoma	After diagnosis	Hypomethylation	3'UTR	Shore
TPM3	Papillary thyroid, ALCL	After diagnosis	Hypomethylation	Promoter	
PIK3R1	Gliobastoma, ovarian, colorectal	After diagnosis	Hypomethylation	Promoter	
NFIB	Adenoid cystic carcinoma, lipoma	After diagnosis	Hypomethylation	Body	
SRGAP3	Pilocytic astrocytoma	Before/after diagnosis	Hypomethylation	Body	Shore
EBF1	Lipoma	After diagnosis	Hypomethylation	Body	Shore
SETD2	Clear cell renal carcinoma	After diagnosis	Hypomethylation	3'UTR	
PAX7	Alveolar rhabdomyosarcoma	Before/after diagnosis	Hypomethylation	Body	
TRIM27	Papillary thyroid	After diagnosis	Hypomethylation	Body	
MUTYH	Colorectal, adenomatous polyposis coli	Before/after diagnosis	Hypomethylation	Body	Shelf
FANCA	AML, leukemia, Fanconi anemia A	Before/after diagnosis	Hypomethylation	Body	
CCND1	CLL, B-ALL, breast	After diagnosis	Hypomethylation	Body	Shore
IKZF1	ALL	After diagnosis	Hypomethylation	Body	
FGFR2	Gastric, NSCLC, endometrial	Before/after diagnosis	Hypomethylation	Promoter	Island
MYH9	ALCL	After diagnosis	Hypomethylation	3'UTR	
MITF	Melanoma	After diagnosis	Hypomethylation	Body	Shelf
SDHB	Paraganglioma, pheochromocytoma, familial paraganglioma	After diagnosis	Hypomethylation	Promoter	Shore
RAD51L1	Lipoma, uterine leiomyoma	After diagnosis	Hypomethylation	Body	
ASPSCR1	Alveolar soft part sarcoma	After diagnosis	Hypomethylation	Body	Shore
MLL2	Medulloblastoma, renal	After diagnosis	Hypomethylation	Body	Shore
MN1	AML, meningioma	After diagnosis	Hypomethylation	Body	Shore
ETV1	Ewing sarcoma, prostate	Before/after diagnosis	Hypomethylation	Promoter	Shore
ERCC4	Skin basal cell, skin squamous cell, melanoma, Xeroderma pigmentosum (F)	Before/after and after diagnosis	Hypomethylation	Body	
CDH11	aneurysmal bone cysts	After diagnosis	Hypomethylation	Promoter	
CIITA	PMBL, Hodgkin lymphoma	After diagnosis	Hypomethylation	Body	Island
BCL3	CLL	After diagnosis	Hypomethylation	Body	Island
SFPQ	Papillary renal cell	After diagnosis	Hypomethylation	Promoter	Island
BCL6	NHL, CLL	After diagnosis	Hypermethylated	Promoter	Shore
PDE4DIP	MPD	After diagnosis	Hypermethylated	Body	Island
ATIC	ALCL	After diagnosis	Hypermethylated	Promoter	Island
MSH2	Colorectal, endometrial, ovarian, hereditary non-polyposis colorectal cancer	After diagnosis	Hypermethylated	Promoter	Island
PAX5	NHL, ALL, B-ALL	After diagnosis	Hypermethylated	Body	Island
FOXP1	ALL	After diagnosis	Hypermethylated	Body	
CBFA2T3	AML	After diagnosis	Hypermethylated	Body	Shore

bcDMP using data obtained from the 450K DNA methylation array of six healthy breast samples. Detecting a hypomethylated region spanning the entire alternative promoter including the bcDMP, we confirmed the functional importance of the identified site (Supplementary Figure S1A, available at *Carcinogenesis* Online).

A strong gain of methylation of *DOK7* could also be detected in three samples (twin pair ID: 9, 12, 24) taken prior to tumor diagnosis (mean of 4.7 years), suggesting it as potential biomarker for early diagnosis or even cancer susceptibility (Figure 3G). Moreover, hypermethylation of *DOK7* was detected 3 years before and 2 years after cancer diagnosis at equal levels in a breast cancer twin pair analyzed at both time points (Supplementary Figure S1B, available at *Carcinogenesis* Online).

To further explore their clinical relevance, we performed locus-specific pyrosequencing in an independent set of 22 matched primary breast tumor samples. Differential methylation analysis revealed a significant gain of methylation for *DOK7* in the cancer samples compared with the matched normal pairs (Wilcoxon signed rank test; $P < 0.05$; Figure 3H), suggesting a profound alteration of the blood-based marker also in primary tumors. As observed before, the significant hypermethylation extended to the entire upstream region (Supplementary Figure S1C, available at *Carcinogenesis* Online). Using a case-control analysis, we still detected significant differences for the array-based bcDMP (Mann–Whitney test; $P < 0.05$) and the entire promoter region (Mann–Whitney test; $P < 0.001$). Analyzing the results in a cancer cell line model system compared with normal primary breast tissue ($n = 5$) also revealed a gain of DNA methylation of *DOK7* in 80% (5 out of 6) of breast cancer cell lines analyzed (Figure 3I), thus presenting consistent hypermethylation profiles in three different breast cancer settings in the face of global demethylation.

To obtain an insight of the role of *DOK7* in cancer types different from breast cancer, we analyzed the DNA methylation profile of the bcDMP and flanking sites in 54 cancer cell lines and normal tissues using the Infinium HumanMethylation450 BeadChip platform. Strikingly, hypermethylation of the bcDMP and additional probes included in the shore region was detected for all cancer types, in particular for melanomas, lung and renal

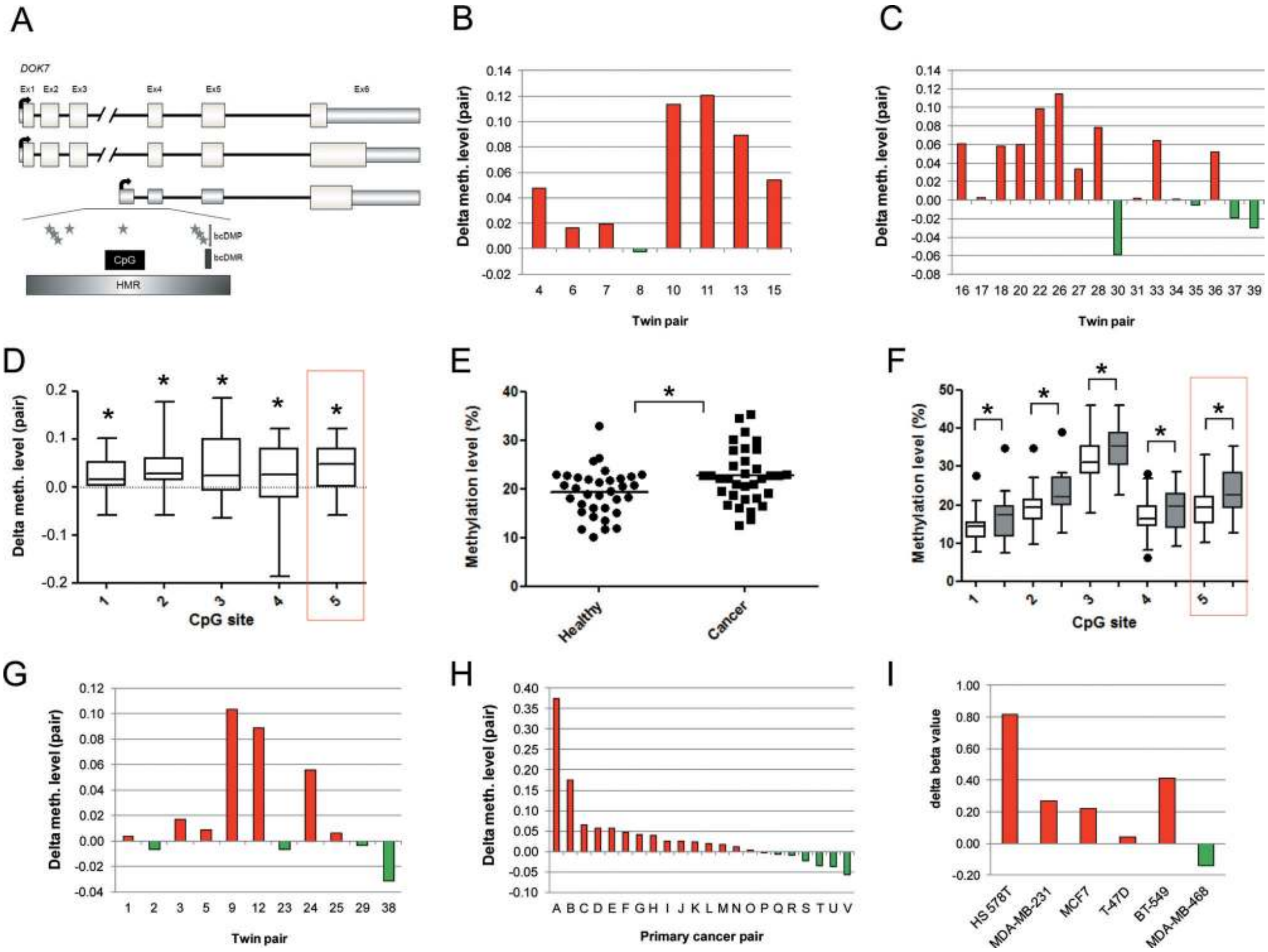

Fig. 3. *DOK7* is hypermethylated in different breast cancer contexts. (**A**) Schematic overview of the *DOK7* gene variants and associated features. Differentially methylated position (DMP; cg15652666) and associated region (DMR) are indicated. Asterisks are indicating transcription factor binding sites. (**B**) Intra-pair difference (cancer-healthy) of the *DOK7* associated bcDMP in eight twins (identification set) postdiagnosis assessed by pyrosequencing. (**C**) Intra-pair difference (cancer-healthy) of the *DOK7* associated bcDMP in 16 twins (validation set) postdiagnosis assessed by pyrosequencing. (**D**) Differences of DNA methylation (all 24 twin pairs postdiagnosis) of CpG site upstream of the bcDMP. Significant consistent differences comparing all twin pairs are indicated (**P* < 0.05). The bcDMP highlighted (red box). (**E**) Unpaired analysis of twin samples comparing healthy and breast cancer blood samples. DNA methylation data were assessed by pyrosequencing, and significance between the groups is indicated (**P* < 0.05). (**F**) Unpaired analysis of twin samples comparing CpG sites upstream of the bcDMP in healthy (white) and breast cancer (gray) blood samples. DNA methylation data were assessed by pyrosequencing, and significance between the groups is indicated (**P* < 0.05). The bcDMP highlighted (red box) and outliers (black circles) were identified by the Tukey test. (**G**) Intra-pair difference (cancer-healthy) of the *DOK7* associated bcDMP in all 11 twin pairs prediagnosis assessed by pyrosequencing. (**H**) Intra-pair difference (cancer-normal) of the *DOK7* associated bcDMP in primary breast tumor samples and matched normal control tissue assessed by pyrosequencing. (**I**) Differences in DNA methylation of six breast cancer cell line displayed relative to the median level of six normal breast samples assessed by the Infinium HumanMethylation450 BeadChip platform.

cancers (Supplementary Figure S2, available at *Carcinogenesis* Online).

Discussion

In this study, we identified DNA hypermethylation of *DOK7* to be consistently detectable in three independent sample settings: blood from twins discordant for breast cancer, primary breast tumors and breast cancer cell lines. Detecting an altered DNA methylation before diagnosis suggests a potential use of *DOK7* promoter methylation as biomarker for the early detection of breast cancer.

Biomarker identification in blood is a challenging task as blood cell-specific events cannot be entirely excluded, and methylation levels of circulating tumor DNA are able to modify the blood-specific DNA methylation profile only marginally. Therefore, alterations are expected to present changes of rather small magnitude, however consistent between cancer patients and control. Although of small magnitude, the integration of multiple epigenetic biomarkers in predictive signatures can be of high translational value. Because DNA methylation was established as crucial factor for cancer formation, it rapidly gained clinical attention as a biomarker for diagnosis and prognosis. In particular, epigenetic markers for prostate, represented by *GSTP1* among others, are close to being approved for clinical use. For sporadic breast cancer, a variety of changes in DNA methylation were detected in primary cancer samples, including the breast cancer susceptibility genes *BRCA1/2*; however, epigenetic markers in biological fluids have previously lacked the sensitivity and specificity seen in other cancer types (25–27). This might be due to the single-gene approaches and low-resolution technologies used to date, which provide limited snapshots of the genome. Here, latest base-pair resolution methylomes and high-resolution array platforms have clearly improved our knowledge of development (28–30) and diseases (21,31–33), including breast cancer (8). The other limiting factor to previous studies analyzing large cohorts of different genetic backgrounds introduces considerable noise and variation due to the interplay between the genetic variability, environmental effects and DNA methylation (10). This is a major problem particularly for the detection of blood-based biomarkers as the expected differences are small.

To improve on previous efforts, we applied the high-resolution Infinium HumanMethylation450 BeadChip platform in this study, previously confirmed to reliably detect methylation changes at about half a million CpG sites (13,14). In addition, we removed genetic noise and reduced other sources of confounding, by analyzing identical twin pairs discordant for breast cancer development. Accordingly, the genome-wide intra-twin pair DNA methylation variability was much lower than in unrelated case-control studies. We detected 403 differentially methylated sites. In line with genome-wide loss of DNA methylation occurring in breast cancer, we detected almost exclusively hypomethylated sites consistently altered between the discordant twin pairs (8). Genes harboring bcDMPs within their promoters were enriched for hallmarks of cancer as well as specific cancer-related pathways (34). Furthermore, genes previously identified to be associated with breast cancer were among the epigenetic candidate genes. Most importantly, we present a set of previously unknown potential blood-based biomarkers, with a subset even able to separate blood from healthy and cancer twins in a hierarchical cluster approach.

Moving from a whole-genome identification approach to a gene-specific validation phase, we analyzed 14 genes, showing an association to cancer or high differences between twin pairs, in more detail. In particular, differential methylation of *DOK7* was confirmed by technical and biological validation, and *HMGB3* and *MYC* revealed a consistent gain or loss of DNA methylation in the validation set, respectively, however not reaching statistical significance. CpG sites upstream of the *DOK7* associated differentially methylated CpG site also gained methylation in the cancer patients, defining it as differentially methylated region of potential functional relevance. Strikingly, the identified DMR is located in a CpG island shore at the border of a hypomethylated region, with both features previously associated to high regulatory potential (21,35,36). With its close proximity to the transcription factor binding sites of HMX1, PAX6 and CREB, it is tempting to speculate that hypermethylation prevents transcription factor binding, so contributing to gene miss-regulation. However, to this point, the functional consequences of an altered *DOK7* DNA methylation at the presented CpG sites remain elusive, and future studies have to address their direct association to gene expression and disease-related phenotype changes.

DOK7 is a docking protein that serves not only as substrate but also as activator of receptor tyrosine kinases (37). Interestingly, the identified differentially methylated CpG site is located in an alternative promoter, controlling the expression of a *DOK7* transcript variant with a truncated open reading frame. However, it might also act as non-coding RNA altering post-transcriptional regulation of the original transcript by absorbing microRNAs targeting the 3′ untranslated region (UTR) of *DOK7* (38). In this context, microRNA-145 presents the most promising candidate, as it was previously reported to be differentially expressed in primary tumors and capable of altering growth of breast cancer cells (39–41).

The potential of *DOK7* as a biomarker is also demonstrated by pre-breast cancer diagnosis blood samples displaying promoter hypermethylation, suggesting alterations of *DOK7* to be an early event in tumorigenesis. However, this has to be confirmed in larger clinical data sets. *DOK7* was not only identified as promising blood biomarker but was also confirmed to be hypermethylated in primary breast cancer specimens and cell lines of different tissue type origin, suggesting a crucial role of the alternative gene product in cancer formation.

Starting with an initial cohort of 15 MZ twin pairs discordant for breast cancer, high-resolution DNA methylation analysis determined a set of differentially methylated CpG sites including known cancer genes involved in disease-specific pathways and also novel candidates with possible implication in breast tumorigenesis. Most importantly, *DOK7* showed the most significant DNA methylation changes in blood-based and primary breast cancer settings, suggesting that it could be useful as a novel biomarker for this tumor type.

Supplementary material

Supplementary Figures 1 and 2 and Tables 1–4 can be found at http://carcin.oxfordjournals.org/

Funding

Ministerio de Ciencia e Innovación (grant SAF2011–22803 to M.E.); the European Research Council (ERC) (advanced grant EPINORC to M.E.); Cellex Foundation, the Health and Science Departments of the Catalan Government (Generalitat de Catalunya), Fondo de Investigaciones Sanitarias (grant PI08–1345 to M.E.) and Fundación Sandra Ibarra de Solidaridad frente al Cancer. M.E. is an ICREA Research Professor. T.D.S. is an European Research Council Principal Investigator. This work was also supported in part by the ERC (250157) as well as core support from the Wellcome Trust and the NIHR BRC award to Guys and St Thomas' foundation Trust and KCL. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Acknowledgments

We would like to thank all the twin volunteers and Raj Gill for collecting information.

Conflict of Interest Statement: None declared.

References

1. Esteller,M. *et al.* (2000) Promoter hypermethylation and *BRCA1* inactivation in sporadic breast and ovarian tumors. *J. Natl. Cancer Inst.*, **92**, 564–569.

2. Stefansson,O.A. *et al.* (2011) CpG island hypermethylation of *BRCA1* and loss of pRb as co-occurring events in basal/triple-negative breast cancer. *Epigenetics*, **6**, 638–649.
3. Birgisdottir,V. *et al.* (2006) Epigenetic silencing and deletion of the *BRCA1* gene in sporadic breast cancer. *Breast Cancer Res.*, **8**, R38.
4. Fong,P.C. *et al.* (2009) Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. *N. Engl. J. Med.*, **361**, 123–134.
5. Veeck,J. *et al.* (2010) *BRCA1* CpG island hypermethylation predicts sensitivity to poly(adenosine diphosphate)-ribose polymerase inhibitors. *J. Clin. Oncol.*, **28**, e563–4; author reply e565.
6. Silver,D.P. *et al.* (2010) Efficacy of neoadjuvant cisplatin in triple-negative breast cancer. *J. Clin. Oncol.*, **28**, 1145–1153.
7. Esteller,M. (2007) Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat. Rev. Genet.*, **8**, 286–298.
8. Hon,G.C. *et al.* (2012) Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. *Genome Res.,* **22***,* 246–258.
9. Esteller,M. (2008) Epigenetics in cancer. *N. Engl. J. Med.*, **358**, 1148–1159.
10. Bell,J.T. *et al.* (2011) DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol.*, **12**, R10.
11. Bell,J.T. *et al.* (2012) Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS Genetics*, **8**, e1002629.
12. Bell,J.T. *et al.* (2011) A twin approach to unraveling epigenetics. *Trends Genet.*, **27**, 116–125.
13. Sandoval,J. *et al.* (2011) Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics*, **6**, 692–702.
14. Bibikova,M. *et al.* (2011) High density DNA methylation array with single CpG site resolution. *Genomics*, **98**, 288–295.
15. Rakyan,V.K. *et al.* (2011) Identification of type 1 diabetes-associated DNA methylation variable positions that precede disease diagnosis. *PLoS Genet.*, **7**, e1002300.
16. Javierre,B.M. *et al.* (2010) Changes in the pattern of DNA methylation associate with twin discordance in systemic lupus erythematosus. *Genome Res.*, **20**, 170–179.
17. Benjamini,Y. *et al.* (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist Soc. B*, **57**, 289–300.
18. Hall, M.A. (1999) Correlation-Based Feature Subset Selection for Machine Learning. Ph.D. Thesis, Department of Computer Science, Waikato University, Waikato, New Zealand.
19. Warde-Farley,D. *et al.* (2010) The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.*, **38**, W214–W220.
20. Bindea,G. *et al.* (2009) ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*, **25**, 1091–1093.
21. Irizarry,R.A. *et al.* (2009) The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat. Genet.*, **41**, 178–186.
22. Doi,A. *et al.* (2009) Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat. Genet.*, **41**, 1350–1353.
23. Hunter,D.J. *et al.* (2007) A genome-wide association study identifies alleles in *FGFR2* associated with risk of sporadic postmenopausal breast cancer. *Nat. Genet.*, **39**, 870–874.
24. Futreal,P.A. *et al.* (2004) A census of human cancer genes. *Nat. Rev. Cancer*, **4**, 177–183
25. Ito,Y. *et al.* (2008) Somatically acquired hypomethylation of IGF2 in breast and colorectal cancer. *Hum. Mol. Genet.*, **17**, 2633–2643.
26. Korshunova,Y. *et al.* (2008) Massively parallel bisulphite pyrosequencing reveals the molecular complexity of breast cancer-associated cytosine-methylation patterns obtained from tissue and serum DNA. *Genome Res.*, **18**, 19–29.
27. Brennan,K. *et al.*; KConFab Investigators. (2012) Intragenic ATM methylation in peripheral blood DNA as a biomarker of breast cancer risk. *Cancer Res.*, **72**, 2304–2313.
28. Lister,R. *et al.* (2011) Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature*, **471**, 68–73.
29. Doi,A. *et al.* (2009) Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nature Genet.,* **41**, 1350–1353.
30. Smith,Z.D. *et al.* (2012) A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature*, **484**, 339–344.
31. Campan,M. *et al.* (2011) Genome-scale screen for DNA methylation-based detection markers for ovarian cancer. *PLoS ONE*, **6**, e28141.
32. Fernandez,A.F. *et al.* (2012) A DNA methylation fingerprint of 1628 human samples. *Genome Res.*, **22**, 407–419.
33. Hansen,K.D. *et al.* (2011) Increased methylation variation in epigenetic domains across cancer types. *Nat. Genet.*, **43**, 768–775.
34. Hanahan,D. *et al.* (2000) The hallmarks of cancer. *Cell*, **100**, 57–70.
35. Molaro,A. *et al.* (2011) Sperm methylation profiles reveal features of epigenetic inheritance and evolution in primates. *Cell*, **146**, 1029–1041.
36. Hodges,E. *et al.* (2011) Directional DNA methylation changes and complex intermediate states accompany lineage specificity in the adult hematopoietic compartment. *Mol. Cell*, **44**, 17–28.
37. Bergamin,E. *et al.* (2010) The cytoplasmic adaptor protein Dok7 activates the receptor tyrosine kinase MuSK via dimerization. *Mol. Cell*, **39**, 100–109.
38. Ebert,M.S. *et al.* (2010) Emerging roles for natural microRNA sponges. *Curr. Biol.*, **20**, R858–R861.
39. Kim,S.J. *et al.* (2011) Development of microRNA-145 for therapeutic application in breast cancer. *J. Control. Release*, **155**, 427–434.
40. Sachdeva,M. *et al.* (2010) MicroRNA-145 suppresses cell invasion and metastasis by directly targeting mucin 1. *Cancer Res.*, **70**, 378–387.
41. Iorio,M.V. *et al.* (2005) MicroRNA gene expression deregulation in human breast cancer. *Cancer Res.*, **65**, 7065–7070.

Received July 2, 2012; revised September 19, 2012; accepted October 2, 2012