Tumor and Stem Cell Biology

A Comprehensive DNA Methylation Profile of Epithelial-to-Mesenchymal Transition


Abstract

Epithelial-to-mesenchymal transition (EMT) is a plastic process in which fully differentiated epithelial cells are converted into poorly differentiated, migratory and invasive mesenchymal cells, and it has been related to the metastasis potential of tumors. This is a reversible process and cells can also eventually undergo mesenchymal-to-epithelial transition. The existence of a dynamic EMT process suggests the involvement of epigenetic shifts in the phenotype. Herein, we obtained the DNA methylomes at single-base resolution of Madin–Darby canine kidney cells undergoing EMT and translated the identified differentially methylated regions to human breast cancer cells undergoing a gain of migratory and invasive capabilities associated with the EMT phenotype. We noticed dynamic and reversible changes of DNA methylation, both on promoter sequences and gene-bodies in association with transcription regulation of EMT-related genes. Most importantly, the identified DNA methylation markers of EMT were present in primary mammary tumors in association with the epithelial or the mesenchymal phenotype of the studied breast cancer samples. Cancer Res; 74(19); 5608-19. ©2014 AACR.

Introduction

Although most human tumors are derived from epithelial cells, it has been observed that, in the natural history of the disease, an epithelial-to-mesenchymal transition (EMT) can occur (2–6). EMT implicates molecular reprogramming of the cells that progressively lose their epithelial characteristics while gaining mesenchymal features that facilitate cellular detachment from the primary tumor, invasion of neighboring stroma and entry into the blood stream (2–6). This might be a reversible process and the transformed cells could eventually undergo mesenchymal–epithelial transition (MET), then proliferate and form distant secondary tumors (2–6). The dynamic nature of EMT suggests that the remodeling of the epigenetic setting, such as DNA methylation and histone modification patterns, could be involved in the process. For histones, changes in H3 Lys9 dimethylation take place in mouse hepatocytes undergoing EMT upon transforming growth factor β (TGFβ) treatment (7). DNA methylation-associated silencing of the miR-200 family, a master regulator of the epithelial phenotype by targeting ZEB1 and ZEB2, also occurs in EMT (8–11). However, a whole-genome DNA methylation study to determine in an unbiased manner the EMT-associated DNA methylation changes has not been reported.

Herein, we have used whole-genome bisulfite sequencing (WGBS) to obtain the complete DNA methylomes at single-base resolution of Madin–Darby canine kidney (MDCK) cells (12). MDCK cells are a widely used model for studying EMT because they have clear apico-basolateral polarity and de...
microarrays and pyrosequencing, we have also validated the obtained differentially methylated regions (DMR) from MDCK cells in human breast cancer.

**Materials and Methods**

**Cell cultures**

MDCK-II cells were purchased from the European Collection of Cell Cultures (catalog no. 00062107; ECACC; Public Health England). EMT was induced by treatment with TGFβ (R&D Systems) at 5 ng/ml for 30 days. Reepithelization was achieved upon TGFβ withdrawal after 30 days. As a human EMT model, we used the human breast cancer MDA-MB-468PT and 468LN cell lines, which were kindly supplied by Ann F. Chambers (London Regional Cancer Program, London, Ontario, Canada). All cell lines were characterized by short tandem repeat analysis (STR) profiling (LGS Standards SLU) within 6 months after receipt. Cell lines were cultured in DMEM supplemented with 10% fetal bovine serum at 37°C and 5% CO2. For epigenetic drug treatments, MDCK-T and MDA-MB-468LN cells were treated with 2 μg 5-aza-2'-deoxycytidine (Sigma) for 72 hours.

**Human primary samples**

Breast cancer samples were obtained from the Department of Pathology, University Hospital, Iceland and the Germans Trias i Pujol Hospital, Barcelona, Catalonia, Spain. Normal tissues were obtained from a distal location to the tumor mass of the same patient at the time of surgery, therefore excluding a bias due to age difference. DNA was isolated by the phenol-chloroform/proteinase-K method. Patient information about the breast cancer samples came from the population-based Icelandic Cancer Registry. Data on clinical staging and histologic grade from the Icelandic Cancer Registry are based on analyses by pathologists at the Department of Pathology, Landspitali University Hospital. For clinical staging, the TNM system was followed (tumor size and nodal status), with histologic grade assessed by the Nottingham system. Informed consent was obtained from all patients and all work was carried out with the permission of the Icelandic Data Protection Commission (2006050307) and Bioethics Committee (VSNh2006050001/03-16). Association of DNA methylation levels and scoring of cytokeratin 5/6 staining analyzed by immunohistochemistry (15) was assessed by ANOVA. DNA methylation data from 586 primary breast carcinomas and 96 normal breast tissues from Illumina 450K arrays available at The Cancer Genome Atlas (TCGA: cancergenome.nih.gov/) were also analyzed.

**WGBS data processing**

We performed sequence alignment and methylation calling with Bismark software, version 0.7.4 (16). We used CanFam3.1 as a reference genome and retrieved genomic information from Biomart (17). To complement the gene annotation, we also made use of the more complete CanFam2 annotation, converting the coordinates with the liftOver University of California, Santa Cruz (UCSC) tool (18). SAM/BAM and BED files were handled with SAMtools, BEDTools, and Tabix (19, 20). We used the smoothing method (21) to depict methylation profiles. DMRs were identified by considering all homozygous CpG dinucleotides, and searching for maximal length sequences of consecutive CpGs with a consistent direction of methylation change in the two samples for all CpGs in the region, and in which the first and last CpG dinucleotides of a region containing a minimum of five CpGs yielded a value of χ² of > 3.84 [degrees of freedom (df) = 1; P < 0.05] for differential methylation of the two residues.

**DNA methylation analyses**

Global methylated-CpG content was quantified by high-performance liquid chromatography (HPLC; ref. 22). Pyrosequencing assays were designed to analyze and validate the results obtained from the WGBS data at intragenic DMRs, and to analyze DNA methylation levels in human breast cancer samples (23). Bisulfite-genomic sequencing (24) was carried out to validate DNA methylation at DMRs overlapping transcription start site-proximal regions. Specific sets of primers were designed using the Methyl Primer Express program (Applied Biosystems). The DNA methylation analysis with Infinium 450K microarray was carried out as previously described (25, 26).

**Gene expression analyses**

We hybridized RNA samples to the Affymetrix Canine Gene 1.0 ST Array (901926). The platform assesses whole-genome expression profile by providing high coverage across the entire transcript. High-capacity cDNA validation of gene expression was carried out by quantitative real-time PCR (qRT-PCR). A reverse transcription kit (Applied Biosystems) was used for reverse transcription. qRT-PCR reactions were performed in triplicate on an Applied Biosystems 7900HT Fast Real-Time PCR system. All data were normalized with respect to 18S RNA, RPL8, and RPS5 expression in dog, and to HPRT1, GAPDH, and PPIA in human.

**Comparison of human and canine CpG methylation**

To compare DNA methylation profiles of dog and human in the same reference tissue, we downloaded reduced representation bisulfite sequencing (RRBS) from adult human kidney, which was generated as part of the Roadmap Epigenomics project. The DNA methylation dataset is available at Gene Expression Omnibus (GEO) database under accession code GSM621713. All genomic tracks were retrieved from the UCSC. For genomic tracks unavailable in the dog (CanFam3.1) genome, human tracks were used to convert the coordinates. For all features conserved at genomic level that were covered by both datasets, we compared DNA methylation levels between human and canine samples. We defined a CpG as having conserved CpG methylation when variation between human and dog was not greater than 33%. We obtained human histone marks for human adult kidney, available at the ENCODE project (GSE19465) and CCAT peak caller (27). All genomic interval handling was performed using BEDTools (28). All human (hg19)–dog (CanFam3.1) coordinates were converted with the UCSC liftOver tool.
Accession code

WGBS and Affymetrix gene expression data have been deposited in GEO repository under accession numbers GSE48527 and GSE56763.

Results

Complete DNA methylomes of MDCK cells undergoing EMT

For WGBS, we generated 767M and 884M raw reads, resulting in 76.7 (MDCK-U) and 88.4 (MDCK-T) raw Gb of paired-end sequence data. Of these, 60.5 (MDCK-U) and 62.1 (MDCK-T) Gb (78.9% and 70.2%, respectively) were successfully aligned to either strand of the reference genome (CanFam 3.1), providing an average 26.0-fold (MDCK-U) and 26.7-fold (MDCK-T) sequencing depth. Of all the cytosines present in the reference genome sequence, 91.3% (MDCK-U) and 88.3% (MDCK-T) of Cs and 93.1% (MDCK-U) and 90.1% (MDCK-T) of CGs were covered. On the basis of alignment with in silico converted non-CpG cytosines, the bisulfite conversion rate was determined to be at least 99.8% (MDCK-U) and 98.5% (MDCK-T), even when assuming that all non-CpG methylcytosines were derived from conversion failure; this ensured reliable ascertainment of CpG methylcytosines with a false-positive rate of <0.5%. The full set of WGBS data from MDCK-U and MDCK-T, obtained using Circos (29), is illustrated in Fig. 1A. The WGBS obtained DNA methylome of the adult canine kidney (MDCK-U) show a high degree of conservation with the DNA methylation profile of adult human kidney (GSM621713) derived from RBBs (Kendall rank correlation test, P < 2.2e–16; Fig. 1B; ref. 30). The WGBS data are freely available for download from NCBI GEO: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=tdubbergoswouzu&acc=GSE48527.

In relation to the global DNA methylation changes in the EMT model, we observed a similar number of methylated CpG dinucleotides (mCpG) in MDCK-U and MDCK-T cells: 16,499,555 versus 16,504,556 on the Watson and Crick strand, respectively (Fig. 1C). HPLC analyses also confirmed the similar 5-methylcytosine DNA content in MDCK-U and MDCK-T cells (Fig. 1D). These results are consistent with the overall nonsignificant differences in methylated CpG sites obtained using a different technology (comprehensive high-throughput arrays for relative methylation) in a mouse EMT model (7). However, the observed absence of massive DNA methylation changes in the MDCK model upon TGFβ treatment did not rule out the possible CpG methylation–mediated reprogramming of these cells to engage in an EMT program: it is entirely feasible that there is a reshuffling of methylated CpG sites between different genomic loci entailing important functional consequences, but the DNA methylation gains and losses compensate each other and, thus, do not alter the global DNA methylation content. Alternatively, the DNA methylation changes can take place in a minor fraction of the studied genomes but targeting genes with a major impact in the phenotype of these cells. To address these issues, we examined for specific DMRs between MDCK-U and MDCK-T cells.

Identification of DMRs in MDCK cells undergoing EMT.

DMRs were identified by considering all homozygous CpG dinucleotides, and searching for maximal length sequences of consecutive CpGs with a consistent change in direction of methylation in the two samples for all CpGs in the region, and for which there was significantly different methylation ($X^2$ test: $P < 0.05$) between the first and last CpG dinucleotide residues of a region containing a minimum of five CpGs (31). Adopting these criteria, we identified 87,296 DMRs between MDCK-U and MDCK-T (17); these are illustrated using Circos (29) in Fig. 2A, and listed in detail in Supplementary Table S1. These DMRs represented 679,699 CpG sites in the reference canine genome. They comprise 2.8% of the total number of CpGs (23,944,073) distributed throughout all dog chromosomes (Supplementary Table S1). The most common DMR change was the presence of an unmethylated sequence in MDCK-U that became methylated in MDCK-T: 52,400 hypermethylated DMRs (representing 441,010 CpGs comprising 1.8% of those in the reference canine genome) versus 34,896 hypomethylated DMRs (238,689 CpGs, representing 1% of the reference genome; Fig. 2B). Thus, the DNA methylene of the MDCK cells underwent selective CpG site methylation changes in discrete regions of the genome upon induction of the EMT phenotype.

We examined in which genomic compartments these DMRs were located. DMRs were identified in intergenic (59%, 51,475 DMRs) and intragenic (41% 35,821 DMRs) regions (Fig. 2B). Among the intergenic loci, nonrepetitive DNA sequences accounted for 18,008 DMRs, whereas DNA repetitive sequences harbored 33,467 DMRs. Among these latter sequences, the most common DMR-associated intergenic sequences were LINE (17,125), SINE (10,407), and LTR (2,826) repeats, whereas the remaining 3,109 corresponded to other types of repetitive sequences (Fig. 2B). For intragenic sequences, the highest frequency of DMRs was observed in gene bodies, particularly in intronic sequences (23,541), but also in 4,390 exonic sequences. DMRs were also found in gene promoters (7,890), associated with a CpG island (5,267) or a CpG-poor sequence (2,623). All the studied genomic compartments presented DMRs in a balanced manner between hypermethylated and hypomethylated events with one significant exception: promoter CpG islands preferentially showed hypermethylation events upon TGFβ-induced EMT (two-tailed Fisher exact test: $P < 0.0001$; Fig. 2B), as would be expected under the conventional assumption that these regions are usually unmethylated in normal tissues (32, 33).

Characterization of candidate genes associated with DMRs in the MDCK-EMT model.

To further identify those DMRs that target genes with functional relevance in the canine EMT model, we complemented the WGBS data with the results of a dog expression microarray experiment. Using a platform containing 590,097 probes for 27,681 genes, we found that 1,400 annotated genes had a more than 2-fold differential expression between MDCK-U and MDCK-T cells (Supplementary Table S2). We then crossed the RNA data with the identified intragenic DMRs to characterize those distinctly methylated sequences that...
influenced gene expression. Overall, 465 genes underwent significant differential expression in our canine EMT system in association with an intragenic DMR [χ² test; OR, 1.82; 95% confidence interval (CI), 2.6–3.1; P < 1e–15; Supplementary Table S3]. Interestingly, more DMR-linked expression changes occurred in gene body (exonic/intronic) sequences (n = 324) than in 5'-end minimal promoters (n = 141) of the corresponding genes. These results highlight recent findings that showed how changes in nonpromoter intragenic methylation also affect the regulation of gene expression (34–36).

We characterized the physiologic relevance of two identified candidate genes undergoing the aforementioned gene body DMR-linked expression change in our MDCk–TGFβ model: the EMT inducer transcription factor 4 (TCF4; ref. 37) and the EMT repressor grainyhead-like-2 (GRHL2; ref. 38; Fig. 3). Using WGBS (Fig. 3A) and pyrosequencing (Fig. 3A), we identified 178 DMRs in the TCF4 gene body encompassing a region from intron 1 to exon 20 (3,352 CpGs) that undergo hypermethylation in MDCk-T relative to MDCk-U cells. Most importantly, the hypermethylation event in the TCF4 exon–intron sequences was associated with upregulation of the transcript, as determined by the expression microarray (Supplementary Table S2 and Fig. 3B) and qRT-PCR (Fig. 3B). The opposite scenario was observed for GRHL2. We found 26 DMRs in the GRHL2 gene body defining a region from exon 1 to intron 15 (a total of 726 CpG sites) that were hypomethylated in MDCk-T relative to MDCk-U cells analyzed by WGBS (Fig. 3C) and pyrosequencing (Fig. 3C). Most importantly, the hypomethylation event in the GRHL2 exon–intron sequences was associated with downregulation of the transcript determined by the expression microarray (Supplementary Table S2 and Fig. 3D) and qRT-PCR (Fig. 3D). These are dynamic processes and, most notably for both genes, when MDCk-T cells underwent a change in a MET phenotype after we withdrew TGFβ from the medium (MDCk-W cells), the cells recovered the TCF4 (Fig. 3A and B) and GRHL2 (Fig. 3C and D) DNA methylation and expression patterns observed in MDCk-U. Additional examples where we found consistent transcriptional silencing or activation upon gene-body hypomethylation or hypermethylation, respectively, determined by WGBS in the described EMT model involved genes related to cell adhesion such as LAMA4 and ANXA10, extracellular matrix remodeling such as LRIG1 or SPARC, cytoskeleton reorganization such as PALLD or TNIK, or invasive properties such as ST8SIA1, TRPS1, or PTGER4.
among others (Supplementary Fig. S1). These results support that DNA methylation shifts in the gene bodies are associated with the opposite expression patterns to those widely reported for 5'-end gene promoters.

Related to the most studied DNA methylation events, those affecting the CpGs surrounding the corresponding transcription start sites, we also validated and characterized four top-ranked genes derived from the promoter-associated DMRs identified here (Supplementary Table S1). To establish their functional relevance in the EMT phenotype and metastasis development, we have further determined the DNA methylation and expression status of TINAGL1 (39), ESYT3 (40), ITGA5 (41), and FKBP10 (42). TINAGL1 and ESYT3 5'-end promoters undergo hypermethylation in MDCK-T in comparison with MDCK-U cells determined by WGBS (Fig. 4A) and pyrosequencing (Fig. 4A). The hypermethylation events in both promoters were associated with downregulation of the transcript determined by the expression microarray (Supplementary Table S2 and Fig. 4B) and qRT-PCR (Fig. 4B). Treatment of the MDCK-T cells with the DNA demethylating agent or induction of MET by TGFβ withdrawal restored TINAGL1 and ESYT3 expression (Fig. 4B). ITGA5 and FKBP10 promoters showed the other side of the coin: both 5'-end regulatory regions became hypomethylated in MDCK-T relative to MDCK-U, as shown by WGBS (Fig. 4C) and pyrosequencing (Fig. 4C). The hypomethylation events in the ITGA5 and FKBP10 promoters were associated with upregulation of the transcripts according to the expression microarray (Supplementary Table S2 and Fig. 4D) and qRT-PCR (Fig. 4D). Thus, these DNA methylation events occurring in the CpGs encompassing the corresponding transcription

![Figure 2. Genome-wide DMRs. A, Circos representation of DMR distribution among canine chromosomes (inner-most circle in gray) for MDCK-U (inner circle) and MDCK-T (outer circle) cells. Average methylation levels in all of the DMRs in 504 5-Mbp-wide windows are expressed as β-values (0–1) and are graded from white (unmethylated) to blue (methylated). B, DMR distribution among the different genomic compartments. Pie-charts illustrate the proportion of hypermethylated (red) and hypomethylated (green) DMRs per genomic feature.](image-url)
start sites adhere to the customarily observed association between the presence of 5-methylcytosine and gene silencing.

To explore the functional implications of DNA methylation changes during EMT, we have performed a functional enrichment analysis of genes differentially expressed in association with differential DNA methylation, both at gene promoters and bodies, using the Bioconductor GOstats package on Biological Process category under R statistical software. Interestingly, upregulated terms are enriched in numerous signaling pathways, including developmental processes, wound-healing, and regulation of cell differentiation. This highly resembles the signaling cascades that are activated during developmental EMT, which ultimately controls the morphogenetic movements and specification patterns in this context. In turn, downregulated pathways are frequently related with cell adhesion, catabolism, and protein transport, which relates to the gain of invasive properties and indicates a metabolic reprogramming associated with EMT (Supplementary Table S4).

Translation of identified DMRs to human breast cancer cells undergoing EMT

We also considered the degree of conservation in humans of the observed epigenetic remodeling patterns that MDCK cells undergo to achieve EMT. To this end, we used an available human model to analyze the EMT process occurring in metastasis, the paired breast cancer cell lines 468PT and 468LN (43). The 468PT cell line was generated by stable transfection of the human breast adenocarcinoma cell line MDA-MB-468 with the pEGFP-C2 expression vector, whereas 468LN was derived from lung metastases dissected from a nude mouse following orthotopic injection with 468PT cells (43). In contrast with the epithelioid phenotype of the parental 468PT cell line, which shows a polygonal shape and forms epithelial structures in confluence, 468LN cells are elongated and spindle shape with a scant cytoplasm, such morphologic changes being associated with EMT (43). We have previously confirmed the epithelial phenotype of 468PT cells by determining a high expression of the microRNA-200 family and the E-cadherin together with a
low expression of ZEB1 and vimentin (11). On the other side, MDA-MB-468LN cells display EMT features such as down-regulation of the microRNA-200 family and the E-cadherin, added to a high expression of ZEB1 and vimentin (11). Most important, the 468LN cells produce extensive spontaneous lymph node metastases following orthotopic injection in nude mice, in contrast with the poorly metastatic MDA-MB-468 (11, 43, 44).

We used WGBS in 468PT and 468LN cells that resulted in approximately 600 million reads per sample and a genome-wide coverage of 15.5°C2 and 14.3°C2, respectively. The WGBS data are available for download from NCBI GEO: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=utsjuogozovmvt&acc=GSE56763. From the identified 87,296 DMRs between MDCK-U and MDCK-T cells, 51,603 (59.1%) DMR sequences were conserved in humans (Supplementary Fig. S2A). From these sequences, 25,092 (48.6%) DMRs underwent the same DNA methylation shift as in the canine model. To be able to assess these cross-species DMRs in an affordable manner in clinical samples, we used the 450K Infinium Human DNA methylation BeadChip microarray (25, 26). Using this platform in 468PT and 468LN cells, we were able to analyze DNA methylation levels of 4,446 canine DMRs, which presented conserved CpGs sequences interrogated on the human platform. Importantly, we observed significant consistency of differential DNA methylation between the canine EMT
model and the human metastatic breast cancer context (OR, 1.65; 95% CI, 1.44–1.89; Fisher P value < 0.31e–13). In particular, we detected 52% of EMT-related DMRs (2,292 out of 4,446) with the same DNA methylation changes in the dog and human contexts (Supplementary Table S5). From the fraction of DMR-associated genes in the canine model with linked expression changes (120 of 465), 49.2% were similarly conserved in humans. We further performed an enrichment analysis of the pathways affected, and found high frequency of terms related with cell motility and extracellular matrix organization, suggesting an active role of DNA methylation changes in driving expression changes in MDCK cells (TINAGL1, ESYT3, ITGA5, FKBP10) in human EMT. Consistent with the findings from the MDCK model, we observed that TINAGL1 and ESYT3 promoters were unmethylated in the epithelial breast cancer cell line 468PT, as revealed by a 450K DNA methylation microarray (Fig. 5A) and bisulfite genomic sequencing (Fig. 5A), in association with a high level of expression of the transcripts (Fig. 5B). However, the spindle-shaped metastatic 468LN cells underwent hypermethylation of both promoters (Fig. 5A), in association with a low level of expression of the transcripts (Fig. 5D). The use of 5-aza-2-deoxycytidine in 468PT cells recovered the expression of both transcripts (Fig. 5B). The observed expression patterns of TINAGL1, ESYT3, ITGA5, and FKBP10 were confirmed in a previously deposited microarray dataset (43).

The translation of DNA methylation markers of EMT from the MDCK model to the human EMT breast cancer system prompted us to search for them in primary mammary tumors. Using the 450K DNA methylation microarray in 53 breast adenocarcinomas, the conserved DMRs identified from the canine and human models were particularly clustered in the basal-subtype of breast cancer (Supplementary Fig. S2B) that has been previously associated with EMT features (45). Among all primary breast tumors, we detected the existence of promoter hypermethylation of ITGA5 (63%), TINAGL1 (33%), ESYT3 (13%), and FKBP10 (13%) in these cases (Fig. 6A). Analyzing the DNA methylation status of these promoters in the breast cancer collection available from the TCGA (cancer-genome.nih.gov; n = 679) showed similar rates of hypermethylation for these four genes (Fig. 6A). The analyses of the ITGA5, TINAGL1, ESYT3, and FKBP10 promoters in normal breast

![Figure 5](https://example.com/figure5.png)

**Figure 5.** Translation to a breast cancer EMT model. A and C, promoter DNA methylation changes were quantified in 468PT and 468LN cells by Infinium 450K DNA methylation arrays (top) and bisulfite genomic sequencing (bottom) for hypermethylated TINAGL1 and ESYT3 (A), and hypomethylated ITGA5 and FKBP10 (C). For the 450K methylation microarray, DNA methylation values were scaled between 0 (unmethylated, green) and 1 (methylated, red). For bisulfite genomic sequencing, presence of unmethylated or methylated CpGs is indicated by white or black squares, respectively. CpG dinucleotides are represented as vertical black lines. B and D, basal gene expression was measured by qRT-PCR experiments, as well as upon treatment with the demethylating agent. Error bars, SD. Asterisks represent CpG sites included in the pyrosequencing assay. TSS, transcription start site.
Promoter hypermethylation frequency (%)

B

CK 5/6 positive

CK 5/6 negative

C

VIMENTIN expression (103)

ITGA5 unmethylated

ITGA5 methylated

SNAI2 expression (103)

CDH1 expression (104)

D

ESYT3

Unmethylated (n = 44)

Methylated (n = 8)

Unmethylated (n = 45)

Methylated (n = 16)

P = 6.3 × 10−11

P = 1.9 × 10−15

P = 3.5 × 10−7

P = 1.6 × 10−16

P = 6.3 × 10−11

P = 0.047

HR, 2.76; 95% CI, 1.01–7.89

Carmona et al.

Cancer Res; 74(19) October 1, 2014

Cancer Research

5616

Published OnlineFirst August 8, 2014; DOI: 10.1158/0008-5472.CAN-13-3659
tissues (n = 20) revealed all samples to be unmethylated (Fig. 6A). The question then arose about whether the observed DNA methylation profiles more closely matched the epithelial or the mesenchymal phenotype of the studied tumors. We addressed this by examining the most frequently methylated gene in our study of primary breast malignancies (ITGA5) and taking advantage of a tissue microarray (TMA) developed in the same samples that included the EMT marker of basal-like breast tumor cytokeratin 5/6 (CK5/6; refs. 15, 46). In this context, we observed that ITGA5 promoter methylation was associated with a lower level of CK5/6 expression (Pearson test, P = 6.3e−11; Fig. 6B), in concordance with the MDCK and MDA-MB-468 results in which ITGA5 epigenetic silencing was also found in the samples with an epithelial phenotype (MDCK-U and MDA-MB-468PT, respectively). We then analyzed the ITGA5 expression levels in the breast cancer TCGA in relation to the expression of additional epithelial and mesenchymal markers. We confirmed that the downregulation of ITGA5 was associated with the loss of expression of mesenchymal markers such as vimentin (Pearson test, P = 2.2e−16; R² = 0.4), ZEB1 (Pearson test, P = 2.2e−16; R² = 0.33), and SNAI2 (Pearson test, P = 2.2e−16; R² = 0.34) and the upregulation of the epithelial marker E-cadherin (Pearson test, P = 0.05; R² = 0.06). The association between promoter methylation and loss of the transcript was also confirmed in these TCGA breast cancer samples (ANOVA, P = 0.00025). Linking the two events, we observed that ITGA5 promoter methylation in the TCGA breast cancer collection was associated with higher levels of expression of the epithelial marker E-cadherin (ANOVA, P = 1.6e−6) and diminished expression of the mesenchymal markers vimentin (ANOVA, P = 1.9e−15), ZEB1 (ANOVA, P = 6.3e−11), and SNAI2 (ANOVA, P = 3.5e−7; Fig. 6C). Finally, because EMT features have been linked with the metastatic capacity of carcinomas (2–6), we wondered whether the observed DNA methylation patterns had any clinical effect on patients with breast cancer. Although we found no obvious effect of ITGA5, TINAGL1, or FKBP10 promoter methylation status on clinicopathologic values, we did find that ESYT3 hypermethylation (linked as described above with the mesenchymal features) was a consistent prognostic factor. Specifically, we found that gain of methylation at ESYT3 promoter region was associated with shorter overall survival in the patients analyzed on the discovery cohort [hazard ratio (HR), 2.76; 95% CI, 1.01–7.89; log-rank test, P = 0.047], as well as with cancer-specific survival on an independent validation cohort (n = 55; HR, 5.64; 95% CI, 1.40–22.7; log-rank test, P = 0.006; Fig. 6D and Supplementary Fig. S2C). In addition to four candidate genes studied in detail, we reanalyzed by multivariate log-rank the latent developmental process of EMT. Even more relevant, we confirmed the occurrence of CpG hypermethylation in TINAGL1, ESYT3, ITGA5, and FKBP10 in human primary breast tumors. Our study supports previous findings on the role of gene-body methylation in modulating gene activity, and uncovers an unprecedented model to deeply characterize its role on genome regulation. Moreover, our data provide an explanation for the mechanisms regulating the expression of genes previously implicated in EMT and/or metastasis. For example, a high expression of TINAGL1 has been previously associated with better distant metastasis-free survival, and stable knockdown of TINAGL1 in the murine breast cancer cell line 4T07 increased metastatic potential after intravenous injection (39). About ITGA5, previous studies have shown that it is not expressed in normal differentiated epithelia, but become expressed in certain “emergency” situations, such as wounding or tumor development and it is centrally implicated in EMT induction (47). Moreover, it has been described that highly invasive cancer cells expressed higher levels of ITGA5 than did poorly invasive cells; and overexpression of ITGA5 enhanced the metastasis capacity (48–50). For FKBP10 and ESYT3, there are no previous studies showing an association with EMT or metastasis, but considering the conservation between canine and other potential prognostic markers. Using this approach, we found that promoter DNA hypermethylation of TMEM59L was also significantly associated with shorter overall survival in the discovery cohort (HR, 2.17; 95% CI, 1.04–4.51; log-rank test, P = 0.034); thus, additional biomarkers for further studies could be found by alternative data mining of the provided DNA methylation profiles. Overall, these last results pave the way for further development and validation of the identified epigenetic EMT markers in the clinical setting.

Discussion

Increasing evidences showing the relevance of EMT and MET in the invasion-metastasis cascade has boosted the research in these processes. Plasticity and reversibility make epigenetic mechanisms an excellent strategy to control these dynamic processes. We take advantage of a classic model of EMT and MET using kidney epithelial canine cells (MDCK) to demonstrate the role of DNA methylation in modulating expression of crucial genes with major impact in the described phenotypic and functional transitions. Furthermore, we observed a significant consistency of differential DNA methylation between the canine and the human EMT model (the paired cell lines 468PT and 468LN), reinforcing the role of the epigenetic mechanisms in the aberrant reactivation of the latent developmental process of EMT. Even more relevant, we confirmed the occurrence of CpG hypermethylation in TINAGL1, ESYT3, ITGA5, and FKBP10 in human primary breast tumors.

Figure 6. DNA methylation changes are correlated with EMT features in vivo. A, promoter DNA methylation frequency measured in our primary breast tumor cohort (PEBC) and in the TCGA breast primary tumor dataset. B, illustrative examples of cytokeratin 5/6 positive (top) and negative (bottom) staining in primary breast tumors. C, correlation between ITGA5 promoter DNA methylation and the expression of mesenchymal (VM, ZEB1, and SNAI2) and epithelial (CDH1) markers in the TCGA primary breast tumor cohort. D, Kaplan–Meier analysis of overall survival in the discovery cohort (top; n = 53) and cancer-specific survival in the validation (bottom; n = 55) cohort according to ESYT3 methylation status. P-values are those from the log-rank test; values of P < 0.05 were considered significant. HRs with a 95% CI were derived from a univariate Cox regression. Ui, unmethylated; M, methylated.
and human EMT models, as well as the results in primary breast tumors, future studies are warranted. FKB10 encodes for FKBP65 that acts as a molecular chaperone and binding partner of type I collagen (51). Interestingly, the only adult mouse tissues that continue to express FKBP65 are reproductive tissues, such as mammary glands, during phases of growth and remodeling (52). These data provide crucial clues about a putative role of FKB10 in EMT, considering the relevance of this process during development, tissue remodeling, and wound healing.

Overall, our results indicate that the EMT-associated plasticity of DNA methylation observed in the in vitro MDCK and MDA-MB-468 models also takes place in primary human tissues undergoing similar phenotypic changes. The occurrence of common DNA methylation switches associated with the acquisition of EMT features in canine and human cells supports the existence of epigenetically conserved mechanisms that contribute to cellular transformation and tumoral progression.

Disclosure of Potential Conflicts of Interest

L.G. Gut is the Director of Centro Nacional de Análisis Genómico (CNAG). No potential conflicts of interest were disclosed by the other authors.

References


DNA Methylation and Epithelial-to-Mesenchymal Transition


A Comprehensive DNA Methylation Profile of Epithelial-to-Mesenchymal Transition

F. Javier Carmona, Veronica Davalos, Enrique Vidal, et al.


Updated version
Access the most recent version of this article at:
doi:10.1158/0008-5472.CAN-13-3659

Supplementary Material
Access the most recent supplemental material at:
http://cancerres.aacrjournals.org/content/suppl/2014/08/08/0008-5472.CAN-13-3659.DC1

Cited articles
This article cites 52 articles, 18 of which you can access for free at:
http://cancerres.aacrjournals.org/content/74/19/5608.full.html#ref-list-1

Citing articles
This article has been cited by 2 HighWire-hosted articles. Access the articles at:
/content/74/19/5608.full.html#related-urls

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.