The FASEB Journal article fj.12-209502. Published online August 28, 2012.

Genome-wide methylation screen in low-grade breast cancer identifies novel epigenetically altered genes as potential biomarkers for tumor diagnosis

Marta Faryna,*[,1] Carolin Konermann,*[,‡,1] Sebastian Aulmann,[‡] Justo Lorenzo Bermejo,[§,1] Markus Brugger,[§] Sven Diederichs,[†,‡] Joachim Rom,[‖] Dieter Weichenhan,* Rainer Claus,*[,#] Michael Rehli,[¶] Peter Schirmacher,[‡] Hans-Peter Sinn,[‡] Christoph Plass,* and Clarissa Gerhauser*[,2]

*Division of Epigenomics and Cancer Risk Factors and [†]Helmholtz University Group Molecular RNA Biology and Cancer, German Cancer Research Center (DKFZ), Heidelberg, Germany; [‡]Institute of Pathology, [§]Institute of Medical Biometry and Informatics, and [‖]Department of Gynecology and Obstetrics, University Hospital Heidelberg, Heidelberg, Germany; [#]Department of Medicine, Division of Hematology nd Oncology, University of Freiburg Medical Center, Freiburg Germany; and [¶]Department of Hematology and Oncology, University Hospital Regensburg, Regensburg, Germany

ABSTRACT **Aberrant DNA methylation constitutes a well-established epigenetic marker for breast cancer. Changes in methylation early in cancer development may be clinically relevant for cancer detection and prognosis-based therapeutic decisions. In the present study, a combination of methyl-CpG immunoprecipitation (MCIp) and human CpG island (CGI) arrays was applied to compare genome-wide DNA methylation profiles in 10 low-grade *in situ* and invasive breast cancers against 10 normal breast samples. In total, 214 CGIs were found to be hypermethylated in ≥6 of 10 tumors. Functional term enrichment analyses revealed an overrepresentation of homeobox genes and genes involved in transcription and regulation of transcription. Significant hypermethylation of 11 selected genes in tumor *vs.* normal tissue was validated in two independent sample sets (45 tumors and 11 controls, 43 tumors and 8 controls) using quantitative EpiTyper technology. In tumors, median methylation levels of *BCAN*, *HOXD1*, *KCTD8*, *KLF11*, *NXPH1*, *POU4F1*, *SIM1*, and *TCF7L1* were ≥30% higher than in normal samples, representing potential biomarkers for tumor diagnosis. Using the 90th percentile of methylation levels in normal tissue as cutoff value, 62–92% of *in situ* samples (n=13), 72–97% of invasive samples from the first validation set (n=32), and 86–100% of invasive samples from the second validation set (n=43) were classified as hypermethylated. Hypermethylation of *KLF11* and *SIM1* might also be associated with increased risk of developing metastases. In summary, early methylation changes are frequent in the low-grade pathway of breast cancer and may be useful in the development of differential diagnostic and possibly also prognostic markers.—Faryna, M., Konermann, C., Aulmann, S., Bermejo, J. L., Brugger, M., Diederichs, S., Rom, J., Weichenhan, D., Claus, R., Rehli, M., Schirmacher, P., Sinn, H.-P., Plass, C., Gerhauser, C. Genome-wide methylation screen in low-grade breast cancer identifies novel epigenetically altered genes as potential biomarkers for tumor diagnosis. *FASEB J.* 26, 000–000 (2012). www.fasebj.org**

Key Words: estrogen receptor-positive · EpiTyper MassArray · CpG island array · methyl-CpG immunoprecipitation · DMR · ductal carcinoma in situ

Abbreviations: ACTB, β-actin; AUC, area under curve; BCAN, brevican; CGI, CpG island; CI, confidence interval; DAC, 2′-deoxy-5-azacytidine (decitabine); DCIS, ductal carcinoma *in situ*; DMR, differentially methylated region; DMSO, dimethyl sulfoxide; ER, estrogen receptor; ESC, embryonic stem cell; FBS, fetal bovine serum; FFPE, formalin-fixed paraffin-embedded; gDNA, genomic DNA; HOXD1, homoeobox D1; HR, hazard ratio; IDC, invasive ductal carcinoma; ILC, invasive lobular carcinoma; KCTD8, potassium channel tetramerization domain containing 8; KLF11, Kruppel-like factor 11; LCIS, lobular carcinoma *in situ*; MCIp, methyl-CpG immunoprecipitation; NXPH1, neurexophilin 1; PCDH10, protocadherin 10; PcG, polycomb group; POU4F1, POU class 4 homeobox 1; PR, progesterone receptor; ROC, receiver operating characteristic; qPCR, quantitative PCR; RYR2, ryanodine receptor 2; SIM1, single-minded homolog 1; TAC1, tachykinin precursor 1; TCF7L1, transcription factor 7-like 1; TMA, tissue microarray; TUB, tubular carcinoma; TUB-LOB, tubulolobular carcinoma

With over 1.38 million diagnosed cases and nearly 460,000 cancer deaths in 2008, breast cancer is the most common cancer type of women worldwide (1). As for many other malignancies, the key factor to a successful

[1] These authors contributed equally to this work.

[2] Correspondence: German Cancer Research Center (DKFZ), Division of Epigenomics and Cancer Risk Factors, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany. E-mail: c.gerhauser@dkfz-heidelberg.de

doi: 10.1096/fj.12-209502

This article includes supplemental data. Please visit *http://www.fasebj.org* to obtain this information.

therapy in breast cancer is early detection, ideally at a preinvasive stage. Ductal carcinoma *in situ* (DCIS) and lobular carcinoma *in situ* (LCIS) are examples of such premalignant lesions. Whereas patients with metastatic breast cancer have an age-standardized 5-yr survival rate of only 21%, detection at an earlier stage increases the rate to 80% for localized invasive cancer and to 98% for *in situ* carcinoma (2). Breast cancer is currently classified into multiple histological phenotypes; the most frequent is invasive ductal carcinoma (IDC; 50–80% of all cases), invasive lobular carcinoma (ILC; 5–15%), tubular carcinoma (TUB, <2%), and others (3). The heterogeneity of breast cancer contributes to the challenge of defining prognosis and adjusting therapy. Expression levels of estrogen receptor (ER) and progesterone receptor (PR) positively correlate with disease-free and overall survival and provide prognostic information (4). Also, gene expression profiling has resulted in the definition of specific molecular signatures associated with disease outcome (5, 6), but predicting tumor progression in an individual case is still problematic. To improve early recognition and treatment of breast cancer, new markers for detection and prognosis need to be identified.

Disease-associated epigenetic modifications such as altered histone modification patterns and aberrant DNA methylation hold promise for developing such markers. Over the past 2 decades, numerous studies have highlighted the contribution of epigenetic alterations to the malignant phenotype in various cancer types. In particular, hypermethylation of promoter CpG islands is a well-established mark for transcriptional inactivation and is often associated with downregulation of tumor suppressor genes, such as *p16*, *BRCA1*, *E-cadherin*, *DAPK1*, and *APC* (7–9). This type of aberrant DNA methylation is, therefore, recognized as one of the key events leading to tumor initiation and progression (10). In addition, due to the early occurrence and stability of DNA methylation, it has a great potential to become a marker for early cancer detection and prognosis.

Several previous methylation studies have analyzed candidate genes for tumor-specific hypermethylation in breast cancer that are involved in cell cycle control, apoptosis, and DNA repair, including *RASSF1A*, *HIN1*, and *Cyclin D2* (reviewed in ref. 11). Starting from the study of Ordway *et al.* (12), the introduction of genome-wide methylation screening methods to breast cancer research has provided a broader overview of the breast cancer methylation landscape. Recent genome-wide breast cancer methylation studies (summary in Supplemental Table S1, reviewed in refs. 13, 14) have demonstrated that aberrant DNA methylation of specific loci is already well established at the DCIS stage and is further maintained in invasive cancers and metastases (15, 16). Moreover, distinct DNA methylation patterns have been associated with hormone receptor status (16–19), expression subtypes (20–23), mutation status (24), and disease progression (16, 18, 19, 23, 25). So far, studies investigating association of methylation with metastatic behavior have reported differing results (18, 19, 25). Most of these studies were based on Illumina Infinium HumanMethylation27 and Illumina Golden Gate Cancer Panel I methylation measurements (Illumina, Inc., San Diego, CA, USA), which interrogated on average 2 CpG loci for a selected number of gene promoters (14,475 promoters for Illumina 27k and 807 promoters for Illumina Golden Gate).

In the present study, we have applied a multistage strategy to identify novel methylation biomarkers in low-grade breast cancer, especially in ER-positive and/or PR-positive cases. In the screening stage, we have combined methyl-CpG immunoprecipitation (MCIp) with human CpG island (CGI) arrays (26, 27). This approach has allowed not only identification of hypermethylated genes but also definition of areas within their promoters that display the most prominent methylation changes compared to normal breast tissue. Subsequently, we have validated and quantified aberrant methylation in two independent sample sets containing various histological subtypes and grades of ER- and/or PR-positive breast cancer using EpiTyper MassArray technology and have correlated methylation levels with clinical parameters.

MATERIALS AND METHODS

Patient samples and cell lines

Breast tumor tissue and control samples were obtained from the Institute of Pathology of the University Hospital, Tissue Bank of the National Center of Tumor Diseases (Heidelberg, Germany). All patients signed a written consent and authorization for use of biological specimens. Sample collection was approved by the ethics committee of the University of Heidelberg. The screening sample set consisted of 2 low-grade preinvasive lesions, 8 low-grade invasive tumors, and 10 normal breast tissue samples (one of them from an affected patient). Validation sample sets 1 (45 tumors and 11 controls) and 2 (43 tumors and 8 controls) were derived from macrodissected tissues. Clinical and pathological data are summarized in **Table 1**. Normal breast tissue samples were obtained from morphologically normal tissue located ≥3 cm away from the tumor (screening set and validation set 1) or from reduction mammoplasty tissue (validation set 2) provided by the Department of Gynecology and Obstetrics, University Hospital Heidelberg.

The MCF7 cell line was provided by the cell line repository of the German Cancer Research Center (H. Löhrke, DKFZ Tumorbank, Heidelberg, Germany) and was maintained in Dulbecco's modified Eagle medium (Gibco; Invitrogen; Life Technologies, Darmstadt, Germany) supplemented with 10% fetal bovine serum (FBS; PAA, Pasching, Austria) and 1% streptomycin/penicillin (Gibco; Invitrogen). T47D cells were derived from American Type Culture Collection (ATCC; Manassas, VA, USA) and were propagated in RPMI 1640 medium (PAA) supplemented with 10% FBS (PAA). Identity of cell lines was verified by single tandem repeat typing.

TABLE 1. *Clinical characterization of analyzed tissues in three sample sets*

Parameter	Screening set	Validation set 1	Validation set 2
Tissue source	Fresh-frozen	FFPE	Fresh-frozen
Menopausal status[a]/Age[b,c]			
Median	NA	All postmenopausal	58
95% CI			55–61
Range			39–85
Cancer type			
In situ	2	13	0
Invasive	8	32	43
Histological grade			
I	8[d]	12	4
II	1	33	25
III	0	0	14
Tumor type			
DCIS	1	7	0
LCIS	1	6	0
IDC	5	15	43
ILC	1 (TUB-LOB)	12	0
TUB	2	5	0
Tumor size (TNM)[e]			
pT1	NA	13	18
pT2		17	20
pT3		2	2
pT4		0	3
Lymphatic invasion			
Yes	NA	9	13
No		36	30
Nodal involvement[e,f]			
Yes	NA	10	20
No		21	23
ER staining[g]			
Negative	1	4	2
Positive	9	41	41
PR staining[g]			
Negative	1	4	4
Positive	9	41	39
Her2 amplification[h]			
Negative	10	42	43
Positive	0	3	0
MIB[a]/K_i-67 staining[b]			
Median	NA	5	10
95% CI		3–7	3–17
Range		0–25	0–90
Recurrences			
Yes	NA	NA	8
No			35
Metastases			
Yes	NA	NA	8
No			35
Normal tissues	10	11	8

All tumors used in the study were classified according to the World Health Organization tumor classification system and staged according to TNM Classification of Malignant Tumors. CI, confidence interval; FFPE, formalin-fixed paraffin-embedded; NA, not available. [a]For validation set 1. [b]For validation set 2. [c]Data for 2 patients not available. [d]Grade for LCIS in screening set not applicable. [e]For invasive cancers only. [f]Data for 1 patient in validation set 1 not available. [g]Staining >0 classified as positive. [h]Staining >2 classified as positive.

DNA isolation

Genomic DNA (gDNA) extraction from the screening set of tissues was performed according to the Tissue Protocol of the QIAamp DNA Mini Kit (Qiagen, Hilden, Germany). After incubation with AL buffer, an additional centrifugation step (2000 *g*, 10 min) was added for fat removal. For DNA extraction from tissues of validation set 1, formalin-fixed paraffin-embedded (FFPE) slides were deparaffinized by xylene, rehydrated using graded alcohols, and subjected to tissue microdissection using glass capillaries. Next, the samples were incubated for 16 h with proteinase K (100 μg/ml final concentration), then heat inactivated and directly used for bisulfite conversion. DNA from validation set 2 was isolated with TRIzol reagent (Invitrogen) according to the manufacturer's protocol. DNA from cell lines was extracted

using a QIAamp DNA Mini Kit (Qiagen) according to the Protocol for Cultured Cells.

MCIp, sample labeling, and array hybridization

Genes differentially methylated between tumor and normal tissue were detected by a combination of methyl-CpG enrichment technique and human CGI microarrays (Agilent Technologies, Santa Clara, CA, USA), similar to formerly applied approaches (15, 26, 28). MCIp experiments were performed as described previously (26, 27) with minor modifications. Briefly, 2 μg gDNA was fragmented by ultrasonication and immunoprecipitated with 60 μg MBD2-Fc protein. Unbound DNA was recovered by centrifugation. Bound DNA was eluted twice with buffers containing increasing NaCl concentrations (400, 500, 550, 600, and 1000 mM NaCl). Desalted eluates were controlled for enrichment of methylated DNA by quantitative PCR (qPCR) using the QuantiTect SYBR Green PCR kit (Qiagen) with primers targeting the imprinted *SNRPN* gene (**Table 2**). For differential labeling with Alexa Fluor 5 and Alexa Fluor 3 (BioPrime Total Genomic Labeling System; Invitrogen), 600 and 1000 mM NaCl fractions were combined. Tumor and normal samples were paired randomly (except for one case with available matching normal tissue) and hybridized to human CGI microarrays (Agilent Technologies) using the Oligo aCGH Hybridization Kit (Agilent Technologies). Array hybridization, washing, and scanning (Agilent scanner) were carried out according to the Agilent Microarray Analysis of Methylated DNA Immunoprecipitation protocol (Version 1.0, May 2008; Agilent Technologies).

TABLE 2. *Primers for qPCR analysis*

Gene	Primers	PCR product size
BCAN	F: GGAGGAGGCGACAAACTTC R: GAGCTGTCTCCTTCCAGAACA	149
HOXD1	F: CCTACCCCAAGTCCGTCTCT R: GTCAGTTGCTTGGTGCTGAA	162
KCTD8	F: CTGGAGCAGCTACACCGAGT R: GCTCATTACAGGAAGTCCCAC	134
KLF11	F: TGCAGCCACACCTGAACTAC R: GGGGAGAAACAGGTGTCCTT	113
NXPH1	F: GTCCCTTCTCGCAGGATTC R: CGGTGCCTTCTTTGAGTCTT	121
PCDH10	F: AGTGATCATGATGCCACCAA R: GGACAAAAGAAGGCATCCAG	118
POU4F1	F: ACAGCAAGCAGCCTCACTTT R: CTGGCGAAGAGGTTGCTC	133
RYR2	F: AGCAGCGATCAGAAGGAGAA R: GTAAGCTGCCGTTGCCATAA	105
SIM1	F: GGACGGTAGGCATGAGAACA R: GTCTCGCGGCATTTAAGGTA	141
TAC1	F: TTACTGGTCCGACTGGTACG R: AAAGAACTGCTGAGGCTTGG	109
TCF7L1	F: CAGTCAAGGACACGAGGTCA R: GGGAGAAGTGGTCATTGCTG	123
ACTB	F: ATTGGCAATGAGCGGTTC R: GGATGCCACAGGACTCCAT	76
SNRPN	F: TACATCAGGGTGATTGCAGTTCC R: TACCGATCACTTCACGTACCTTCG	126

F, forward; R, reverse.

Microarray data analysis and functional gene enrichment analysis

Microarray images were processed with Feature Extraction Software 10.5 (Agilent Technologies) according to the CGH_105_Dec08 protocol. Array data were corrected for GC content as reported previously (29). Using the Tibco Spotfire 2.2.0 software (Spotfire, Gothenburg, Sweden) and Excel scripts (Microsoft, Redmond, WA, USA), hypermethylated CGIs were identified. Probes most frequently detected as positive within a CGI, designated as differentially methylated regions (DMRs), were selected, and their representation in the various tumor samples was determined. The genes that were associated with 214 CGIs and hypermethylated in ≥6 of 10 arrays (reported in Supplemental Table S2) were used as input for the DAVID Functional Annotation tool (30, 31). Of the gene annotations analyzed on the Agilent human CGI array background, 165 were recognized, with the following parameters: similarity term overlap 4, similarity threshold 0.7, initial group membership 5, final group membership 5, and multilinkage threshold 0.5. The data discussed in this publication have been deposited in the U.S. National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (32) and are accessible through GEO Series accession number GSE35263.

Quantitative methylation analysis

Genomic DNA was subjected to sodium bisulfite treatment using the EZ methylation kit (Zymo Research, Orange, CA, USA) according to the manufacturer's instructions. Quantitative DNA methylation analysis was performed by MassArray EpiTyper technology (Sequenom, San Diego, CA, USA) as described previously (33, 34). Data were corrected using DNA methylation standards (0, 20, 40, 60, 80, and 100% methylated genomic DNA). Primer sequences are listed in **Table 3**. The location of EpiTyper amplicons relative to respective genes is depicted in Supplemental Fig. S2.

2′-Deoxy-5-azacytidine (DAC) treatment

Breast cancer cell lines were treated for 96 h (T47D) or 72 h (MCF7) with 0.5 or 1 μM DAC (decitabine; Sigma-Aldrich, Munich, Germany) dissolved in dimethyl sulfoxide (DMSO), respectively. Medium containing DAC or DMSO as solvent control (0.005 and 0.01% final DMSO concentration for T47D and MCF7, respectively) was renewed every 24 h. Each experiment was independently repeated 3 times.

RNA isolation, reverse transcription, and qPCR analysis

RNA was isolated with TRIzol (Invitrogen; MCF7) and RNeasy Mini Kit (Qiagen; T47D), including an additional DNase treatment (RNase-Free DNase Set; Qiagen). M-MLV Reverse Transcriptase set (Promega, Madison, WI, USA) was used for reverse transcription. Real-time qPCR analysis was performed with the QuantiTect SYBR Green PCR kit using a program of 15 min at 95°C followed by 45 cycles of 10 s at 95°C, 15 s at 60°C, and 20 s at 72°C on a Roche Lightcycler 480 Real-Time PCR System (Roche, Mannheim, Germany). β-*Actin* (*ACTB*) was used as a housekeeping gene. Primer sequences are available in Table 2.

Tissue microarray protein expression analysis

A tissue microarray (TMA) was assembled from the same collection of FFPE donor blocks as used for validation sample

TABLE 3. *Primers for EpiTyper validation*

Gene	Primers
BCAN	F: aggaagagagGGTGTTTTTAGAAGGGTTGTTGGT
	R: cagtaatacgactcactatagggagaaggctTACTTCCAACTCTACCAATAACAACC
HOXD1	F: aggaagagagTGGTYGTTGAGGAGGAAAGAGT
	R: cagtaatacgactcactatagggagaaggctAACCTAATACCRCACCTACC
KCTD8	F: aggaagagagGGAAGAAAATGTATTYGGTGTAGTTG
	R: cagtaatacgactcactatagggagaaggctTTCACCTACTTAAAACAAACCTTTAATC
KLF11	F: aggaagagagGTGATATTATAGGGGGTTTAGTGG
	R: cagtaatacgactcactatagggagaaggctTTTCCTCTCTATCTACCATACACC
NXPH1	F: aggaagagagGTATATTTTTTTGGGGGTTGTTTGTA
	R: cagtaatacgactcactatagggagaaggctCCTTTAACATAAACAACCCCTAAC
PCDH10	F: aggaagagagGAGGATAGAGTAGGTAAGTTGGG
	R: cagtaatacgactcactatagggagaaggctCTCCTACACCCTAACTATAATAACTC
POU4F1	F: aggaagagagGGTTTGAAAGGATGGTTTTTGTTTTGG
	R: cagtaatacgactcactatagggagaaggctCCTACAACTACAAAACAACCTCTTC
RYR2	F: aggaagagagGGTAGTAGAAGTAGAAGGTAG
	R: cagtaatacgactcactatagggagaaggctCCCCTTCCCCTAACAACACAC
SIM1	F: aggaagagagGGGTGGTTGAGYGTTTAGTTTAG
	R: cagtaatacgactcactatagggagaaggctAAATAACTCCCCAAACCRACCTC
TAC1	F: aggaagagagAATTTAATTGGGTTTAGATGTTATGGG
	R: cagtaatacgactcactatagggagaaggctTTATCCCAACCTCCTTAAACCTC
TCF7L1	F: aggaagagagTTTTTAATTTGAAGAATGTTTTGGG
	R: cagtaatacgactcactatagggagaaggctCTAATTCCTAAAAAACCCAATTTTC

F, forward; R, reverse.

set 1 and contained 5 normal tissues, 13 *in situ* tumors, and 32 invasive lesions. For each sample, two representative tissue cores 0.6 mm in diameter were obtained and combined in one TMA using a semiautomated array system (Beecher Instruments, Silver Spring, MD, USA). Immunostains were performed on 1-μm sections of the TMAs after heat pretreatment (1 mM EDTA, pH 9) using a polymer-based detection system (Zytomed, Berlin, Germany). Primary antibodies against neurexophilin 1 (NXPH1; ProteinTech Group, Chicago, IL, USA), tachykinin precursor 1 (TAC1; Sigma, St. Louis, MO, USA), and brevican (BCAN; Sigma) were diluted 1:100, 1:25, and 1:30, respectively. For NXPH1 and TAC1, 6 additional normal tissues were stained separately according to the same protocol.

Statistical analyses

The non-normal distribution of methylation and expression data and the relatively small number of observations for some experiments motivated the use of nonparametric statistics. Tumor hypermethylation was investigated by Mann-Whitney *U* tests and quantified by median differences with 1-sided 95% confidence intervals (CIs). The correlation among methylation levels in neighboring CpG units and the relationship between methylation and clinical characteristics was measured by Spearman's rank correlation ρ. Kruskal-Wallis tests were used to evaluate associations between methylation and expression.

Metastasis-free survival was calculated from the date of operation to detection of metastasis. Patients who did not experience metastasis by last contact were classified as censored. The median follow-up time of patients with metastasis (n=11) was 6 yr. The association between methylation and risk of metastasis was assessed by log-rank tests and represented by Kaplan-Meier plots. To account for the influence of established prognostic factors, hazard ratios (HRs) and 95% CIs were adjusted in a multivariate Cox proportional hazard regression. To quantify the accuracy of methylation as diagnostic marker, the area under the receiver operating characteristic (ROC) curve was estimated with 95% CIs based on the Mann-Whitney *U* statistic. Amplicon-wise estimates were obtained by bootstrapping, taking into account the intrasample correlation of methylation values (10,000 bootstrap samples). The reported *P* values are not corrected for multiplicity. The focus on amplicon-wise instead of single CpG-unit effects should alleviate the multiplicity problem, and together with the investigation of independent sample sets increase the likelihood of external replication of our findings.

RESULTS

Genome-wide methylation analyses identify homeobox and transcriptional regulatory genes as frequently hypermethylated in breast cancer

We applied a combination of MCIp and human CGI microarrays to identify aberrant DNA hypermethylation in 8 low-grade invasive carcinomas and 2 preinvasive *in situ* tumors, mostly ER- and/or PR-positive, against 10 normal breast tissues. The analysis yielded a list of 214 CGIs, which were detected as hypermethylated in ≥6 of 10 arrays (Supplemental Table S2). The CGI associated with *TAC1* was the only hotspot of aberrant tumor hypermethylation common for all 10 samples. Twelve CGIs were identified as hypermethylated in 9 samples, followed by 22, 62, and 117 CGIs hypermethylated in 8, 7, and 6 samples, respectively. Functional annotation analysis using DAVID software (30, 31) revealed a significant overrepresentation of homeobox genes (enrichment score 21.8) and genes involved in transcription and regulation of transcription among hypermethylated genes (16.8). Further functional terms included cell morphogenesis and neuron development (5.8), positive regulation of transcription (5.5), regula-

tion of neurogenesis (4.3), negative regulation of biosynthetic processes (4.0), cadherins (3.3), and cell migration and motility (2.4) (Supplemental Table S3). Analysis of the relative location of the CGIs indicated that 53% coincided with gene promoters (defined as a 5-kb region upstream of a gene transcription start site, according to Agilent probe annotations) and 62% overlapped with gene bodies (defined as the whole gene from the transcription start site to the end of the transcript). This observation supports a possible role of methylation in epigenetic gene silencing of the associated genes.

Quantitative EpiTyper methylation analysis verifies genome-wide methylation screen

To validate our microarray results, we selected a series of candidate genes: *BCAN*, *homoeobox D1 (HOXD1)*, *potassium channel tetramerization domain containing 8* (*KCTD8*), *Kruppel-like factor 11 (KLF11)*, *NXPH1*, *POU class 4 homeobox 1* (*POU4F1*), *ryanodine receptor 2* (*RYR2*), *single-minded homologue 1* (*SIM1*), *TAC1*, and *transcription factor 7-like 1* (*TCF7L1*) for quantitative EpiTyper analysis using the initial sample collection. Locations of most common DMRs covered by EpiTyper amplicons are indicated in Supplemental Fig. S2. Gene selection was based on the extent and frequency of methylation changes and proximity to gene promoter, as well as existence of specific antibodies. In addition, *protocadherin 10* (*PCDH10*) was chosen since it has been previously reported as a tumor suppressor gene (35–37).

EpiTyper analysis confirmed the presence of tumor hypermethylation in the analyzed DMRs, with each DMR being represented by 3 to 8 CpG units (single CpG sites or groups of 2 or more CpG sites; **Fig. 1**, top panel). Methylation in normal tissue was generally low and consistent with data on DNA methylation in normal breast luminal epithelial cells provided by the Roadmap Epigenomics Human Epigenome Atlas Data [ref. 38; compare methylated DNA immunoprecipitation (MeDIP) summary track in Supplemental Fig. S2]. Median methylation differences between tumor and normal samples are presented as absolute values; *i.e.*, methylation was measured on a 0 to 1 scale (Fig. 1, bottom panel). Over 85% single CpG units showed median methylation differences >0.1, representing significant hypermethylation in tumor samples at a single CpG unit resolution. The amplicon-wise combination of methylation values for single CpG units revealed a median methylation difference (with the lower limit of the 1-sided 95% CI) of 0.37 (0.23) for *BCAN*, 0.26 (0.09) for *HOXD1*, 0.23 (0.11) for *KCTD8*, 0.24 (0.17) for *KLF11*, 0.32 (0.18) for *NXPH1*, 0.17 (0.09) for *PCDH10*, 0.12 (0.05) for *POU4F1*, 0.28 (0.15) for *RYR2*, 0.25 (0.17) for *SIM1*, 0.15 (0.08) for *TAC1*, and 0.32 (0.20) for *TCF7L1*.

DAC treatment leads to demethylation and gene reexpression in two breast cancer cell lines

Next, we tested whether hypermethylation of the 11 candidate genes was associated with gene silencing. For demethylation and reexpression analyses, we

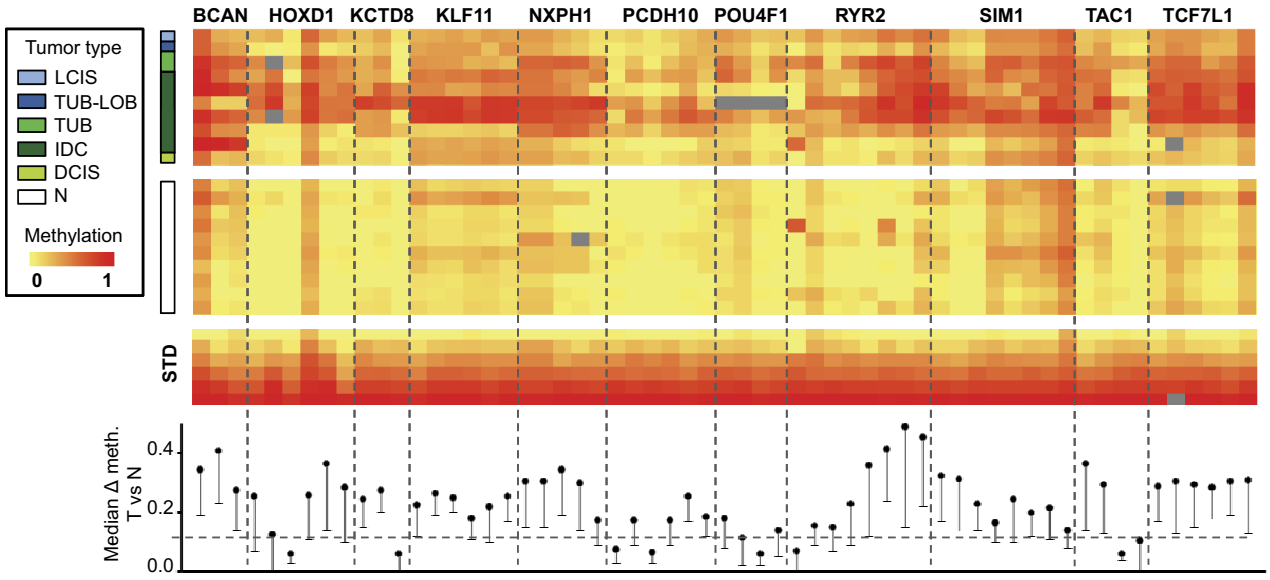


Figure 1. Validation of microarray data by quantitative methylation analysis. Heatmap illustrates the level of methylation for selected candidate genes in the initial sample set as measured by EpiTyper analysis. Each row of the heatmap represents a sample; each column depicts a CpG unit (single CpG or a group of CpGs). Tumor samples are located in the top part of the heatmap, followed by normal samples and a 6-point methylation standard (STD). Color coding at left of the heatmap indicates histological subtype of each tissue: LCIS, lobular carcinomas *in situ*; TUB-LOB, tubulolobular carcinoma; TUB, tubular carcinoma; IDC, invasive ductal carcinoma; DCIS, ductal carcinomas *in situ*; N, normal tissue. Within the heatmap, degree of methylation is indicated by a color gradient from yellow (not methylated) to red (fully methylated). Gray represents missing values. Median methylation differences between tumor group and normal tissues for each CpG unit (with 1-sided 95% CI) are depicted in the bottom panel. Dashed line marks a methylation difference of 0.1 (10%).

treated the ER-positive breast cancer cell lines T47D and MCF7 with optimized concentrations of DAC (**Fig. 2**). EpiTyper analysis revealed elevated median methylation levels in the range of 0.3 to 1.0 (30 to 100%) for all candidate DMRs in both cell lines. DAC treatment reduced methylation of DMRs by 5 to 50% compared to solvent control samples. Accordingly, the majority of the genes showed either an increase in expression (*BCAN*, *HOXD1*, *NXPH1*, and *PCDH10*) or a change from no expression in the control samples to detectable mRNA levels in DAC-treated samples (*KCTD8*, *POU4F1*, *RYR2*, *SIM1*, and *TAC1*). Exceptions were *KLF11* and *TCF7L1*, for which expression was already high in both cell lines before DAC treatment and was not further increased following demethylation. These results were consistent between both investigated cell lines and indicated that methylation indeed contributes to regulation of expression of most of the candidate genes.

Candidate genes are early and frequently hypermethylated in breast carcinomas

To determine whether the selected genes are commonly targeted by hypermethylation in breast cancer, we quantified methylation levels in a larger sample set (validation set 1) consisting of 45 macrodissected low-grade, ER-positive, and/or PR-positive *in situ* and invasive carcinomas of various histological subtypes and 11 normal tissues (patient characteristics in Table 1). Consistent with quantitative methylation data obtained for the screening sample set, methylation was increased in both *in situ* and invasive cancers compared to normal samples (**Fig. 3**, top panel). No statistically significant methylation differences were found among cancer subtypes (IDC, ILC, and TUB; data not shown). Notably, 9 genes were already significantly hypermethylated in *in situ* samples compared to normal tissue (Fig. 3, bottom panel, and **Table 4**). Methylation levels in the *in situ* samples were almost as high as in invasive tumors and showed a similar pattern for the majority of the genes. These data suggest that the investigated genes were hypermethylated in early stages of tumor development and hypermethylation was maintained during breast carcinogenesis. Median methylation differences between invasive breast cancer and normal samples were highest for *TCF7L1* (0.58), *NXPH1* (0.54), *BCAN* (0.51), and *KCTD8* (0.49).

To assess the frequency of hypermethylated samples in validation set 1, we defined gene-specific thresholds based on the 90th methylation percentile in normal samples (Table 4). *KCTD8*, *KLF11*, *NXPH1*, *POU4F1*, and *SIM1* (respective thresholds: 0.03, 0.30, 0.39, 0.16, and 0.35) were identified as the most commonly hypermethylated among our candidate genes in invasive tumors, with hypermethylation frequencies ranging from 91 to 97%. These genes were also hypermethylated in 62 to 92% of the *in situ* samples. We have thus validated 11 genes as aberrantly methylated in both *in situ* and invasive carcinomas in comparison to normal breast tissue. Our results indicate that the investigated DMRs are frequently hypermethylated already in early stages of breast cancer.

NXPH1 hypermethylation is associated with decreased protein expression

Using TMAs representing samples of validation sample set 1, we examined whether methylation of selected DMRs affected protein expression levels of the associated genes. Commercially available antibodies against BCAN, NXPH1, and TAC1 showed consistent and specific immunohistochemical staining. Staining intensity was interpreted semiquantitatively at a score scale of 0 to 3. Epithelial cells in normal samples (**Fig. 4*A–C***) demonstrate strong NXPH1 staining (score 3); one of the invasive tumors (Fig. 4*D*) shows moderate NXPH1-staining (score 2), while both other invasive tumor samples (Fig. 4*E*, *F*) lack detectable NXPH1-expression (score 0). Staining for BCAN, NXPH1 and TAC1 was less frequent in tumor samples than normal tissues (**Table 5**). For example, NXPH1 protein was detectable

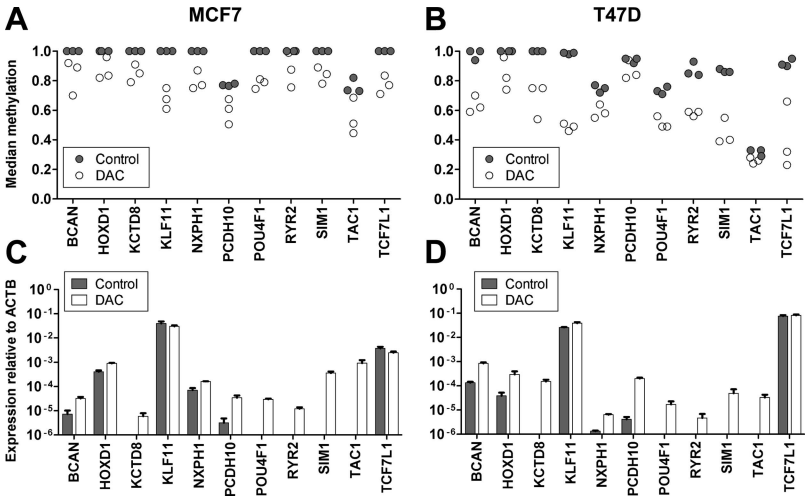


Figure 2. Demethylation and gene reexpression after DAC treatment. MCF7 cells (*A*, *C*) were treated with 1 μM DAC for 72 h, and T47D cells (*B*, *D*) were treated with 0.5 μM of DAC for 96 h. Degree of methylation was measured by quantitative EpiTyper analyses. *A*, *B*) Median amplicon methylation of solvent control and DAC-treated samples is depicted for each candidate gene. *C*, *D*) mRNA expression was determined by RT–qPCR. Relative expression levels were calculated in relation to BCAN assuming 90% primer efficiency. Data depict mean ± SD values of 3 independent experiments.

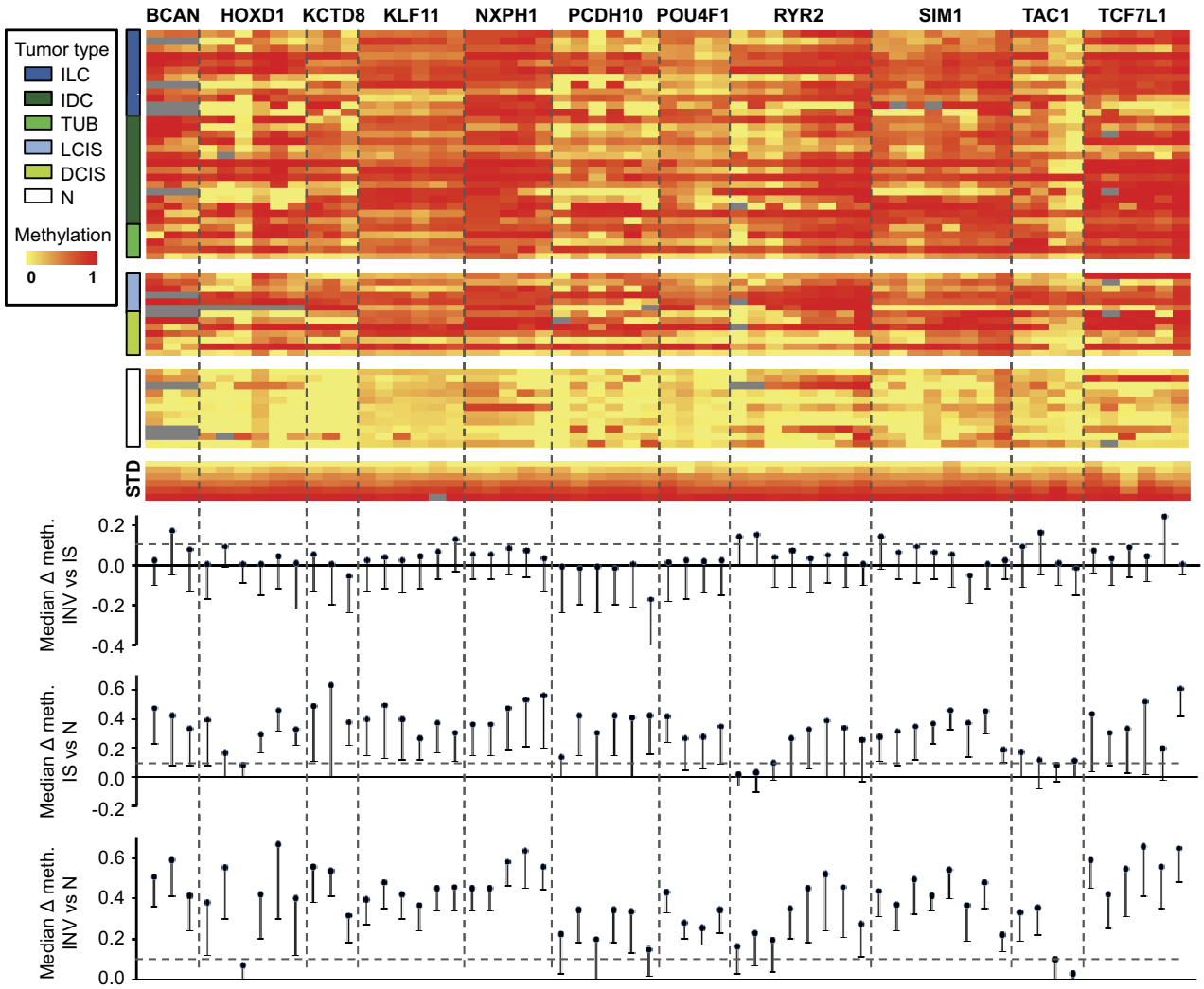


Figure 3. Verification of hypermethylation in validation set 1. Degree of methylation was measured by EpiTyper analysis. Color coding schemes indicating methylation gradient and the histological tissue type are same as in Fig. 1. Bottom panels represent median methylation differences between the following groups: invasive tumors (INV) and *in situ* carcinomas (IS), IS and normal tissues (N), and INV and N. Corresponding 1-sided 95% CIs are also represented. dashed line marks a methylation difference of 0.1 (10%).

in 90% of normal tissues, 73% of *in situ* carcinomas, and 57% of invasive tumors. Staining intensity increased with decreasing *NXPH1* methylation (P=0.0073, Kruskal-Wallis test; Fig. 4*G*), indicating an inverse relationship between *NXPH1* hypermethylation and protein expression.

Hypermethylation is confirmed in low- and high-grade breast carcinomas

We quantitatively analyzed methylation levels in a second independent sample set (validation set 2). This set comprised 43 IDC ER- and/or PR-positive cases and included both low- and high-grade tumor samples (patient characteristics in Table 1). Quantitative methylation analysis confirmed methylation differences between invasive tumors and normal samples (**Table 6** and Supplemental Fig. S1). Using the same thresholds as in Table 4, *BCAN*, *KCTD8*, and *SIM1* hypermethylation correctly identified 93, 100, and 95% of the invasive carcinomas, respectively. To evaluate the diagnostic accuracy of methylation in the investigated genes, we performed ROC curve analyses exemplarily for *BCAN*, *KCTD8*, *KLF11*, *NXPH1*, and *SIM1* (**Table 7**). In validation set 1, area under curve (AUC) estimates for ROC were in the range of 0.82–0.88 for *in situ* carcinomas and 0.88–0.96 for invasive breast cancers. High discriminating power was confirmed for invasive cancers in validation set 2, with AUC estimates of 0.91–0.98.

The combined study of both validation sample sets had 90% power to detect a hypermethylation difference > 0.34. This was determined by multiplying the combined sample size for a 2-sample t test [1-sided, α=0.0045 (0.05 divided by 11 genes), 88 tumor, and 19 normal samples, common SD of 0.3] by the asymptotic relative efficiency for a Mann-Whitney U test under normality (0.95).

TABLE 4. *Methylation differences between* in situ *(IS) and invasive (INV) lesions compared to normal (N) samples and hypermethylation frequencies in validation set 1*

	IS (n=13) *vs.* N (n=11)		INV (n=32) *vs.* N		INV *vs.* IS			Frequency	
Gene	Δ Methylation	*P*	Δ Methylation	*P*	Δ Methylation	*P*	Threshold	IS	INV
BCAN	0.46 (0.19; ∞)	0.007	0.51 (0.40; ∞)	<0.001	0.06 (−0.08; ∞)	0.628	0.33	70	89
HOXD1	0.36 (0.15; ∞)	0.032	0.41 (0.26; ∞)	<0.001	0.03 (−0.11; ∞)	0.723	0.25	67	72
KCTD8	0.42 (0.11; ∞)	0.039	0.49 (0.39; ∞)	<0.001	0.01 (−0.20; ∞)	0.818	0.03	77	97
KLF11	0.38 (0.15; ∞)	0.005	0.43 (0.35; ∞)	<0.001	0.05 (−0.09; ∞)	0.690	0.30	62	91
NXPH1	0.44 (0.23; ∞)	0.003	0.54 (0.42; ∞)	<0.001	0.06 (−0.05; ∞)	0.667	0.39	77	97
PCDH10	0.36 (0.11; ∞)	0.042	0.30 (0.20; ∞)	0.002	−0.04 (−0.23; ∞)	0.953	0.15	77	81
POU4F1	0.33 (0.11; ∞)	0.029	0.34 (0.27; ∞)	<0.001	0.02 (−0.15; ∞)	0.835	0.16	69	94
RYR2	0.23 (0.01; ∞)	0.181	0.36 (0.22; ∞)	0.002	0.08 (−0.07; ∞)	0.672	0.50	46	66
SIM1	0.37 (0.22; ∞)	0.003	0.42 (0.34; ∞)	<0.001	0.05 (−0.08; ∞)	0.749	0.35	92	94
TAC1	0.12 (0.01; ∞)	0.363	0.21 (0.13; ∞)	0.013	0.06 (−0.08; ∞)	0.743	0.29	46	66
TCF7L1	0.42 (0.15; ∞)	0.025	0.58 (0.41; ∞)	<0.001	0.06 (−0.01; ∞)	0.643	0.42	69	84

n = number of analyzed samples. Δ Methylation indicates median methylation difference between groups, with 1-sided 95% CI in parenthesis. Methylation was measured on a 0 (0%) to 1 (100%) scale. *P* values are for a median hypermethylation < 0.1 (10%). Threshold was used to calculate frequencies of hypermethylation in tumors. Threshold for each amplicon was defined as the 90th percentile of median methylation levels for combined normal tissues from validation set 1 and validation set 2. Frequency indicates percentage of samples exceeding the threshold value.

Hypermethylation of selected genes correlates with tumor characteristics and increased risk of metastasis

We further explored potential relationships between hypermethylation and tumor characteristics for the invasive tumor groups. We found an association between tumor grade and *BCAN* methylation in both validation sets (P =0.04 for validation set 1; P=0.01 for validation set 2). In addition, *NXPH1* methylation positively correlated with tumor proliferation (determined by MIB or Ki-67 staining), as reflected by Spearman's rank correlation ρ = 0.32 (P=0.02) in validation set 1 and ρ = 0.24 (P=0.13) in validation set 2. *KCTD8* methylation and tumor proliferation were correlated with ρ = 0.24 (P=0.15) in validation set 1, and ρ = 0.34 (P=0.01) in validation set 2. *BCAN*, *HOXD1*, and *SIM1* methylation levels positively correlated with *EZH2* mRNA expression.

We were interested in the prognostic relevance of aberrant methylation in the selected DMRs and investigated whether methylation levels were associated with an increased risk to develop metastases using data from validation set 2 with clinical follow-up available. A weak association between methylation and risk of metastasis was found for *KLF11* (P=0.06) and *SIM1* (P=0.04). Survival analyses indicated that a 0.1 (10%) increase in *KLF11* methylation increases the risk of metastasis by 37% to 1.37 (95% CI: 0.99–2.28), while for *SIM1*, the increase in risk was 29% to 1.29 (95% CI: 1.02–1.88). Metastasis-free survival also correlated with the expression of tumor proliferation marker Ki-67 in the investigated patient collective (P=0.01). After adjustment for Ki-67 staining, the HR associated with a 10% increase in methylation was 1.38 for *KLF11* (95% CI: 0.97–2.59) and 1.25 for *SIM1* (95% CI: 0.96–1.91). These data suggest that methylation levels of both genes may have prognostic significance.

In addition to HRs based on amplicon-wise methylation, we illustrate the potential prognostic value of single CpG-unit methylation for *KLF11* using a Kaplan-Meier plot (**Fig. 5*A*** and **Table 8**). Taking CpG unit 5 of *KLF11* as an example, the patient cohort was divided into 2 groups according to median methylation. This methylation value was selected because it divides the sample set into 2 subsets of similar size with 21 and 22 women. Two metastasis events occurred in the group with methylation < 0.62, whereas 9 metastasis events were detected in the remaining patients (P=0.009; log-rank test). Figure 5*B* represents the estimated correlations among methylation values in neighboring CpG-units of *KLF11*. This figure has a descriptive purpose similar to linkage disequilibrium plots in association studies. The plot reflects a strong within-amplicon dependence of methylation values among CpG-units. The lowest estimated correlation amounted 0.82 and the highest was 0.97.

DISCUSSION

Genome-wide methylation screen and validation of hypermethylation as diagnostic marker

In the present study, we aimed at identifying genes hypermethylated in breast cancer as potential biomarkers for detection and prognosis. We primarily focused on the biology of low-grade ER- and/or PR-positive breast cancer using samples from various histological subtypes and stages, including DCIS, IDC, TUB, LCIS, and ILC. Our array-based genome-wide methylation screen identified >200 hypermethylated CGIs that may constitute novel targets for biomarker discovery and harbor pathological relevance. In concordance with earlier reports (15, 18, 39, 40), homeobox genes and protocadherins were enriched among the hypermethylated genes. Our list of 214 CGIs contains 197 unique gene names. Among them, 21 genes overlapped with a

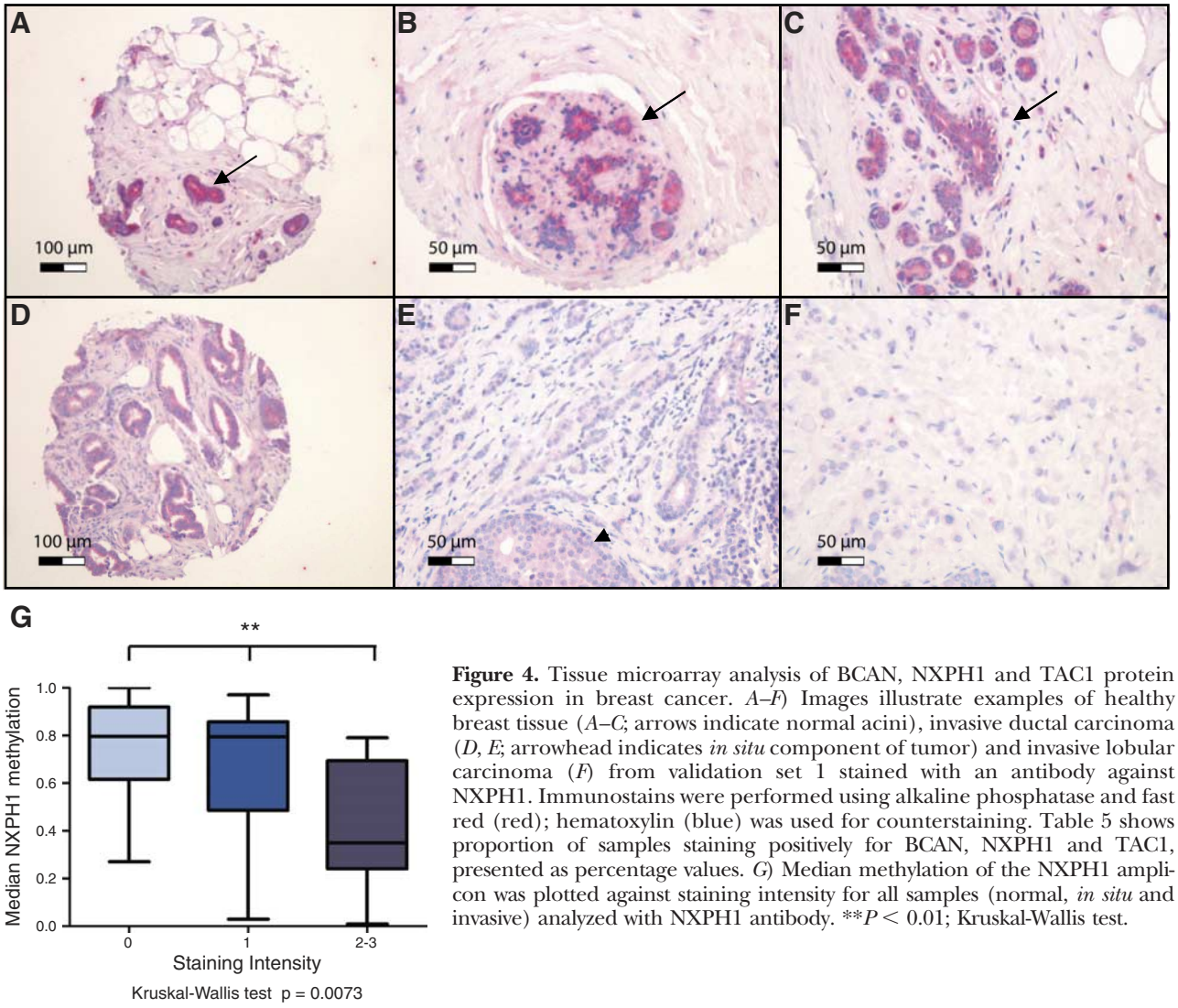


Figure 4. Tissue microarray analysis of BCAN, NXPH1 and TAC1 protein expression in breast cancer. *A–F*) Images illustrate examples of healthy breast tissue (*A–C*; arrows indicate normal acini), invasive ductal carcinoma (*D, E*; arrowhead indicates *in situ* component of tumor) and invasive lobular carcinoma (*F*) from validation set 1 stained with an antibody against NXPH1. Immunostains were performed using alkaline phosphatase and fast red (red); hematoxylin (blue) was used for counterstaining. Table 5 shows proportion of samples staining positively for BCAN, NXPH1 and TAC1, presented as percentage values. *G*) Median methylation of the NXPH1 amplicon was plotted against staining intensity for all samples (normal, *in situ* and invasive) analyzed with NXPH1 antibody. $^{**}P < 0.01$; Kruskal-Wallis test.

list of 79 hypermethylated genes identified in the study of Tommasi *et al.* (15). Similarly, a set of 21 genes overlapped with 264 hypermethylated genes provided by Hill *et al.* (19), whereas 60 genes were identified as recurrently hypermethylated in comparison with 852 hypermethylated genes reported by Van der Auwera *et al.* (20). These parallels between different studies using various platforms (see Supplemental Table S1) underscore the universal character of the observed hypermethylation events and support their applicability as detection biomarkers. We selected 11 candidates for quantitative methylation analysis and validated hypermethylation at representative DMRs in the screening sample set and in two independent validation sample sets. Previous single-gene studies on genes such as *14-3-3σ*, *RASSF1A*, *HIN-1*, *GSTP1*, *APC*, *CDH1*, *CTNNB1*, and others (41–49) suggest that methylation changes occur early in the development of breast cancer and are frequent already at the *in situ* stage. From our 11 candidate genes, *TAC1*, *NXPH1*, and *KLF11* were identified as hypermethylated in DCIS in an earlier genome-wide screen (15). In agreement with these studies, we

TABLE 5. *Proportion of samples in Fig. 4 staining positively for BCAN, NXPH1, and TAC1*

	Positive samples (%)		
Tissue type	BCAN	NXPH1	TAC1
Normal	75 (3/4)	90 (9/10)	89 (8/9)
In situ carcinoma)	45 (5/11	73 (8/11)	73 (8/11)
Invasive cancer)	36 (10/28)	57 (16/28)	79 (23/29)

Values in parentheses indicate number of tissues with detectable staining in comparison to total number of analyzed tissues.

TABLE 6. *Methylation changes in invasive lesions (INV) relative to normal (N) samples in validation set 2*

	INV (n=43) *vs.* N (n=8)			
Gene	Δ Methylation	*P*	Threshold	Frequency
BCAN	0.48 (0.39; ∞)	<0.001	0.33	93
HOXD1	0.40 (0.35; ∞)	<0.001	0.25	88
KCTD8	0.52 (0.46; ∞)	<0.001	0.03	100
KLF11	0.46 (0.40; ∞)	<0.001	0.30	86
NXPH1	0.47 (0.43; ∞)	<0.001	0.39	86
PCDH10	0.29 (0.24; ∞)	<0.001	0.15	79
POU4F1	0.40 (0.34; ∞)	<0.001	0.16	88
RYR2	0.53 (0.48; ∞)	<0.001	0.50	72
SIM1	0.37 (0.33; ∞)	<0.001	0.35	95
TAC1	0.32 (0.20; ∞)	<0.001	0.29	65
TCF7L1	0.58 (0.55; ∞)	<0.001	0.42	86

n = number of analyzed samples. Δ Methylation indicates median methylation difference between groups, with 1-sided 95% CI in parenthesis. Methylation was measured on a 0 (0%) to 1 (100%) scale. *P* values are for a median hypermethylation < 0.1 (10%). Threshold was used to calculate frequencies of hypermethylation in tumors. Threshold for each amplicon was defined as the 90th percentile of median methylation levels for combined normal tissues from validation set 1 and validation set 2. Frequency indicates percentage of samples exceeding the threshold value.

confirmed frequent hypermethylation of all the candidate genes already at the *in situ* stage (frequency ranging from 46 to 92%). We did not detect significant differences in methylation levels between *in situ* carcinomas and invasive carcinomas although the methylation frequency was ~2–29% lower in *in situ* cases compared to invasive tumors (see Fig. 3 and Table 4). Methylation levels were not different between distinct tumor subtypes. Our data suggest that at least some targets of aberrant methylation are common for many histological subtypes of ER- and/or PR-positive breast cancer and are frequently hypermethylated in both early and late stages of breast cancer. Hypermethylation of these genes may therefore be regarded as an attractive cancer diagnostic biomarker. The predictive value of our study is restricted by the limited number of *in situ* samples and the fact that some *in situ* cases were derived from samples with adjacent invasive tumors. However, especially early detection of *in situ* stage tumors may constitute a significant clinical advantage. Present results may help to design future studies to validate current findings. Our estimated ROC AUC estimates for *BCAN*, *KCTD8*, *KLF11*, *NXPH1* and *SIM1* methylation in *in situ* carcinoma were 0.85, 0.84, 0.88, 0.83, and 0.82 (see Table 7), respectively, which indicates the high diagnostic value of methylation at these genes.

Hypermethylation as a potential prognostic marker for metastasis risk?

In earlier studies, high methylation of single candidate genes or groups of genes identified in genome-wide screens was associated with poor overall survival (18), occurrence of distant metastasis (20), or tumor relapse (19). Here, hypermethylation of two candidate genes (*KLF11* and *SIM1*) was weakly associated with an elevated risk of developing metastases, although the observed prognostic value may be dependent on tumor proliferation. Thus it is important to further assess the prognostic relevance of these findings using a larger sample collection.

Specific function of candidate genes in breast carcinogenesis?

What remains to be answered is whether our identified candidate genes, in addition to serving as potential biomarkers, might have tumor suppressor properties or specific functions in mammary development or tissue differentiation that are lost during breast carcinogenesis as a result of hypermethylation. In cell culture experiments, hypermethylation of these genes was in general associated with gene silencing that was reversed

TABLE 7. *Amplicon-wise area under ROC curve estimates with 95% CIs*

	Validation set 1		Validation set 2
Gene	*In situ*	Invasive	Invasive
BCAN	0.85 (0.72–0.95)	0.93 (0.86–0.97)	0.94 (0.90–0.98)
KCTD8	0.84 (0.71–0.94)	0.93 (0.88–0.97)	0.97 (0.94–0.99)
KLF11	0.88 (0.75–0.97)	0.96 (0.90–0.99)	0.91 (0.83–0.97)
NXPH1	0.83 (0.69–0.95)	0.93 (0.85–0.98)	0.98 (0.96–1.00)
SIM1	0.82 (0.71–0.91)	0.88 (0.83–0.93)	0.92 (0.88–0.95)

Values in parentheses represent 95% CIs.

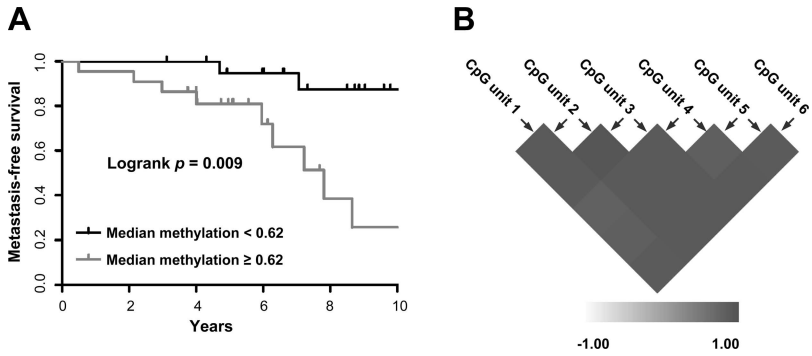


Figure 5. Association of *KLF11* methylation with metastasis-free survival. *A*) Kaplan-Meier curves of metastasis-free survival were plotted for methylation of *KLF11* CpG unit 5 (compare Supplemental Fig. S1). Patients in validation set 2 were divided into 2 groups according to median methylation at CpG unit 5 (0.62), separating the sample set into 2 subgroups of equal size. Groups are represented by black and gray lines for methylation levels < 0.62 and ≥ 0.62, respectively. Table 8 summarizes numbers of patients followed over time for each group. *B*) Spearman's rank correlations ρ among methylation values of single CpG units within the *KLF11* amplicon were calculated based on methylation data from validation sample set 2 to illustrate the within amplicon dependence of methylation values. Color code ranges from white ($\rho=-1$) to dark gray ($\rho=1$).

after DAC treatment (see Fig. 2). DAC treatment may lead to gene reexpression either through direct demethylation or through an indirect effect as the result of reexpression or activation of transcription factors, *etc.* Despite reduced DNA methylation, TCF7L1 and KLF11 mRNA levels were not up-regulated following DAC treatment. In comparison with the reexpressed genes, these genes had already high basal expression in the investigated cell lines before DAC treatment. In both cases, the genomic region investigated by EpiTyper analyses was located outside of the canonical promoter, indicating that this most differentially methylated region might not contribute to the regulation of expression of these genes.

Further supporting the inverse link between methylation and gene expression, protein expression of BCAN, NXPH1, and TAC1 was reduced in tumor samples compared to normal tissue as indicated by immunohistochemical staining of TMA sections (see Fig. 4). According to the NCBI gene database, the primary function of *BCAN*, *HOXD1*, *KCTD8*, *NXPH1*, *PCDH10*, *POU4F1*, *SIM1,* and *TAC1* involves developmental and neural processes. *PCDH10* has putative tumor suppressor function in nasopharyngeal, esophageal, gastric and other cancers (36, 50), whereas *RYR2* is overexpressed in malignant melanomas and derived cell lines (51). So far, no tumor-related specific functions of our candidate genes in breast development have been established, but it is possible that loss of expression of these genes plays a yet-unknown role in breast carcinogenesis.

Widschwendter *et al.* (52) and Ohm *et al.* (53) postulated an alternative hypothesis addressing the function of genes hypermethylated in cancer. They suggest that the methylation profile could reflect a stem cell origin of the tumor. In stem cells, sets of developmental genes are reversibly silenced by bivalent chromatin marks of activating histone 3 lysine 4 dimethylation (H3K4me2) and inactivating histone 3 lysine 27 trimethylation (H3K27me3). During carcinogenesis, the bivalent chromatin becomes methylated and thus locked in an inactive state. Whether this process is a bystander effect of tumorigenesis or whether it may actively contribute to tumor development (*e.g.*, by conserving the stem cell-like phenotype through permanent inactivation of differentiation programs) is still unclear. The fact that the top 214 aberrantly methylated CGIs identified in our study are significantly enriched for homeobox genes might provide further support for this model. Homeobox genes are known to be frequent targets of polycomb group (PcG) proteins in embryonic stem cells (ESCs) (54) and these targets, in turn, tend to be *de novo* methylated in cancer (55, 56). *EZH2* is a PcG protein that has been reported to be overexpressed in breast cancer (57). This observation was confirmed in a subset of samples of validation set 2 with significant increase of *EZH2* mRNA levels in tumor samples compared to normal tissues ($P=0.031$; *t* test). *BCAN*, *HOXD1*, and *SIM1* are known PcG protein targets in ESCs (54), and their methylation levels correlated with *EZH2* expression. This may point to a functional link between EZH2 and gene silencing (58), but does not exclude independent mechanisms (59).

In summary, we have identified aberrant methylation of multiple genomic regions. These changes are frequent in early stages of breast cancer, as well as in its most common subtypes. They are also detected in high-grade tumors and might serve as biomarkers for the detection of breast cancer. FJ

TABLE 8. *Association of* KLF11 *methylation with metastasis-free survival*

	Women at risk in year					
Median methylation	0	2	4	6	8	10
<0.62	21	21	20	16	11	5
≥ 0.62	22	21	17	8	3	2

Summary of data from Fig. 5*A*: number of patients followed over time for each group.

The authors thank Manuela Zucknick for statistical support in array analysis and Julia Strathmann for providing DNA from MCF7 cells. The authors thank Oliver Mücke, Sandra Hahn, and Tina Philipp for excellent technical assistance. This project was funded by the FRONTIER Innovation Fund as part of the Excellence Initiative of Heidelberg University. J.L.B. and M.B. were supported by a grant of the Deutsche Forschungsgemeinschaft (DFG; SFB/TRR77, project Z2). Research in the S.D. laboratory is supported by the Helmholtz Society (VH-NG-504), the German Research Foundation

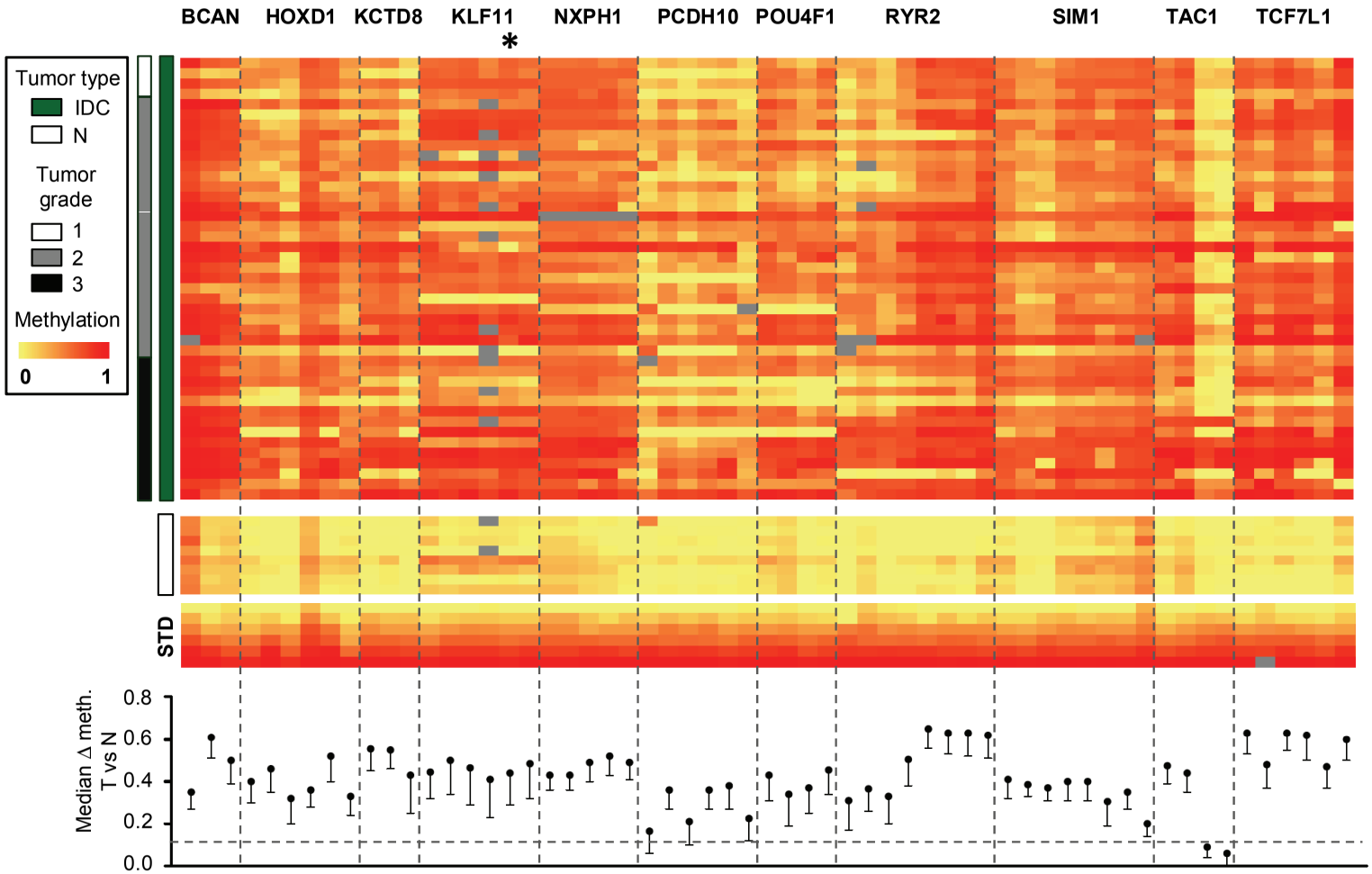
(DFG Transregio TRR77, TP B03) and the Marie Curie Program of the European Commission. R.C. was supported by a DFG grant (CL 427/2-1). The authors declare no conflicts of interest.

REFERENCES

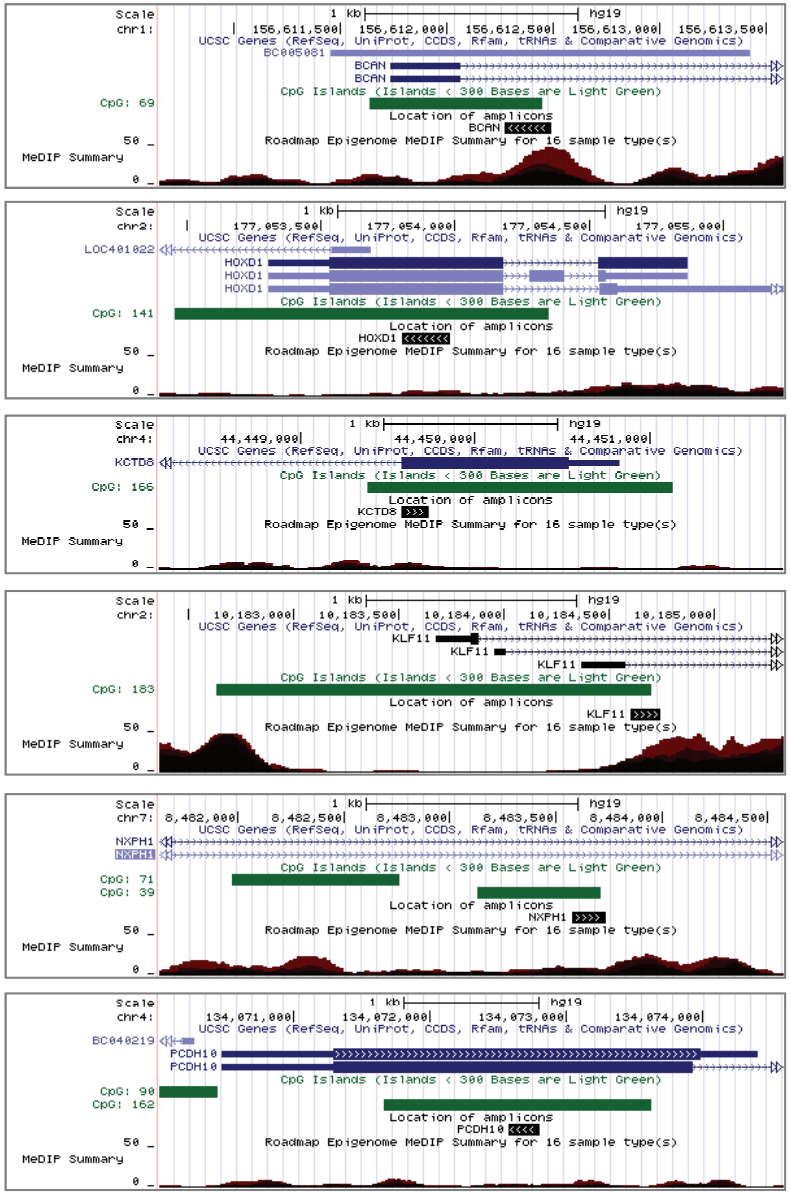
1. Ferlay, J., Shin, H. R., Bray, F., Forman, D., Mathers, C., and Parkin, D. M. (2010) GLOBOCAN 2008 v1.2, Cancer incidence and mortality worldwide: IARC CancerBase No. 10 [online], International Agency for Research on Cancer>, Lyon, France, http://globocan.iarc.fr
2. Holleczek, B., Arndt, V., Stegmaier, C., and Brenner, H. (2011) Trends in breast cancer survival in Germany from 1976 to 2008: a period analysis by age and stage. *Cancer Epidemiol.* **35,** 399–406
3. Weigelt, B., and Reis-Filho, J. S. (2009) Histological and molecular types of breast cancer: is there a unifying taxonomy? *Nat. Rev. Clin. Oncol.* **6,** 718–730
4. Vollenweider-Zerargui, L., Barrelet, L., Wong, Y., Lemarchand-Beraud, T., and Gomez, F. (1986) The predictive value of estrogen and progesterone receptors' concentrations on the clinical behavior of breast cancer in women. Clinical correlation on 547 patients. *Cancer* **57,** 1171–1180
5. Sorlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Thorsen, T., Quist, H., Matese, J. C., Brown, P. O., Botstein, D., Eystein Lonning, P., and Borresen-Dale, A. L. (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 10869–10874
6. Sorlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J. S., Nobel, A., Deng, S., Johnsen, H., Pesich, R., Geisler, S., Demeter, J., Perou, C. M., Lonning, P. E., Brown, P. O., Borresen-Dale, A. L., and Botstein, D. (2003) Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 8418–8423
7. Esteller, M. (2007) Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat. Rev. Genet.* **8,** 286–298
8. Jones, P. A., and Baylin, S. B. (2007) The epigenomics of cancer. *Cell* **128,** 683–692
9. Portela, A., and Esteller, M. (2010) Epigenetic modifications and human disease. *Nat. Biotechnol.* **28,** 1057–1068
10. Feinberg, A. P., Ohlsson, R., and Henikoff, S. (2006) The epigenetic progenitor origin of human cancer. *Nat. Rev. Genet.* **7,** 21–33
11. Widschwendter, M., and Jones, P. A. (2002) DNA methylation and breast carcinogenesis. *Oncogene* **21,** 5462–5482
12. Ordway, J. M., Budiman, M. A., Korshunova, Y., Maloney, R. K., Bedell, J. A., Citek, R. W., Bacher, B., Peterson, S., Rohlfing, T., Hall, J., Brown, R., Lakey, N., Doerge, R. W., Martienssen, R. A., Leon, J., McPherson, J. D., and Jeddeloh, J. A. (2007) Identification of novel high-frequency DNA methylation changes in breast cancer. *PLoS One* **2,** e1314
13. Huang, Y., Nayak, S., Jankowitz, R., Davidson, N. E., and Oesterreich, S. (2011) Epigenetics in breast cancer: what's new? *Breast Cancer Res.* **13,** 225
14. Dedeurwaerder, S., Fumagalli, D., and Fuks, F. (2011) Unravelling the epigenomic dimension of breast cancers. *Curr. Opin. Oncol.* **23,** 559–565
15. Tommasi, S., Karm, D. L., Wu, X., Yen, Y., and Pfeifer, G. P. (2009) Methylation of homeobox genes is a frequent and early epigenetic event in breast cancer. *Breast Cancer Res.* **11,** R14
16. Killian, J. K., Bilke, S., Davis, S., Walker, R. L., Jaeger, E., Killian, M. S., Waterfall, J. J., Bibikova, M., Fan, J. B., Smith, W. I., Jr., and Meltzer, P. S. (2011) A methyl-deviator epigenotype of estrogen receptor-positive breast carcinoma is associated with malignant biology. *Am. J. Pathol.* **179,** 55–65
17. Li, L., Lee, K. M., Han, W., Choi, J. Y., Lee, J. Y., Kang, G. H., Park, S. K., Noh, D. Y., Yoo, K. Y., and Kang, D. (2010) Estrogen and progesterone receptor status affect genome-wide DNA methylation profile in breast cancer. *Hum. Mol. Genet.* **19,** 4273–4277
18. Fackler, M. J., Umbricht, C. B., Williams, D., Argani, P., Cruz, L. A., Merino, V. F., Teo, W. W., Zhang, Z., Huang, P., Visvananthan, K., Marks, J., Ethier, S., Gray, J. W., Wolff, A. C., Cope, L. M., and Sukumar, S. (2011) Genome-wide methylation analysis identifies genes specific to breast cancer hormone receptor status and risk of recurrence. *Cancer Res.* **71,** 6195–6207
19. Hill, V. K., Ricketts, C., Bieche, I., Vacher, S., Gentle, D., Lewis, C., Maher, E. R., and Latif, F. (2011) Genome-wide DNA methylation profiling of CpG islands in breast cancer identifies novel genes associated with tumorigenicity. *Cancer Res.* **71,** 2988–2999
20. Van der Auwera, I., Yu, W., Suo, L., Van Neste, L., van Dam, P., Van Marck, E. A., Pauwels, P., Vermeulen, P. B., Dirix, L. Y., and Van Laere, S. J. Array-based DNA methylation profiling for breast cancer subtype discrimination. *PLoS One* **5,** e12616
21. Holm, K., Hegardt, C., Staaf, J., Vallon-Christersson, J., Jonsson, G., Olsson, H., Borg, A., and Ringner, M. (2010) Molecular subtypes of breast cancer are associated with characteristic DNA methylation patterns. *Breast Cancer Res.* **12,** R36
22. Bediaga, N. G., Acha-Sagredo, A., Guerra, I., Viguri, A., Albaina, C., Ruiz Diaz, I., Rezola, R., Alberdi, M. J., Dopazo, J., Montaner, D., de Renobales, M., Fernandez, A. F., Field, J. K., Fraga, M. F., Liloglou, T., and de Pancorbo, M. M. (2010) DNA methylation epigenotypes in breast cancer molecular subtypes. *Breast Cancer Res.* **12,** R77
23. Dedeurwaerder, S., Desmedt, C., Calonne, E., Singha, S. K., Haibe-Kains, B., Defrance, M., Michiels, S., Volkmar, M., Deplus, R., Luciani, J., Lallemand, F., Larsimont, D., Toussaint, J., Haussy, S., Rothe, F., Rouas, G., Metzger, O., Majjaj, S., Saini, K., Putmans, P., Hames, G., Baren, N. V., Coulie, P. G., Piccart, M., Sotiriou, C., and Fuks, F. (2011) DNA methylation profiling reveals a predominant immune component in breast cancers. *EMBO Mol. Med.* **3,** 726–741
24. Flanagan, J. M., Cocciardi, S., Waddell, N., Johnstone, C. N., Marsh, A., Henderson, S., Simpson, P., da Silva, L., Khanna, K., Lakhani, S., Boshoff, C., and Chenevix-Trench, G. (2010) DNA methylome of familial breast cancer identifies distinct profiles defined by mutation status. *Am. J. Hum. Genet.* **86,** 420–433
25. Fang, F., Turcan, S., Rimner, A., Kaufman, A., Giri, D., Morris, L. G., Shen, R., Seshan, V., Mo, Q., Heguy, A., Baylin, S. B., Ahuja, N., Viale, A., Massague, J., Norton, L., Vahdat, L. T., Moynahan, M. E., and Chan, T. A. (2011) Breast cancer methylomes establish an epigenomic foundation for metastasis. *Sci. Transl. Med.* **3,** 75ra25
26. Gebhard, C., Schwarzfischer, L., Pham, T. H., Schilling, E., Klug, M., Andreesen, R., and Rehli, M. (2006) Genome-wide profiling of CpG methylation identifies novel targets of aberrant hypermethylation in myeloid leukemia. *Cancer Res.* **66,** 6118–6128
27. Schilling, E., and Rehli, M. (2007) Global, comparative analysis of tissue-specific promoter CpG methylation. *Genomics* **90,** 314–323
28. Rauch, T. A., Zhong, X., Wu, X., Wang, M., Kernstine, K. H., Wang, Z., Riggs, A. D., and Pfeifer, G. P. (2008) High-resolution mapping of DNA hypermethylation and hypomethylation in lung cancer. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 252–257
29. Schilling, E., El Chartouni, C., and Rehli, M. (2009) Allele-specific DNA methylation in mouse strains is mainly determined by cis-acting sequences. *Genome Res.* **19,** 2028–2035
30. Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37,** 1–13
31. Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4,** 44–57
32. Edgar, R., Domrachev, M., and Lash, A. E. (2002) Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30,** 207–210
33. Ehrich, M., Nelson, M. R., Stanssens, P., Zabeau, M., Liloglou, T., Xinarianos, G., Cantor, C. R., Field, J. K., and van den Boom, D. (2005) Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 15785–15790
34. Goeppert, B., Schmezer, P., Dutruel, C., Oakes, C., Renner, M., Breinig, M., Warth, A., Vogel, M. N., Mittelbronn, M., Mehrabi, A., Gdynia, G., Penzel, R., Longerich, T., Breuhahn, K., Popanda, O., Plass, C., Schirmacher, P., and Kern, M. A. (2010) Down-regulation of tumor suppressor A kinase anchor protein

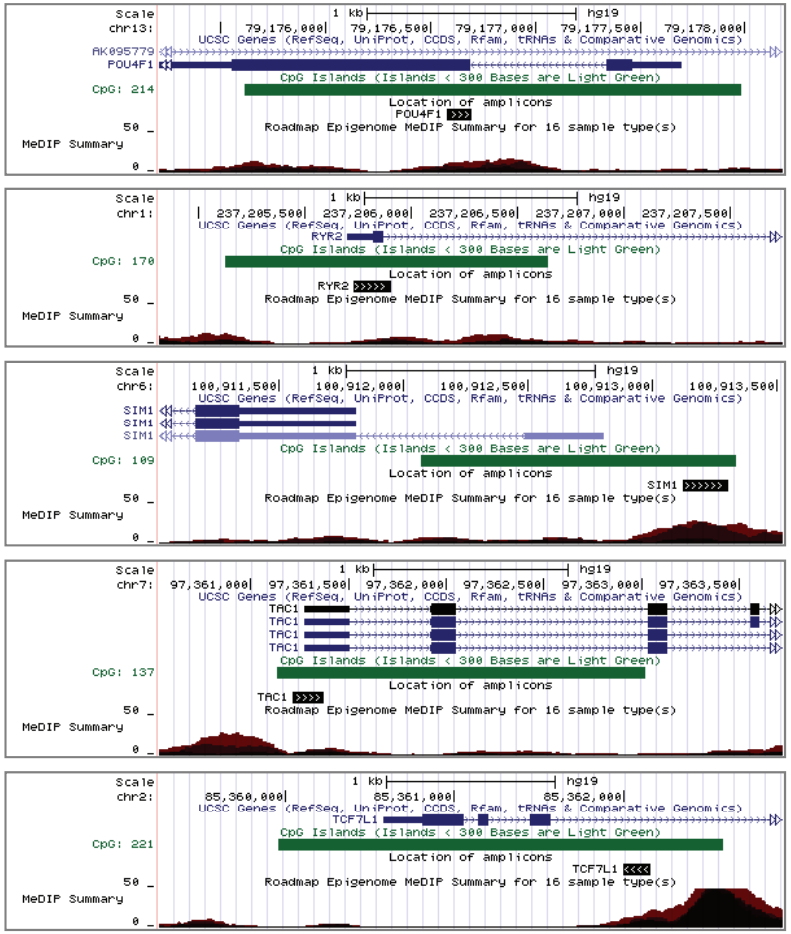
12 in human hepatocarcinogenesis by epigenetic mechanisms. *Hepatology* **52,** 2023–2033
35. Narayan, G., Freddy, A. J., Xie, D., Liyanage, H., Clark, L., Kisselev, S., Un Kang, J., Nandula, S. V., McGuinn, C., Subramaniyam, S., Alobeid, B., Satwani, P., Savage, D., Bhagat, G., and Murty, V. V. (2011) Promoter methylation-mediated inactivation of PCDH10 in acute lymphoblastic leukemia contributes to chemotherapy resistance. *Genes Chromosomes Cancer* **50,** 1043–1053
36. Yu, J., Cheng, Y. Y., Tao, Q., Cheung, K. F., Lam, C. N., Geng, H., Tian, L. W., Wong, Y. P., Tong, J. H., Ying, J. M., Jin, H., To, K. F., Chan, F. K., and Sung, J. J. (2009) Methylation of protocadherin 10, a novel tumor suppressor, is associated with poor prognosis in patients with gastric cancer. *Gastroenterology* **136,** 640–651 e641
37. Narayan, G., Scotto, L., Neelakantan, V., Kottoor, S. H., Wong, A. H., Loke, S. L., Mansukhani, M., Pothuri, B., Wright, J. D., Kaufmann, A. M., Schneider, A., Arias-Pulido, H., Tao, Q., and Murty, V. V. (2009) Protocadherin PCDH10, involved in tumor progression, is a frequent and early target of promoter hypermethylation in cervical cancer. *Genes Chromosomes Cancer* **48,** 983–992
38. UCSC Genome Browser Track Data Hub: Roadmap Epigenomics Data Complete Collection at Washington University in St. Louis, Assembly hg19, accessed August 2012 at http://vizhub.wustl.edu/VizHub/RoadmapReleaseAll.txt
39. Novak, P., Jensen, T., Oshiro, M. M., Watts, G. S., Kim, C. J., and Futscher, B. W. (2008) Agglomerative epigenetic aberrations are a common event in human breast cancer. *Cancer Res.* **68,** 8616–8625
40. Kamalakaran, S., Varadan, V., Giercksky Russnes, H. E., Levy, D., Kendall, J., Janevski, A., Riggs, M., Banerjee, N., Synnestvedt, M., Schlichting, E., Karesen, R., Shama Prasada, K., Rotti, H., Rao, R., Rao, L., Eric Tang, M. H., Satyamoorthy, K., Lucito, R., Wigler, M., Dimitrova, N., Naume, B., Borresen-Dale, A. L., and Hicks, J. B. (2011) DNA methylation patterns in luminal breast cancers differ from non-luminal subtypes and can identify relapse risk independent of other clinical variables. *Mol. Oncol.* **5,** 77–92
41. Umbricht, C. B., Evron, E., Gabrielson, E., Ferguson, A., Marks, J., and Sukumar, S. (2001) Hypermethylation of 14-3-3 sigma (stratifin) is an early event in breast cancer. *Oncogene* **20,** 3348–3353
42. Lehmann, U., Langer, F., Feist, H., Glockner, S., Hasemeier, B., and Kreipe, H. (2002) Quantitative assessment of promoter hypermethylation during breast cancer development. *Am. J. Pathol.* **160,** 605–612
43. Fackler, M. J., McVeigh, M., Evron, E., Garrett, E., Mehrotra, J., Polyak, K., Sukumar, S., and Argani, P. (2003) DNA methylation of RASSF1A, HIN-1, RAR-beta, Cyclin D2 and Twist in in situ and invasive lobular breast carcinoma. *Int. J. Cancer* **107,** 970–975
44. Lee, J. S. (2007) GSTP1 promoter hypermethylation is an early event in breast carcinogenesis. *Virchows Arch.* **450,** 637–642
45. Pasquali, L., Bedeir, A., Ringquist, S., Styche, A., Bhargava, R., and Trucco, G. (2007) Quantification of CpG island methylation in progressive breast lesions from normal to invasive carcinoma. *Cancer Lett.* **257,** 136–144
46. Hoque, M. O., Prencipe, M., Poeta, M. L., Barbano, R., Valori, V. M., Copetti, M., Gallo, A. P., Brait, M., Maiello, E., Apicella, A., Rossiello, R., Zito, F., Stefania, T., Paradiso, A., Carella, M., Dallapiccola, B., Murgo, R., Carosi, I., Bisceglia, M., Fazio, V. M., Sidransky, D., and Parrella, P. (2009) Changes in CpG islands promoter methylation patterns during ductal breast carcinoma progression. *Cancer Epidemiol. Biomarkers Prev.* **18,** 2694–2700
47. Muggerud, A. A., Ronneberg, J. A., Warnberg, F., Botling, J., Busato, F., Jovanovic, J., Solvang, H., Bukholm, I., Borresen-Dale, A. L., Kristensen, V. N., Sorlie, T., and Tost, J. (2010) Frequent aberrant DNA methylation of ABCB1, FOXC1, PPP2R2B and PTEN in ductal carcinoma in situ and early invasive breast cancer. *Breast Cancer Res.* **12,** R3
48. Moelans, C. B., Verschuur-Maes, A. H., and van Diest, P. J. (2011) Frequent promoter hypermethylation of BRCA2, CDH13, MSH6, PAX5, PAX6 and WT1 in ductal carcinoma in situ and invasive breast cancer *J. Pathol.* **225,** 222–231
49. Park, S. Y., Kwon, H. J., Lee, H. E., Ryu, H. S., Kim, S. W., Kim, J. H., Kim, I. A., Jung, N., Cho, N. Y., and Kang, G. H. (2011) Promoter CpG island hypermethylation during breast cancer progression. *Virchows Arch.* **458,** 73–84
50. Ying, J., Li, H., Seng, T. J., Langford, C., Srivastava, G., Tsao, S. W., Putti, T., Murray, P., Chan, A. T., and Tao, Q. (2006) Functional epigenetics identifies a protocadherin PCDH10 as a candidate tumor suppressor for nasopharyngeal, esophageal and multiple other carcinomas with frequent methylation. *Oncogene* **25,** 1070–1080
51. Deli, T., Varga, N., Adam, A., Kenessey, I., Raso, E., Puskas, L. G., Tovari, J., Fodor, J., Feher, M., Szigeti, G. P., Csernoch, L., and Timar, J. (2007) Functional genomics of calcium channels in human melanoma cells. *Int. J. Cancer* **121,** 55–65
52. Widschwendter, M., Fiegl, H., Egle, D., Mueller-Holzner, E., Spizzo, G., Marth, C., Weisenberger, D. J., Campan, M., Young, J., Jacobs, I., and Laird, P. W. (2007) Epigenetic stem cell signature in cancer *Nat. Genet.* **39,** 157–158
53. Ohm, J. E., and Baylin, S. B. (2007) Stem cell chromatin patterns: an instructive mechanism for DNA hypermethylation? *Cell Cycle* **6,** 1040–1043
54. Lee, T. I., Jenner, R. G., Boyer, L. A., Guenther, M. G., Levine, S. S., Kumar, R. M., Chevalier, B., Johnstone, S. E., Cole, M. F., Isono, K., Koseki, H., Fuchikami, T., Abe, K., Murray, H. L., Zucker, J. P., Yuan, B., Bell, G. W., Herbolsheimer, E., Hannett, N. M., Sun, K., Odom, D. T., Otte, A. P., Volkert, T. L., Bartel, D. P., Melton, D. A., Gifford, D. K., Jaenisch, R., and Young, R. A. (2006) Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125,** 301–313
55. Ohm, J. E., McGarvey, K. M., Yu, X., Cheng, L., Schuebel, K. E., Cope, L., Mohammad, H. P., Chen, W., Daniel, V. C., Yu, W., Berman, D. M., Jenuwein, T., Pruitt, K., Sharkis, S. J., Watkins, D. N., Herman, J. G., and Baylin, S. B. (2007) A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. *Nat. Genet.* **39,** 237–242
56. Schlesinger, Y., Straussman, R., Keshet, I., Farkash, S., Hecht, M., Zimmerman, J., Eden, E., Yakhini, Z., Ben-Shushan, E., Reubinoff, B. E., Bergman, Y., Simon, I., and Cedar, H. (2007) Polycomb-mediated methylation on Lys27 of histone H3 premarks genes for de novo methylation in cancer. *Nat. Genet.* **39,** 232–236
57. Kleer, C. G., Cao, Q., Varambally, S., Shen, R., Ota, I., Tomlins, S. A., Ghosh, D., Sewalt, R. G., Otte, A. P., Hayes, D. F., Sabel, M. S., Livant, D., Weiss, S. J., Rubin, M. A., and Chinnaiyan, A. M. (2003) EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 11606–11611
58. Vire, E., Brenner, C., Deplus, R., Blanchon, L., Fraga, M., Didelot, C., Morey, L., Van Eynde, A., Bernard, D., Vanderwinden, J. M., Bollen, M., Esteller, M., Di Croce, L., de Launoit, Y., and Fuks, F. (2006) The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* **439,** 871–874
59. Kondo, Y., Shen, L., Cheng, A. S., Ahmed, S., Boumber, Y., Charo, C., Yamochi, T., Urano, T., Furukawa, K., Kwabi-Addo, B., Gold, D. L., Sekido, Y., Huang, T. H., and Issa, J. P. (2008) Gene silencing in cancer by histone H3 lysine 27 trimethylation independent of promoter DNA methylation. *Nat. Genet.* **40,** 741–750

Received for publication June 27, 2012.
Accepted for publication August 13, 2012.



Supplementary Figure 1. Methylation levels in validation sample set 2.
The degree of methylation of invasive ductal carcinomas (IDC), normal tissues (N) and methylation standard (STD) was assessed by EpiTYPER analysis. Each row represents a sample and each column depicts a CpG unit (single CpG or a group of CpGs). The degree of methylation is indicated by a color gradient from yellow (not methylated) to red (fully methylated). Grey color represents missing values. Grade and type of respective tumors are represented by the color coding on the left side of the heatmap. Asterisk marks the CpG unit corresponding to Kaplan-Meier curve presented in Figure 5. The lower panel depicts median methylation differences with one-sided 95% confidence interval between tumor samples and normal tissues for each CpG unit.





Supplementary Figure 2. Location of Epityper amplicons relative to the CGI and comparison with Roadmap Epigenome MeDIP data. From top to bottom: Candidate genes are marked in blue/black/light blue. The location of the corresponding CGI is depicted in green. Black bars indicate the location of amplicons used for quantitative Epityper methylation analysis. “MeDIP summary” tracks represent sequencing data of immunoprecipitated methylated DNA derived from normal breast luminal epithelial cells of four donors. Signal intensity reflects the likelihood that the underlying region is methylated.

Supplementary Table 1. Key findings of genome-wide methylation studies in breast cancer.

Author, Date	Focus	No. of tumor / normal samples	Screening platforms
Ordway, 2007 (12)	Molecular detection markers	9 (IDCs) / 9	McrBC[a] + OGHAv1.0 microarray[b]
Novak, 2008 (37)	Long-range methylation patterns	16 / 5[c]	MeDIP[d] + Affymetrix Promoter[e]
Tommasi, 2009 (15)	Early-stage molecular markers	6 (DCIS) / ND	MIRA[f] + Agilent CpG Island[g]
Van der Auwera, 2010 (20)	Molecular subtype definition	62 / 10	Illumina Infinium 27k[h]
Holm, 2010 (21)	Molecular subtype definition	189 / 4	Illumina GoldenGate[i]
Bediaga, 2010 (22)	Molecular subtype definition	28 / 28[j]	Illumina Golden Gate[i]
Li, 2010 (17)	Markers for HR[k] status	24 / 0	Illumina Infinium 27k[h]
Flanagan, 2010 (24)	Familial breast tumors	33 / 0	MeDIP + Affymetrix Promoter[e]
Kamalakaran, 2010 (38)	Molecular subtype definition and markers for prognosis	108 / 11	MOMA[l] + Custom CpG Island tiling arrays (390k format, Nimblegen)
Fang, 2011 (25)	B-CIMP signature as a prognostic marker	39 (IDCs) / 0	Illumina Infinium 27k[h]
Fackler, 2011 (18)	Markers for HR status and prognosis	103 / 21	Illumina Infinium 27k[h]
Hill, 2011 (19)	Markers for HR status and prognosis	43 (IDCs) / 4	Illumina Infinium 27k[h]
Killian, 2011 (16)	Molecular subtype definition and markers for prognosis	351 / 47	Illumina GoldenGate[i]
Dedeurwaerder, 2011 (23)	Molecular subtype definition and markers for prognosis	119 (IDCs) / 4	Illumina Infinium 27k[h]

[a]McrBC restriction cleavage.
[b]Coverage: 21,294 unique probes for 19,595 transcription start sites.
[c]Additionally human mammary epithelial cells and 4 cell lines.
[d]Methylated DNA ImmunoPrecipitation using an antibody against methylated C.
[e]Affymetrix GeneChip Human Promoter 1.0R Array. Coverage: >4.6 million probes for > 25,500 promoters.
[f]Methylated-CpG Island Recovery Assay.
[g]Agilent Human CpG Island Array. Coverage: 237,220 probes for 27,800 CpG islands.
[h]Illumina Infinium HumanMethylation27 BeadChip. Coverage: 27,578 CpG sites for 14,475 promoters.
[i]Illumina GoldenGate Methylation Cancer Panel I. Coverage: 1,505 CpG sites for 807 genes
[j]Additionally 4 peritumoral samples.
[k]Hormone receptor.
[l]Methylation Oligonucleotide Microarray Analysis

Supplementary Table 2. Genes associated with the 214 CGIs hypermethylated in ≥ 6 /10 arrays.

10 / 10:
TAC1[A,C]

9 / 10:
AK094793
BCOR[A], AK125468
FOXD3[B,C]
HOXD1
KCTD8[C]
KLF11[A,B]
NR2E1[A,C]
NXPH1[A]
POU4F1[B,C]
SIM1
TCF7L1
Y12726/ ZFP161

8 / 10:
ADCYAP1[C]
C1orf164
CCDC4
chr1:119344555-119344959
chr5:72629900-72631526
chr8:23618369-23621154
DMRTA2
DSC3[C]
FAT4
HTR1B[B,C]
IGF2BP1
LBXCOR1
LHX1[C], AATF
ONECUT2[C]
PCDH8[C]
PRDM14[A,C]
RYR2[B,C]
SLC32A1
SSTR1
TBX20[B,C]
TLX3[C]
ZIC4

7 / 10:
AFAP1L1
AK056817
AK091058, SOX1[C]
AK123758
AK125792
AK289802,
AB007886
AP1M2
AVPR1A
AY166699,
AK054822
AY354498
BC009340
BC025350
BC156660
BCAN
BCL2L11
BMI1, SPAG6[C]
C1orf114[A,B,C]
CA3[B,C]
CBLN4[C]
CDO1[B,C]
DIO3[C]
DMRT1[C]
EMX2
EVX2[A]
FAM19A2
FERD3L[B,C]
FZD10[B]
HNF1B
hsa-mir-183
HTR2C
INA[C]
KCNJ3
MGC29506
NEFL
NEFM
NFIX
NID2[C]
NRN1[C]
OLIG3
OTX2[A]
PAX2[A]
PAX6
PAX6
PAX9[C]
PCDH17
PCDH7[A]
PCDHGC5
PENK[C]
PHOX2B
POU4F2[B,C]
RBP1[B]
SALL3[C]
SIX6[B,C]
SLITRK1
SPON1[C]
TBX20[B,C]
TBX20[B,C]
TCHH
TRIM67
TUBB6[C]
XKR4
ZIC1

6 / 10:
ACTA1[A]
AK026266
AK055805
AK056817
AK096498
AK124214, PAX9[C]
AK126832
AK131233
AK131264
BARHL2[A,C]
BC034002,
AK097396
BC042562
BX647670,
AK090984,
CR936715
C14orf39
C1orf14
C5orf39
CCDC140
CCK[B,C]
CDX2[B C], PRHOXNB
chr10:5922269-5922737
chr16:3178896-3179552
chr4:144840235-144841406
chr5:172604958-172605640
chr6:105494900-105496201
chr7: 117641276-117641909, ANKRD7
chr7:35460636-35461278
chr8:23622749-23623674
chr8:68035885-68038207
chr9:2231303-2232072
chrX:287322-289636
CLDN11[B]
CNIH3
CYP26A1
DBX1
DGKI[C]
DLX4
DMRTA2
DQ227570
EBF2
EFNA3
EIF5A2[C]
ELOVL2
EN2
EPS8L1
ERICH1
FN3K
FOXA2[B,C]
FOXG1
GAD1
GAD2[C]
GATA6
GFI1[A,C]
GLYATL1
HAND2[C]
HERC5
HIST3H2BB
HMGA2
HOXD12[A,C], HOXD11[B,C]
HOXD9[B,C]
HS3ST2[C]
HSPA2[C]
ID4[C]
IRX1[A]
KCNC2
KLF6
KLHL26
LHX2[A]
LHX4
LHX5
LOXL2
LRRTM1[C]
MPPED2
MSC[C]
MSC[C]
NFATC1[A]
NKAPL[A]
NKX2-8[A]
NKX3-2
NMBR[C]
OLFML2B
ONECUT1
ONECUT2[C]
PAX6
PCDH10
PCDHAC1[B,C]
PDX1
PHYHIPL
PM20D1
POU3F3[C]
POU4F1[B,C]
PRKCZ
PRLHR[C]
RAB3A
RBM4
RGS20
SLC12A5[C]
SLC22A3[C]
SOX1[C]
SOX17
SOX9
SPAG6[C]
TBX18
TBX18
TBX2[C]
TFAP2C
TMEM101
UNCX
USP44
WNK4
WT1[A,C]
ZIC1
ZIC4
ZNF232[C]
ZNF385A
ZNF397OS
ZNF518B
ZNF613

In case that no gene was associated, the exact genomic location of hypermethylated region is given.
The detailed information on location of CGIs relative to genes, genomic location of DMRs and their representation in the various tumor samples is available from the authors.
A, B, C – genes overlapping with gene lists provided in references (15), (19) and (20), respectively.

Supplementary Table 3. DAVID analysis of 214 top CGIs identified in the genome-wide screen.

Annotation cluster	Functional Term	Enrichment score	Median Gene Count
1	Homeobox	21.82	29
2	Transcription and regulation of transcription	16.77	65
3	Cell morphogenesis and neuron development	5.76	16
4	Positive regulation of transcription	5.53	22
5	Regulation of neurogenesis	4.29	10
6	Negative regulation of biosynthetic processes	3.96	19
7	Cadherin	3.32	8
8	Cell migration and motility	2.41	10
9	Neurological transmission	1.88	7
10	Ear morphogenesis	1.68	6
11	Zinc finger, LIM-type	1.53	4

The complete table with output of the DAVID analysis is available from the authors.