SUPPLEMENTARY MATERIALS AND METHODS

Genomic DNA extraction, sodium bisulfite conversion and quality assurance: DNA extractions and proteinase K treatment of fresh frozen tissues were performed as described previously and according to manufacturer's directions: Sodium bisulfite conversion of DNA was done with the EZ DNA methylation Kit (Zymo Research, D5002). Purified DNA 2 µg or lysate from 6-8 tissue sections (10 µm thick) was mixed with 7.5 µl M-dilution buffer in a final volume of 42.5 µl and incubated at 42^{o}C 30 min. Conversion was accomplished after the addition of 97.5 µl CT Conversion Reagent (prepared by adding 750 µl water and 185 µl M-dilution buffer to a vial of CT Conversion Reagent intended for 10 reactions) and samples were incubated in a thermocycler overnight (95^{o}C 30 sec, 50^{o}C 1 hr, 16 cycles). Bisulfite-converted DNA was column-purified using a ZymoSpin IC column and eluted in 12 ul of water. DNA was quantified using a Nanodrop-1000 and 1µl of a 1:5 dilution of DNA was tested to insure amplification potential with a panel of markers developed for QM-MSP which overlap array CpG loci (e.g. AKR1B1, Figure 5D). Any sample that amplified poorly was not used.

Illumina Infinium Human Methylation27 BeadChip

Bisulfite-converted DNA was analyzed using Illumina Infinium Human Methylation27 BeadChip Kit (WG-311-1202) in the DNA Microarray Core, Johns Hopkins University. The BeadChip contains 27,578 CpG loci covering more than 14,000 human RefSeq genes. The methylation of a single CpG is detected with two bead type probes per CpG (recognizing U = unmethylated, and M = methylated DNA) at single-base resolution by using primer extension with a labeled nucleotide in single color. For array, samples were adjusted to 75 ng/µl with water and 4 µl was processed according to manufacturer's directions. Data were extracted using GenomeStudio Methylation Module v1.0 software. The methylation value for each 50 bp CpG locus is expressed as a β-value, representing a continuous measurement from 0 (completely unmethylated) to 1 (completely methylated) according to the following calculation: β value =

(signal intensity of M probe)/[(signal intensity of M + U probes) +100]; the average β-value is based on the average intensity of all U and M CpG probes for a given locus. Probe performance is reported as the p-detection value, reflecting the extent of variation among replicates of a single bead type. Only loci with p-detection values < 0.0001 were considered. As controls for normal and tumor samples, we used replicate samples of a normal breast organoid (ORG) and a breast cancer cell line MDA-MB-231, (provided by NCI as the ICBP45 breast cancer cell line set through ATCC) with a highly methylated genome.

Test of array reproducibility

Additional replicate samples (two groups of 4 samples each) were used to assess assay reproducibility. Two identical aliquots each of MDA-MB-231 cell line DNA and of ORG DNA were processed independently through sodium bisulfite conversion. Two unrelated tumor samples were microdissected using laser capture, processed to completion with sodium bisulfite and then halved. This resulted in 2 identical groups of 4 samples, MDA-MB-231, ORG, Tumor1 and Tumor 2. The 8 samples were run together on a HumanMethylation27BeadChip then analyzed. Results showed excellent correlation between groups of samples (r^2=0.991; Supplementary Figure 1A)

Data analysis

T-tests, ANOVA, Wilcoxon rank sum tests and chi-squared tests were used to identify significant association between methylation and such covariates as tumor subtype and ER status. Cox regression was used to model associations between methylation levels and time to recurrence, in the presence of relevant clinical covariates. Associations to binary outcomes were modeled using logistic regression. Where possible, we have used empirical Bayes linear models(1) in place of standard t-tests, ANOVA and linear regressions. These methods use modified estimates of variance to reduce false positive rates in the analysis of microarray data. Data were also analyzed using GenomeStudio software (Illumina, Inc., San Diego, CA), Bioconductor

in R (http://www.bioconductor.org), and GraphPad Prism v5.02 (GraphPad Software, San Diego, CA; http://www.graphpad.com). For differential analysis, results are reported as DiffScore within GenomeStudio Methylation module, after computing for false discovery (rate 0.05). To convert DiffScore to adjusted p-value: p-value =1/ [10^(ABS(DiffScore)/(10)], where p-value of 0.05 is equal to a DiffScore of +/-13, p-value of 0.01 is equal to a DiffScore of +/-20, and p-value of 0.001 is equal to a DiffScore of +/-30. Methylation Score derivation: We derived a methylation score for ER-subtype specific markers by calculating the average methylation over all 40 probes, after standardizing the data for each CpG locus by subtracting its mean methylation and dividing by its standard deviation. Hypermethylation at some loci was associated with the ER-positive phenotype while other loci were hypermethylated in the ER-negative samples, so standardized methylation levels for the latter loci were multiplied by -1, so that all 40 genes would vote in the same direction.

QM-MSP (Quantitative-Multiplex Methylation-Specific PCR)

Details of this method have been reported previously (2-4). Briefly, sodium bisulfite converted template DNA is pre-amplified by PCR with gene-specific forward and reverse primers, hybridizing to regions lacking CpG residues and up to 12 genes are co-amplified with amplicons ranging to 300 bp. QM-MSP primers were designed to overlap array CpG loci. An aliquot of this reaction is diluted between 1:5 and 1:10,000 with water and then used as template for real-time methylation-specific PCR which was performed on an Applied Biosystems 7500 system. Separately for each gene, the copy number of methylated and unmethylated alleles is determined by absolute quantitation in a single well using two-color fluorescent labeling. The relative amount of methylation is calculated as % M = 100 X [no. of copies of methylated DNA/ (no. of copies of methylated + unmethylated DNA)]. Cumulative methylation of a panel of genes is calculated as the sum of the % of each gene, where 10 genes in a panel would have a maximal possibility of 1000 cumulative methylation units per sample. Concordant quantitative

methylation levels are observed between the Human Methylation27 BeadChip array and QM-MSP on primary tumors for AKR1B1 (Figure 4C). AKR1B1 primers and probes: AKR1B1_F_Ext: gYGtaattaattagaaggtttttt, AKR1B1_R_Ext: aacacctaccttccaaatac, AKR1B1_FM: gCGCGttaatCGtaggCGttt, AKR1B1_RM: cccaataCGataCGaccttaac, AKR1B1_FUM: TGgTGTGttaatTGtaggTGtttt, AKR1B1_RUM: cccaataCAataCAaccttaacC, AKR1B1_M_Probe: VIC- CGtacctttaaataaccCGtaaaatCGa-TAMRA, and AKR1B1_U_Probe: 6FAM- ACAtacctttaaataaccCAtaaaatCAac-TAMRA.

SUPPLEMENTARY FIGURE LEGENDS

Supplementary Figure 1: Distribution of methylation in primary invasive breast tumors compared to normal breast tissue. A. Performance of the HumanMethylation27 array was assessed by evaluating duplicate groups of 3 primary tissue samples, and one cell line, MDA-MB-231. High level of reproducibility was observed with r^2=0.991 for 27549 probes with detection p-value <0.0001. B. The distribution of CpG methylated loci in tumors (left; mean = 0.42 ± .21)) and normal breast (right; mean = 0.39 ± .26) is shown for 8374 probes which varied in methylation above SD = 0.100 among tumor samples (n = 103). Of these 8374 locus probes (Y-axis), methylation levels (beta value; x-axis) ≥0.15 were observed for 7539 probes in tumor and 6032 probes in normal breast, indicating that methylation is not limited to breast cancer for most of the probes. C. Array based methylation levels in primary invasive breast tumor were further compared to that of normal breast by calculating the ratio of median methylation across tumor samples (n=103) divided by median methylation values found in normal samples (n =15) at each locus. For display on the X-axis, ratio results were transformed by log (10) and divided into 100 bins (loci having the highest T/N methylation ratio at far right on the X-axis). The number of loci at each bin is plotted on the Y-axis. Of 8374 CpG array loci, 5334 loci were found

more highly methylated in the cancer group compared to normal breast and 3042 loci were found more highly methylated in the normal breast group compared to cancer.

Supplementary Figure 2: Loci Specifically Hypermethylated in ER subtypes of Breast Cancer. Individual samples are shown on the X axis, methylation (β-value) is shown on the Y-axis for the discovery set (n = 103 tumors; left) and the validation set (n = 50 tumors; right). ER-negative and ER-positive samples are as indicted. This figure is a continuation of manuscript Figure 3.

Supplementary Figure 3: Additional CpG Loci Associated with Breast Cancer Recurrence. CpG loci were identified which were associated with a higher rate of recurrence when methylated in primary tumors. Kaplan-Meier plots indicate the 32 CpG loci showing the strongest associations between hypermethylation and high rate of disease progression among 82 tumors. Recurrence was observed as distant metastases and/or death from breast cancer at least 6 months after and within 5 years of diagnosis for 7/41 ER-positive (dashed line), and 11/41 ER negative tumors (solid line). High (red line) and low (blue line) methylation was defined relative to the median methylation level for a given CpG loci within ER subtype. This figure accompanies Figures 4A and 4B. Since the differential methylation analysis was designed in such a way to find loci most highly methylated in recurrent tumors, we did not observe hypomethylated loci associating with recurrence.

SUPPLEMENTARY REFERENCES

1. Smyth GK. Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments. Statistical Applications in Genetics and Molecular Biology. 2004;3:1-25.

2. Fackler MJ, Rivers A, Teo WW, Mangat A, Taylor E, Zhang Z, et al. Hypermethylated genes as biomarkers of cancer in women with pathologic nipple discharge. Clinical cancer research : an official journal of the American Association for Cancer Research. 2009;15:3802-11.

3. Swift-Scanlan T, Blackford A, Argani P, Sukumar S, Fackler MJ. Two-color quantitative multiplex methylation-specific PCR. Biotechniques. 2006;40:210-9.

4. Fackler MJ, McVeigh M, Mehrotra J, Blum MA, Lange J, Lapides A, et al. Quantitative multiplex methylation-specific PCR assay for the detection of promoter hypermethylation in multiple genes in breast cancer. Cancer Res. 2004;64:4442-52.

5. Dan S, Shirakawa M, Mukai Y, Yoshida Y, Yamazaki K, Kawaguchi T, et al. Identification of candidate predictive markers of anticancer drug sensitivity using a panel of human cancer cell lines. Cancer science. 2003;94:1074-82.

6. Lefrancois-Martinez AM, Bertherat J, Val P, Tournaire C, Gallo-Payet N, Hyndman D, et al. Decreased expression of cyclic adenosine monophosphate-regulated aldose reductase (AKR1B1) is associated with malignancy in human sporadic adrenocortical tumors. The Journal of clinical endocrinology and metabolism. 2004;89:3010-9.

7. Dietrich D, Krispin M, Dietrich J, Fassbender A, Lewin J, Harbeck N, et al. CDO1 promoter methylation is a biomarker for outcome prediction of anthracycline treated, estrogen receptor-positive, lymph node-positive breast cancer patients. BMC Cancer. 2010;10:247.

8. Morris MR, Ricketts C, Gentle D, Abdulrahman M, Clarke N, Brown M, et al. Identification of candidate tumour suppressor genes frequently methylated in renal cell carcinoma. Oncogene. 2010;29:2104-17.

9. Paulus W, Baur I, Schuppan D, Roggendorf W. Characterization of integrin receptors in normal and neoplastic human brain. Am J Pathol. 1993;143:154-63.

10. Vachani A, Nebozhyn M, Singhal S, Alila L, Wakeam E, Muschel R, et al. A 10-gene classifier for distinguishing head and neck squamous cell carcinoma and lung squamous cell carcinoma. Clinical cancer research : an official journal of the American Association for Cancer Research. 2007;13:2905-15.

11. Pan SH, Chao YC, Chen HY, Hung PF, Lin PY, Lin CW, et al. Long form collapsin response mediator protein-1 (LCRMP-1) expression is associated with clinical outcome and lymph node metastasis in non-small cell lung cancer patients. Lung Cancer. 2010;67:93-100.

12. Fu DY, Wang ZM, Wang BL, Chen L, Yang WT, Shen ZZ, et al. Frequent epigenetic inactivation of the receptor tyrosine kinase EphA5 by promoter methylation in human breast cancer. Hum Pathol. 2010;41:48-58.

13. Zou H, Osborn NK, Harrington JJ, Klatt KK, Molina JR, Burgart LJ, et al. Frequent methylation of eyes absent 4 gene in Barrett's esophagus and esophageal adenocarcinoma. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2005;14:830-4.

14. Oster B, Thorsen K, Lamy P, Wojdacz TK, Hansen LL, Birkenkamp-Demtroder K, et al. Identification and validation of highly frequent CpG island hypermethylation in colorectal adenomas and carcinomas. International journal of cancer. 2011;129.

15. Peng DF, Razvi M, Chen H, Washington K, Roessner A, Schneider-Stock R, et al. DNA hypermethylation regulates the expression of members of the Mu-class glutathione S-transferases and glutathione peroxidases in Barrett's adenocarcinoma. Gut. 2009;58:5-15.

16. Starkova J, Zamostna B, Mejstrikova E, Krejci R, Drabkin HA, Trka J. HOX gene expression in phenotypic and genotypic subgroups and low HOXA gene expression as an adverse prognostic factor in pediatric ALL. Pediatr Blood Cancer. 2010;55:1072-82.

17. Tommasi S, Karm DL, Wu X, Yen Y, Pfeifer GP. Methylation of homeobox genes is a frequent and early epigenetic event in breast cancer. Breast Cancer Res. 2009;11:R14.

18. Rodriguez BA, Cheng AS, Yan PS, Potter D, Agosto-Perez FJ, Shapiro CL, et al. Epigenetic repression of the estrogen-regulated Homeobox B13 gene in breast cancer. Carcinogenesis. 2008;29:1459-65.

19. Ghoshal K, Motiwala T, Claus R, Yan P, Kutay H, Datta J, et al. HOXB13, a target of DNMT3B, is methylated at an upstream CpG island, and functions as a tumor suppressor in primary colorectal tumors. PLoS One. 2010;5:e10338.

20. Ma XJ, Salunga R, Dahiya S, Wang W, Carney E, Durbecq V, et al. A five-gene molecular grade index and HOXB13:IL17BR are complementary prognostic factors in early stage breast cancer. Clin Cancer Res. 2008;14:2601-8.

21. Sgroi DC. The HOXB13:IL17BR gene-expression ratio: a biomarker providing information above and beyond tumor grade. Biomark Med. 2009;3:99-102.

22. Kim YH, Lee HC, Kim SY, Yeom YI, Ryu KJ, Min BH, et al. Epigenomic Analysis of Aberrantly Methylated Genes in Colorectal Cancer Identifies Genes Commonly Affected by Epigenetic Alterations. Ann Surg Oncol. 2011.

23. Kamalakaran S, Varadan V, Giercksky Russnes HE, Levy D, Kendall J, Janevski A, Riggs M, Banerjee N, Synnestvedt M, Schlichting E, Kåresen R, Shama Prasada K, Rotti H, Rao R, Rao L, Eric Tang MH, Satyamoorthy K, Lucito R, Wigler M, Dimitrova N, Naume B,

Borresen-Dale AL, Hicks JB. DNA methylation patterns in luminal breast cancers differ from non-luminal subtypes and can identify relapse risk independent of other clinical variables. Mol Oncol. 2011 Feb;5(1):77-92.Epub 2010 Dec 2. PubMed PMID: 21169070.

24. Tong WG, Wierda WG, Lin E, Kuang SQ, Bekele BN, Estrov Z, et al. Genome-wide DNA methylation profiling of chronic lymphocytic leukemia allows identification of epigenetically repressed molecular pathways with clinical impact. Epigenetics. 2010;5:499-508.

25. Sato N, Fukushima N, Hruban RH, Goggins M. CpG island methylation profile of pancreatic intraepithelial neoplasia. Mod Pathol. 2008;21:238-44.

26. Herbst A, Rahmig K, Stieber P, Philipp A, Jung A, Ofner A, et al. Methylation of NEUROG1 in Serum Is a Sensitive Marker for the Detection of Early Colorectal Cancer. Am J Gastroenterol. 2011.

27. Lai HC, Lin YW, Huang TH, Yan P, Huang RL, Wang HC, et al. Identification of novel DNA methylation markers in cervical cancer. Int J Cancer. 2008;123:161-7.

28. Rahmatpanah FB, Carstens S, Guo J, Sjahputera O, Taylor KH, Duff D, et al. Differential DNA methylation patterns of small B-cell lymphoma subclasses with different clinical behavior. Leukemia. 2006;20:1855-62.

29. Katoh M, Katoh M. Integrative genomic analyses of ZEB2: Transcriptional regulation of ZEB2 based on SMADs, ETS1, HIF1alpha, POU/OCT, and NF-kappaB. Int J Oncol. 2009;34:1737-42.

30. Fevre-Montange M, Champier J, Szathmari A, Wierinckx A, Mottolese C, Guyotat J, et al. Microarray analysis reveals differential gene expression patterns in tumors of the pineal region. J Neuropathol Exp Neurol. 2006;65:675-84.

31. Matsubara S, Take M, Pedraza C, Muramatsu T. Mapping and characterization of a retinoic acid-responsive enhancer of midkine, a novel heparin-binding growth/differentiation factor with neurotrophic activity. J Biochem. 1994;115:1088-96.

32. Dalgin GS, Drever M, Williams T, King T, DeLisi C, Liou LS. Identification of novel epigenetic markers for clear cell renal cell carcinoma. J Urol. 2008;180:1126-30.

33. Aruga J, Yokota N, Mikoshiba K. Human SLITRK family genes: genomic organization and expression profiling in normal brain and brain tumor tissue. Gene. 2003;315:87-94.

34. Fang F, Flegler AJ, Du P, Lin S, Clevenger CV. Expression of cyclophilin B is associated with malignant progression and regulation of genes implicated in the pathogenesis of breast cancer. Am J Pathol. 2009;174:297-308.

35. Singer S, Malz M, Herpel E, Warth A, Bissinger M, Keith M, et al. Coordinated expression of stathmin family members by far upstream sequence element-binding protein-1 increases motility in non-small cell lung cancer. Cancer Res. 2009;69:2234-43.

36. Jin Z, Cheng Y, Gu W, Zheng Y, Sato F, Mori Y, et al. A multicenter, double-blinded validation study of methylation biomarkers for progression prediction in Barrett's esophagus. Cancer Res. 2009;69:4112-5.