

# Convergence of Mutation and Epigenetic Alterations Identifies Common Genes in Cancer That Predict for Poor Prognosis

Timothy A. Chan<sup>1,2\*</sup>, Sabine Glockner<sup>1,3</sup>, Joo Mi Yi<sup>1</sup>, Wei Chen<sup>1,4</sup>, Leander Van Neste<sup>5</sup>, Leslie Cope<sup>6</sup>, James G. Herman<sup>1</sup>, Victor Velculescu<sup>7,8</sup>, Kornel E. Schuebel<sup>1\*</sup>, Nita Ahuja<sup>1,3\*</sup>, Stephen B. Baylin<sup>1,4\*</sup>

**1** Cancer Biology Program, The Johns Hopkins Kimmel Cancer Center, Baltimore, Maryland, United States of America, **2** Department of Radiation Oncology and Molecular Radiation Sciences, The Johns Hopkins Kimmel Cancer Center, Baltimore, Maryland, United States of America, **3** Department of Surgery, Johns Hopkins Hospital, Baltimore, Maryland, United States of America, **4** Predoctoral Training Program in Human Genetics, The Johns Hopkins University, Baltimore, Maryland, United States of America, **5** Department of Molecular Biotechnology, Ghent University, Ghent, Belgium, **6** Biometry and Clinical Trials Division, The Johns Hopkins Kimmel Cancer Center, Baltimore, Maryland, United States of America, **7** Ludwig Center for Cancer Genetics and Therapeutics, The Johns Hopkins Kimmel Cancer Center, Baltimore, Maryland, United States of America, **8** Howard Hughes Medical Institute at The Johns Hopkins Kimmel Cancer Center, Baltimore, Maryland, United States of America

**Funding:** This work was supported by National Institute of Environmental Health Sciences grant ES11858 and National Cancer Institute grant CA043318. Funding institutions had no role in study design, data collection, data analysis, decision to publish, or preparation of manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

**Academic Editor:** Adi F. Gazdar, University of Texas Southwestern Medical Center, United States of America

**Citation:** Chan TA, Glockner S, Yi JM, Chen W, Van Neste L, et al. (2008) Convergence of mutation and epigenetic alterations identifies common genes in cancer that predict for poor prognosis. *PLoS Med* 5(5):e114. doi:10.1371/journal.pmed.0050114

**Received:** September 12, 2007

**Accepted:** April 15, 2008

**Published:** May 27, 2008

**Copyright:** © 2008 Chan et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Abbreviations:** CAN, candidate cancer gene; DAC, 5-aza-deoxycytidine; DNMT, DNA methyltransferase; LOH, loss of heterozygosity; MSP, methylation-specific PCR; RT-PCR, reverse-transcription PCR; TSA, trichostatin A

\* To whom correspondence should be addressed. E-mail: sbaylin@jhmi.edu (SBB); kornels@jhmi.edu (KES); nahuja@jhmi.edu (NA)

\* Current address: Human Oncology and Pathogenesis Program and Department of Radiation Oncology, Memorial Sloan Kettering Cancer Center, New York, New York, United States of America

## ABSTRACT

### Background

The identification and characterization of tumor suppressor genes has enhanced our understanding of the biology of cancer and enabled the development of new diagnostic and therapeutic modalities. Whereas in past decades, a handful of tumor suppressors have been slowly identified using techniques such as linkage analysis, large-scale sequencing of the cancer genome has enabled the rapid identification of a large number of genes that are mutated in cancer. However, determining which of these many genes play key roles in cancer development has proven challenging. Specifically, recent sequencing of human breast and colon cancers has revealed a large number of somatic gene mutations, but virtually all are heterozygous, occur at low frequency, and are tumor-type specific. We hypothesize that key tumor suppressor genes in cancer may be subject to mutation or hypermethylation.

### Methods and Findings

Here, we show that combined genetic and epigenetic analysis of these genes reveals many with a higher putative tumor suppressor status than would otherwise be appreciated. At least 36 of the 189 genes newly recognized to be mutated are targets of promoter CpG island hypermethylation, often in both colon and breast cancer cell lines. Analyses of primary tumors show that 18 of these genes are hypermethylated strictly in primary cancers and often with an incidence that is much higher than for the mutations and which is not restricted to a single tumor-type. In the identical breast cancer cell lines in which the mutations were identified, hypermethylation is usually, but not always, mutually exclusive from genetic changes for a given tumor, and there is a high incidence of concomitant loss of expression. Sixteen out of 18 (89%) of these genes map to loci deleted in human cancers. Lastly, and most importantly, the reduced expression of a subset of these genes strongly correlates with poor clinical outcome.

### Conclusions

Using an unbiased genome-wide approach, our analysis has enabled the discovery of a number of clinically significant genes targeted by multiple modes of inactivation in breast and colon cancer. Importantly, we demonstrate that a subset of these genes predict strongly for poor clinical outcome. Our data define a set of genes that are targeted by both genetic and epigenetic events, predict for clinical prognosis, and are likely fundamentally important for cancer initiation or progression.

*The Editors' Summary of this article follows the references.*

## Introduction

It is widely accepted that loss of tumor suppressor function leads to the initiation and progression of human cancer [1,2]. Inactivation of tumor suppressor genes can result from both genetic mechanisms such as mutation or epigenetic mechanisms such as DNA hypermethylation [3–5]. Identification of these genes provides insight into the biological processes underlying oncogenesis and is useful for developing new therapeutic and diagnostic modalities. Recently, several efforts to examine the cancer genome utilizing large-scale sequencing have revealed that a large number of genes undergo somatic mutation in cancer [6,7]. Sjoblom et al. sequenced 13,023 human genes in breast and colon cancer and identified 1,149 that harbored somatic mutations. Through statistical analysis, they showed that the majority of these changes were passenger mutations and that 189 genes were likely selected for during tumorigenesis (candidate cancer genes [CAN]). Interestingly, for virtually all of the newly discovered mutations, the frequencies in each tumor type were low—in the range of 5% to 15%. Furthermore, the vast majority of these mutations were heterozygous missense mutations. Thus, it is difficult to know whether each mutation conveys an oncogenic or tumor suppressor function. Moreover, if the genes are tumor suppressors, the heterozygous nature of the mutations would provide loss of function effects through a state of haploinsufficiency. This has been seen for a number of cancer genes including *APC* and *MSH2* [8,9]. It may also be possible that many of the heterozygous mutations are dominant and oncogenic. Similarly, Greenman et al. demonstrated that the mutational spectrum of protein kinases in tumors is highly variable and that mutations in a large number of cancer genes are operative in human tumors [6]. Again, it is unknown whether most of the mutated genes are oncogenes or tumor suppressors. Finally, most of the mutations identified in breast cancers were not present in colon tumors and vice versa [7], suggesting that the mutational spectrum is highly tumor-type specific.

Epigenetic silencing is a prevalent mechanism by which abnormal gene inactivation can occur in cancer. These epigenetic abnormalities can cooperate with genetic alterations to effect aberrant gene function that results in cancer [5]. A predominant mode of epigenetic alteration in cancer is gene silencing via CpG island promoter hypermethylation (henceforth called hypermethylation). Hypermethylation has now been observed to result in abnormal silencing of a number of tumor suppressors in many human malignancies such as cancers of the breast, prostate, colon, stomach, esophagous, blood, central nervous system, and lung [10]. Hypermethylation acts by recruiting methyl-cytosine-binding proteins and histone deacetylases, which in a coordinated fashion modify nucleosomes to form transcriptionally repressive chromatin [11,12]. Repressive histone marks such as methylation of lysine-9 on histone 3 (H3K9) may initiate and help maintain this state of repression [13,14]. Hypermethylation is heritable and thus constitutes a form of cellular memory. As such, abnormal silencing of tumor suppressor genes can help drive clonal selection during tumorigenesis.

Given the above unanticipated characteristics of the newly discovered mutations—especially their low frequency and high rate of heterozygosity—we performed an initial survey to characterize the epigenetic status of genes in colon cancer

on a genome-wide level. To this end, we developed an expression microarray technique to characterize the spectrum of hypermethylated genes in cancers [15]. In the present paper, we have utilized our global approach to comprehensively compare the epigenetic alteration of the CAN genes in both breast and colon cancers, and included analyses in the specific breast cancer lines where individual mutations were identified by Sjoblom et al. [7].

## Methods

### Cell Culture

MCF7, MDA-MB-231, MDA-MB-468, T-47D, HT-29, Caco-2, Colo320, SW480, RKO, and HCT116 cells and isogenic DNA methyltransferase (DNMT)1/3b genetic knockout derivatives were maintained in culture as recommended by American Type Tissue Culture (ATCC). All HCC series lines used were obtained from ATCC. For drug treatments, log phase cells were cultured in the appropriate media (Invitrogen) containing 10% FBS and 1× penicillin/streptomycin with 5 μM 5-aza-deoxycytidine (DAC) (Sigma; stock solution: 1 mM in PBS) for 96 h, replacing media and DAC every 24 h. Cell treatment with 300 nM Trichostatin A (Sigma; stock solution: 1.5 mM dissolved in ethanol) was performed for 18 h. Control cells underwent mock treatment in parallel with addition of equal volume of PBS or ethanol without drugs.

### Microarray Analysis

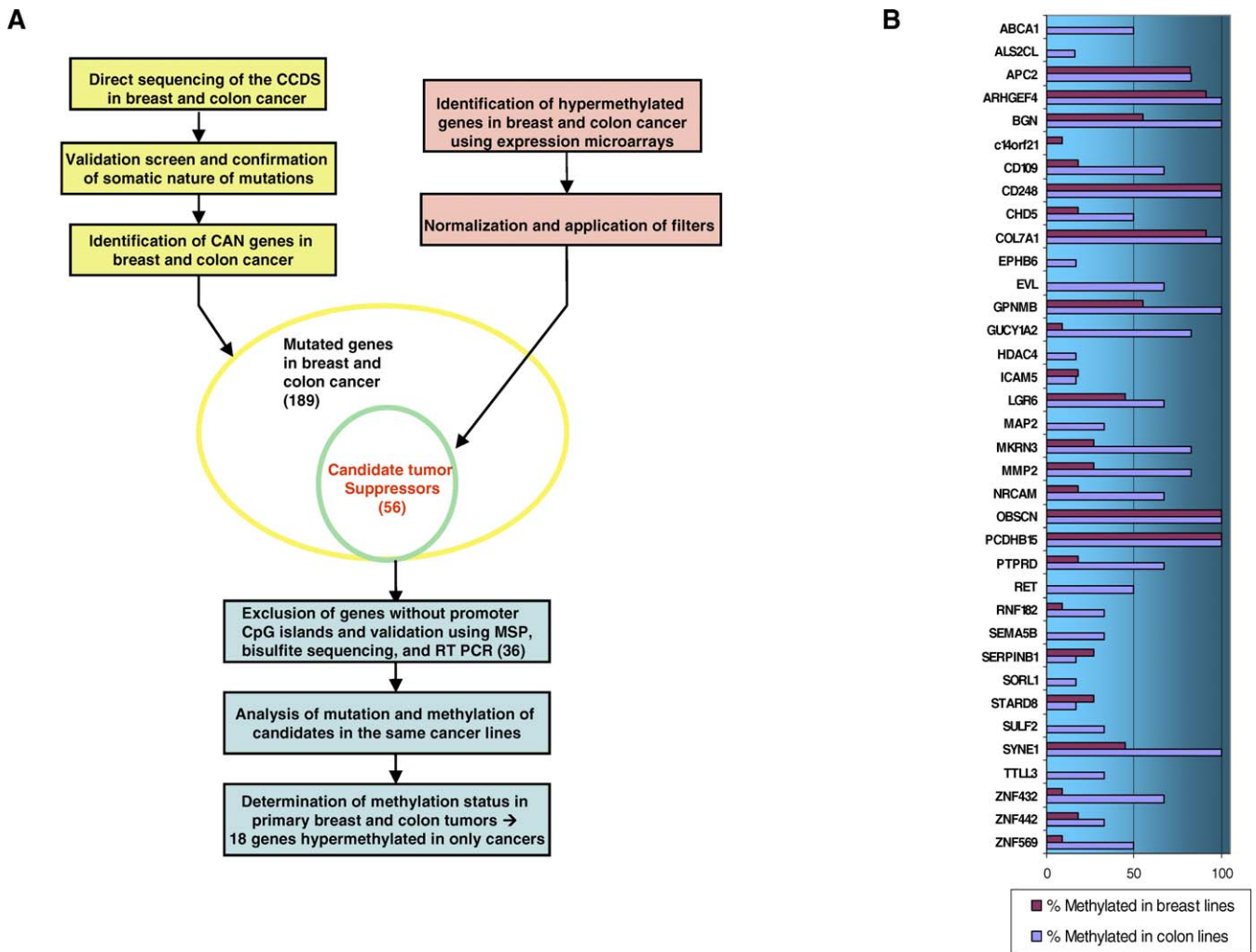
Total RNA was harvested from log phase cells using the Qiagen RNeasy kit according to the manufacturer's instructions. Sample amplification and labeling procedures were carried out using the Low RNA Input Fluorescent Linear Amplification kit (Agilent Technologies) according to the manufacturer's instructions. Hybridization was carried out according to the Agilent microarray protocol. Scanning was performed with the Agilent G2565BA microarray scanner.

### Data Analysis

All arrays were subject to quality checks recommended by the manufacturer. All calculations were performed using the R statistical computing platform [16] and packages from Bioconductor bioinformatics software project [17]. The log ratio of red signal to green signal was calculated after background-subtraction and LoEss normalization as implemented in the limma package from Bioconductor [18]. Individual arrays were scaled to have the same interquartile range (75th percentile to 25th percentile). Patient information, including clinical data and gene expression data, was obtained from and analyzed using Oncomine (<http://www.oncomine.org>). Our analysis included microarray databases such as the Netherlands Cancer Institute breast cancer database [19]. The microarray meta-analysis algorithms and statistical analysis used were as previously described [20,21]. *p*-Values were calculated using adjustment for multiple testing and false discovery as described at <http://www.oncomine.org> [21].

### Methylation and Gene Expression Analysis

RNA was isolated with TRIzol Reagent (Invitrogen) according to the manufacturer's instructions. For reverse transcription-PCR (RT-PCR), 1 μg of total RNA was reverse transcribed by using Ready-To-Go You-Prime First-Strand Beads (Amersham Biosciences) with addition of random hexamers (0.2 μg per reaction). Bisulfite modification of



**Figure 1.** Strategy to Identify Common Gene Targets of Mutation and Promoter Hypermethylation in Cancer

(A) Large-scale sequencing of breast and colon cancers identified 189 CAN genes as reported by Sjoblom et al. [7]. Candidate hypermethylated genes were identified via expression microarray analysis as described in the text. The cell lines used in the analyses were the breast cancer lines MCF7, MDA-MB-231, MDA-MB-468, and T-47D; and the colon cancer lines SW480, RKO, HCT116, Caco-2, Colo320, and HT-29. Filters were used as discussed in the text. Genes without promoter CpG islands were excluded.

(B) Frequency of promoter methylation in the 36 CAN genes that are subject to hypermethylation. x-Axis denotes percent methylation. Methylation status was determined in the six colorectal cancer cell lines and in the 15 breast cancer lines described in the text (four lines used for microarray analysis plus 11 “Discovery Phase” breast cancer lines).

doi:10.1371/journal.pmed.0050114.g001

genomic DNA was carried out using the EZ DNA methylation kit (Zymo Research). Selection of primers used for methylation-specific PCR (MSP) and determinants for CpG island localization and designation was accomplished using MSPPrimer (<http://www.mspprimer.org>) [22]. MSP was performed as previously described [23]. Primer sequences are listed in Table S3. Bisulfite sequencing and RT-PCR was performed as previously described [24]. Gene expression quantitation was performed using RT-PCR and the 1D software package (Kodak). For Table S1, decreased expression was defined as expression that was not detectable with RT-PCR or decreased by two-thirds compared to expression levels in normal tissue measured using the 1D software to quantitate bands. Quantitative PCR was performed using the Invitrogen SYBR Green qPCR kit according to the manufacturer’s instructions. Real-time PCR reactions were performed using the Mastercycler Realplex machine (Eppendorf). Readings were normalized using GAPDH.

## Human Tumor Analysis

Formalin-fixed, paraffin-embedded tissues from primary breast ( $n = 30$ ) and colorectal cancers ( $n = 20$ ) were obtained from the archive of the Department of Pathology of the Johns Hopkins Hospital. Analysis of breast tumors was performed on 12 early stage and 12 late stage primary cancers (Dataset S1). Ten stage 2 and ten stage 3 colon cancers were analyzed. Approval was obtained by the Medical Ethical Committee of Johns Hopkins Hospital. DNA was isolated using the Puregene DNA isolation kit (Gentra Systems). MSP analysis was performed as described above.

## Results

### Discovery of Common Targets of Mutation and Hypermethylation in Breast and Colon Cancer

We chose to examine breast and colon cancers because of their substantial epidemiological prevalence and clinical

**Table 1.** Genes Targeted by Mutation and Hypermethylation

| Gene            | Gene Name  | Accession Number | Mutated In | Methylated in Cell Lines | Comment  |
|-----------------|--|------------------|------------|--------------------------|--|
| <i>ABCA1</i>    | ATP-binding cassette, sub-family A, member 1                 | NM_005502.2      | CR         | CR                       | Not methylated in primary CR                         |
| <i>ALS2CL</i>   | ALS2 C-terminal-like   | NM_147129.2      | BR         | CR                       | Not methylated in primary BR, CR                     |
| <i>APC2</i>     | Adenomatosis polyposis coli 2                                | NM_005883.2      | BR         | BR, CR                   | Cancer-specific methylation                          |
| <i>ARHGEF4</i>  | Rho guanine nucleotide exchange factor 4                     | NM_015320.2      | BR         | BR, CR                   | Methylated in normal BR, CR                          |
| <i>BGN</i>      | Biglycan   | NM_001711.3      | BR         | BR, CR                   | Methylated in normal BR, CR                          |
| <i>C14orf21</i> | Chromosome 14 open reading frame 21                          | NM_174913.1      | BR         | BR                       | Cancer-specific methylation                          |
| <i>CD109</i>    | CD109 molecule   | NM_133493.2      | CR         | BR, CR                   | Cancer-specific methylation                          |
| <i>CD248</i>    | CD248 molecule, endosialin                                   | NM_020404.2      | CR         | BR, CR                   | Methylated in normal BR, CR                          |
| <i>COL7A1</i>   | Collagen, type VII, alpha 1                                  | NM_000094.2      | BR         | BR, CR                   | Cancer-specific methylation                          |
| <i>EPHB6</i>    | Ephrin receptor B6   | NM_004445.2      | CR         | CR                       | Not methylated in primary BR, CR                     |
| <i>EVL</i>      | Enah/Vasp-like   | NM_016337.2      | CR         | CR                       | Cancer-specific methylation                          |
| <i>GNPMB</i>    | Glycoprotein (transmembrane) nmb                             | NM_001005340.1   | BR         | BR, CR                   | Cancer-specific methylation                          |
| <i>GUCY1A2</i>  | Guanylate cyclase 1, soluble, alpha 2                        | NM_000855.1      | CR         | BR, CR                   | Cancer-specific methylation                          |
| <i>HDAC4</i>    | Histone deacetylase 4  | NM_006037.2      | BR         | CR                       | Not methylated in primary BR, CR                     |
| <i>ICAM5</i>    | Intercellular adhesion molecule 5, telencephalin             | NM_003259.2      | BR         | BR, CR                   | Cancer-specific methylation                          |
| <i>LGR6</i>     | Leucine-rich repeat-containing G protein-coupled Receptor 6  | NM_021636.2      | CR         | BR, CR                   | Cancer-specific methylation                          |
| <i>MAP2</i>     | Microtubule-associated protein 2                             | NM_002374.3      | CR         | CR                       | Not methylated in primary CR                         |
| <i>MKRN3</i>    | Makorin, ring finger protein 3                               | NM_005664.2      | CR         | BR, CR                   | Methylated in normal BR, CR                          |
| <i>MMP2</i>     | Matrix metalloproteinase 2                                   | NM_004530.2      | CR         | BR, CR                   | Cancer-specific methylation                          |
| <i>NRCAM</i>    | Neuronal cell adhesion molecule                              | NM_005010.3      | BR         | BR, CR                   | Cancer-specific methylation                          |
| <i>OBSCN</i>    | Obscurin   | NM_052843.2      | BR, CR     | BR, CR                   | Methylated in normal BR, CR                          |
| <i>PCDH15</i>   | Protocadherin beta 15  | NM_018935.2      | BR         | BR, CR                   | Methylated in normal BR, CR                          |
| <i>PTPRD</i>    | Protein tyrosine phosphatase, receptor-type, D               | NM_130391.2      | CR         | BR, CR                   | Cancer-specific methylation                          |
| <i>RET</i>      | Ret proto-oncogene   | NM_020975.4      | CR         | CR                       | Cancer-specific methylation                          |
| <i>RNF182</i>   | Ring finger protein 182                                      | NM_152737.2      | CR         | BR, CR                   | Cancer-specific methylation                          |
| <i>SEMA5B</i>   | Semaphorin 5B  | NM_001031702.2   | BR         | CR                       | Not methylated in primary BR, CR                     |
| <i>SERPINB1</i> | Serpin peptidase inhibitor, clade B, member 1                | NM_030666.2      | BR         | BR, CR                   | Not methylated in primary BR, CR                     |
| <i>SORL1</i>    | Sortilin-related receptor, L(DLR) class A-repeats-containing | NM_003105.3      | BR, CR     | CR                       | Not methylated in primary BR, CR                     |
| <i>STARDB8</i>  | START domain contain 8                                       | NM_014725.2      | BR         | BR, CR                   | Cancer-specific methylation                          |
| <i>SULF2</i>    | Sulfatase 2  | NM_018837.2      | BR         | CR                       | Not methylated in primary CR                         |
| <i>SYNE1</i>    | Spectrin repeat containing, nuclear envelope 1               | NM_182961.1      | CR         | BR, CR                   | Cancer-specific methylation                          |
| <i>TLL3</i>     | Tubulin tyrosine ligase-like family member 3                 | NM_015644.3      | CR         | CR                       | Not methylated in primary CR                         |
| <i>ZNF432</i>   | Zinc finger protein 432                                      | NM_014650.2      | BR         | BR, CR                   | Cancer-specific methylation                          |
| <i>ZNF442</i>   | Zinc finger protein 442                                      | NM_030824.2      | CR         | BR, CR                   | Methylated in normal BR not methylated in primary CR |
| <i>ZNF569</i>   | Zinc finger protein 569                                      | NM_152484.2      | BR         | BR, CR                   | Methylated in normal BR                              |

BR, breast; CR, colorectal.

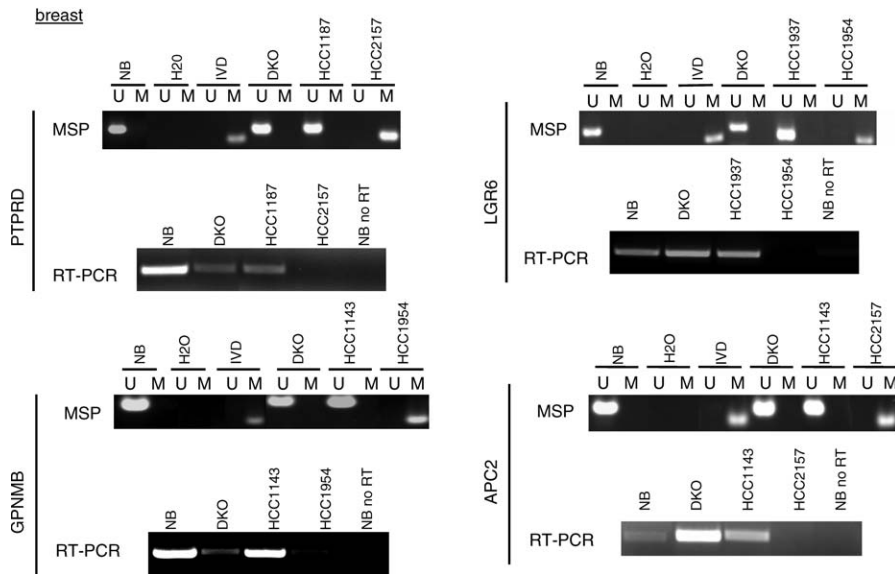
doi:10.1371/journal.pmed.0050114.t001

significance, and because extensive genome-wide mutational analysis has been conducted in these tumor types [7]. The strategy we utilized to identify the common gene targets of mutation and hypermethylation is depicted in Figure 1A. The utility of our microarray screen was previously validated and enables the identification of hypermethylated genes that are re-expressed following treatment with the DNMT inhibitor DAC, but not following treatment with the HDAC I/II inhibitor trichostatin A (TSA) alone [15,24]. Following drug treatments, the cells were subjected to microarray analysis, and we then searched for CAN genes that fell either in the top (expression change >2-fold after DAC treatment and <1.4-fold for TSA) or next tier (expression change >1.4-fold after DAC treatment and <1.4-fold following TSA). We assigned them status as candidate DNA hypermethylated genes to be analyzed further. Genes without promoter CpG islands were excluded from further analysis. This approach allows the identification of greater than 70% of the hypermethylated genes in cell lines with a false negative rate of 9% [15]. In total, we identified 56 (out of 189 CAN genes) that met our microarray criteria for candidate DNA hypermethylated

genes, with an approximately equal fraction originating from breast and colon lines.

We next analyzed in the laboratory, using MSP [23] and RT-PCR, the DNA methylation and expression status of the above 56 genes, including the response of the latter to treatment of cells with DAC, in four breast cancer cell lines (MCF7, MDA-MB-231, MDA-MB-468, and T-47D) and six colorectal cancer (CRC) cell lines (SW480, RKO, HCT116, Caco-2, Colo320, and HT-29). Importantly, we also focused, additionally, on the exact 11 breast cancer lines used in the large-scale sequence analysis reported by Sjoblom et al. to identify the CAN genes (HCC38, HCC1954, HCC1008, HCC1143, HCC1187, HCC1395, HCC1937, HCC2218, HCC2157, Hs578T, and HCC1599). Thirty-six of the genes were found to be hypermethylated in the original breast and colon cancer cell lines in which they were identified as candidates and in the 11 breast cancer lines in which the original CAN gene mutations were found (Figure 1B, Table 1). RT-PCR analysis demonstrated markedly reduced or no expression accompanying this hypermethylated state and re-expression after DAC treatment (examples in Figures 2 and 3 and summarized in Table S1).

From the analysis of methylation and expression status of



**Figure 2.** Common Gene Targets of Mutation and Hypermethylation in Breast Cancer

MSP and RT-PCR expression analysis of selected genes. Each set of MSP and expression analyses are labeled with the corresponding gene name. U denotes the unmethylated band, and M denotes the methylated band. In vitro methylated DNA (IVD) was used as a positive control for methylation. cDNA from normal breast (NB) and normal colon (NC) (Figure 3) was used to determine the expression of the genes in normal tissue. DKO corresponds to DNMT 1/3b double knockout HCT116 cells. All RT-PCR experiments were performed in parallel without reverse transcriptase as a control, and in all cases, no PCR product was generated (unpublished data). The control experiment performed without RT for normal breast is shown (NB no RT) for each gene. RT-PCR was performed for all cancer lines with beta-actin primers to control for the amount of cDNA (Figure 3, bottom). doi:10.1371/journal.pmed.0050114.g002

the 36 common target genes, it is clear that, although hypermethylation was accompanied by loss of gene expression in nearly all cases, loss of gene expression in the overlap genes can occur by mechanisms other than methylation (Table S1). Potential mechanisms include repressive chromatin modifications, mutational changes outside the coding regions that destabilize the mRNA, or coordinate downregulation of relevant pathways. For example, p21 expression is frequently decreased in tumors with inactivating p53 mutations [25].

To confirm our MSP results, we analyzed the methylation status of selected genes using sequencing of bisulfite-treated genomic DNA from samples that were used in the MSP studies. In all cases, bisulfite sequencing confirmed the results obtained with MSP (Figure 4). For another control, we also studied the methylation and expression status of the genes in a derivative of HCT116 human colon cancer cells in which the DNMT1 and DNMT3b DNA methyltransferases were homozygously deleted (DKO cells). These simultaneous deletions result in nearly complete lack of DNA methylation and the promoter DNA demethylation and re-expression of all known DNA hypermethylated genes examined in these cells [26]. In all cases, the loss or absence of methylation at the promoter regions of the 36 DNA hypermethylated loci in DKO cells was associated with expression of these genes (Figures 2 and 3 and unpublished data). Importantly, in all cases, these 36 overlap genes were also found to be expressed in normal colon and breast tissue (Figures 2 and 3 and Table S1).

### Hypermethylation of Common Target Genes is More Frequent Than Mutation and is Not Restricted to a Single Tumor Type

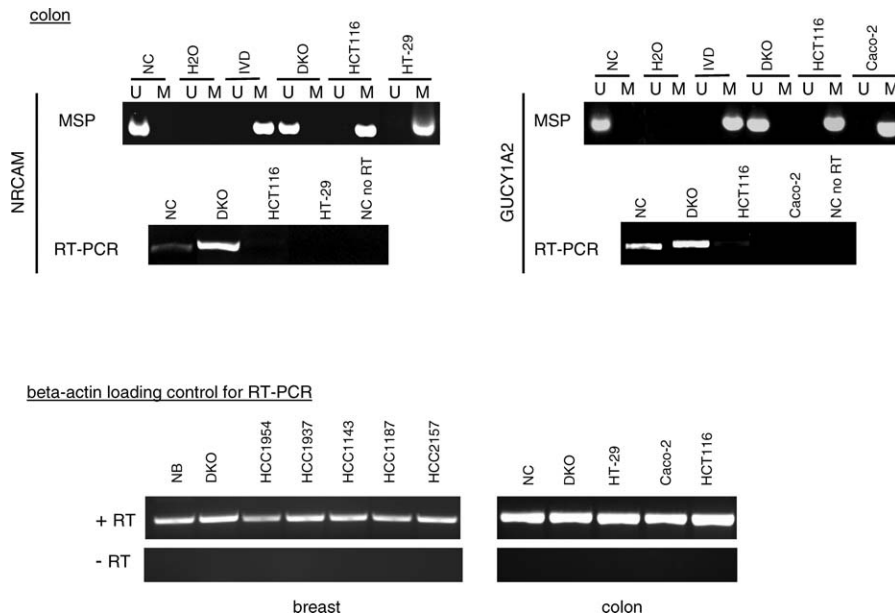
We determined whether the 36 common target genes were hypermethylated in primary colon and breast tumors. Eighteen of these genes were methylated in cell lines only, but not in primary breast or colorectal tumors, or were methylated in

normal breast or colon tissue. Importantly, the other 18 genes showed cancer-specific methylation—being hypermethylated in primary tumors but not in normal colon or breast tissue (Figure 5, Table 2). Among these 18 genes, the frequency of methylation (defined as the percentage of tumor samples that demonstrate promoter CpG island methylation) varied between breast and colon cancers. Some genes were methylated in breast but not colon cancers, while others were methylated in tumors from colon but not breast. Most importantly, however, while only 6% of all CAN genes (12/189) are mutated in both colon and breast cancers [7], of the 18 CAN genes with cancer-specific methylation, 44% were hypermethylated in both colon and breast tumors (8/18) (Table 2). This is a highly significant difference as determined by the chi-square test ( $p = 0.004$ ).

The above results suggest that when epigenetic silencing is taken into consideration, the biological alterations of a significant number of genes in breast and colon cancers may share more similarities than is apparent from mutational analysis alone. Figure 6 shows the number of genes (from among the 18 genes showing cancer-specific methylation) that are hypermethylated per tumor for both breast and colon cancers. A greater proportion of the genes are hypermethylated in colon cancer compared to breast cancer. It is highly likely that this represents an intrinsic difference between breast and colon tumors and not a statistical bias because approximately equal numbers of the genes found to be methylated were methylated in each tumor type.

### Concomitant Analysis of Genetic and Methylation Events

We next compared the genetic and epigenetic status of the 18 genes directly in the 11 “Discovery Phase” breast cancer lines in which the mutations for CAN genes were documented by Sjoblom et al. (Figure 7) [7]. Several observations are apparent. First, genes that are mutated in colon cancer but



**Figure 3.** Common Gene Targets of Mutation and Hypermethylation in Colon Cancer

MSP and RT-PCR expression analysis of selected genes. Each set of MSP and expression analyses are labeled with the corresponding gene name. U denotes the unmethylated band, and M denotes the methylated band. In vitro methylated DNA (IVD) was used as a positive control for methylation. cDNA from normal breast (NB) and normal colon (NC) was used to determine the expression of the genes in normal tissue. DKO corresponds to DNMT 1/3b double knockout HCT116 cells. All RT-PCR experiments were performed in parallel without reverse transcriptase as a control, and in all cases, no PCR product was generated (unpublished data). The control experiment performed without RT for normal colon is shown (NC no RT) for each gene. RT-PCR was performed for all cancer lines with beta-actin primers to control for the amount of cDNA. doi:10.1371/journal.pmed.0050114.g003

not in breast may instead be hypermethylated in breast cancer and vice versa. Furthermore, while the breast cancer mutations for a given gene are usually found in only one line, hypermethylation of most of the genes occurs in multiple lines for both breast and colon cancers (Figure 7A and 7B).

Second, within a single cancer, mutation and methylation of a specific gene are not always mutually exclusive events. For example, in HCC2157, the *APC2* gene undergoes both mutation (in one allele) and hypermethylation (of both alleles). In some cancers, partial methylation of genes such as *GPNMB* and *COL7A1* occur in the same cells harboring heterozygous mutations. Sequencing of cDNA from these cancers revealed they possess partial methylation on both mutant and wild-type alleles and not selective hypermethylation of the wild-type allele (unpublished data). We measured expression levels of *APC2*, *GPNMB*, and *COL7A1* using real-time PCR in these cell lines with partial methylation (as well as lines with unmethylated and fully methylated alleles of these genes) (Figure S1). In each case, higher levels of methylation were associated with decreasing levels of expression. It is possible that multiple events contribute to the incremental inactivation of the genes in these cancers. Alternatively, it is also possible that these genes have a limited role in cancer and are not under the same selective constraints as bona fide tumor suppressors.

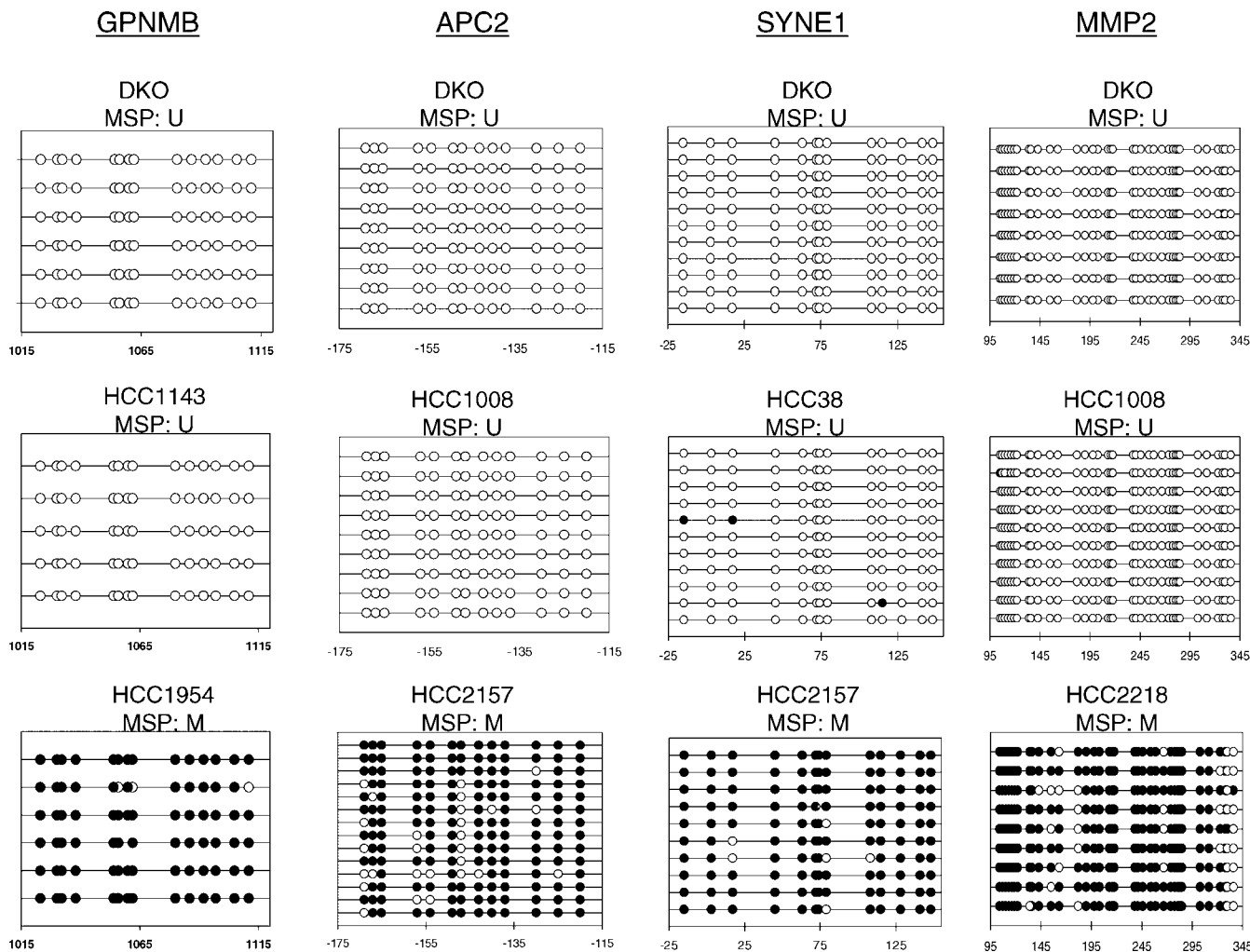
In previous studies, methylation has been observed to result in a functional loss of heterozygosity (LOH) for heterozygously mutated genes such as p16 where the mutated allele is expressed [27]. These results are consistent with the finding that, rarely, genetic alterations and methylation can converge on the same allele [27,28]. However, we did not observe this to be the case with the majority of the CAN genes

currently under study. We actually observed that for five out of 11 breast CAN genes, mutation and methylation converged in the same tumor (Figure 7A). Such convergence strongly suggests that these common target genes play an important role in tumor suppression, but that the methylation may complement the loss of function of these genes in a way different than with other mutations previously examined.

The above data may be particularly significant because the vast majority of CAN gene mutations identified are heterozygous missense mutations. It is possible that many of these mutations result in haploinsufficiency or cause small decreases in protein or transcript abundance that may be functionally meaningful for driving tumorigenesis. Small changes in expression of a number of tumor suppressor genes (such as *APC*, *SMAD4*, *MSH2*, etc.) have now been well described to have tumorigenic effects [8,29,30]. Consistent with the hemizygous nature of the mutations, it seems likely that DNA methylation deepens the haploinsufficiency status of the genes when both changes are found in the same tumor or, more often just with the epigenetic changes alone. The accompanying loss of function can compound, and thus progressively contribute to, tumorigenesis.

### Functional Associations of Common Target Genes

Next, we analyzed the functional associations of the 18 common target genes that demonstrate cancer-specific hypermethylation. We utilized Gene Ontology (GO) classification and available data in the literature to describe the functional associations of the genes (Figure 8). A number of genes are involved in cell adhesion and motility and/or signal transduction. Examination of the literature revealed that at least six of these genes possess known tumor suppressive



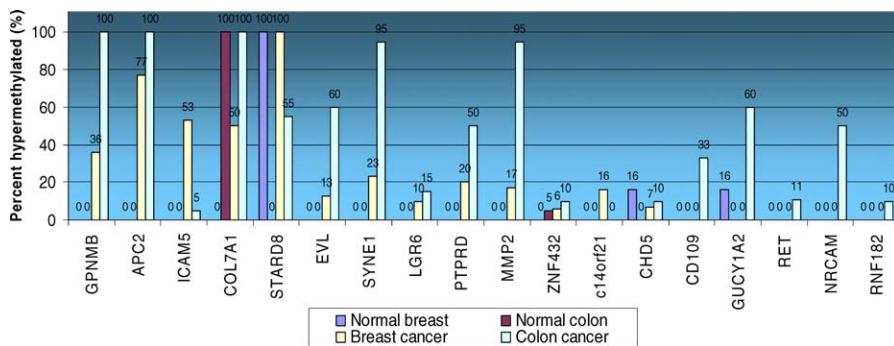
**Figure 4.** Bisulfite-Sequencing Results

Location relative to the transcriptional start site is shown on the x-axis of each plot. Empty circles represent unmethylated CpGs, and black circles represent methylated CpGs. The corresponding results obtained with MSP are noted above each plot.

doi:10.1371/journal.pmed.0050114.g004

properties, defined as the ability to modulate known tumor suppressor (i.e., p53, wnt, etc.) function or inhibit cancer cell growth in vitro and/or in vivo (Figure 8). We then determined the chromosomal location of these 18 genes. We compared these chromosomal locations to those that have been shown to be deleted in primary human tumors in the literature using

standard genetic mapping or comparative genomic hybridization (CGH). Sixteen out of 18 (89%, next to last panel) of these genes map to loci that have been found to be deleted in cancers (including colon, breast, prostate, Wilm's tumor, hematopoietic tumors, and medulloblastoma). For example, Yang et al. demonstrated that 19p13.3, the location of *APC2*,



**Figure 5.** Frequency of Cancer-Specific Methylation of CAN Genes in Primary Breast and Colon Tumors

For each gene, results are shown in the following order: normal breast, normal colon, breast cancer, colon cancer.

doi:10.1371/journal.pmed.0050114.g005

**Table 2.** Genetic Versus Epigenetic Events in Breast and Colon Cancer

| Gene            | Cancer-Specific <sup>a</sup> Methylation in Breast Cancer | Cancer-Specific <sup>a</sup> Methylation in Colon Cancer | Mutated in Breast Cancer | Mutated in Colon Cancer |
|-----------------|---|--|--------------------------|-------------------------|
| <i>APC2</i>     | ×   | ×  | ×                        | —                       |
| <i>GPNMB</i>    | ×   | ×  | ×                        | —                       |
| <i>ICAM5</i>    | ×   | ×  | ×                        | —                       |
| <i>EVL</i>      | ×   | ×  | —                        | ×                       |
| <i>SYNE1</i>    | ×   | ×  | —                        | ×                       |
| <i>LGR6</i>     | ×   | ×  | —                        | ×                       |
| <i>MMP2</i>     | ×   | ×  | —                        | ×                       |
| <i>PTPRD</i>    | ×   | ×  | —                        | ×                       |
| <i>COL7A1</i>   | ×   | —  | ×                        | —                       |
| <i>ZNF432</i>   | ×   | —  | ×                        | —                       |
| <i>c14orf21</i> | ×   | —  | ×                        | —                       |
| <i>CD109</i>    | —   | ×  | —                        | ×                       |
| <i>GUCY1A2</i>  | —   | ×  | —                        | ×                       |
| <i>RET</i>      | —   | ×  | —                        | ×                       |
| <i>RNF182</i>   | —   | ×  | —                        | ×                       |
| <i>CHD5</i>     | —   | ×  | ×                        | —                       |
| <i>NCAM</i>     | —   | ×  | ×                        | —                       |
| <i>STARD8</i>   | —   | ×  | ×                        | —                       |

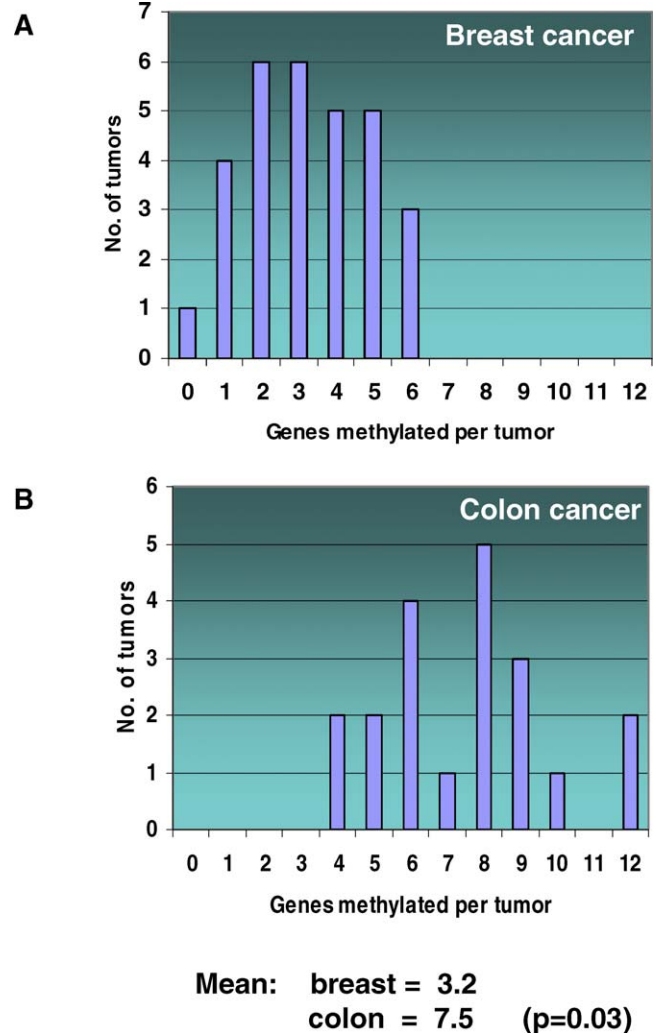
<sup>a</sup>Cancer-specific hypermethylation in primary tumors but not in normal tissue counterparts. X, positive detection of the type of event indicated. doi:10.1371/journal.pmed.0050114.t002

is a common site of LOH or deletion in breast carcinoma [31]. Similarly, *SYNE1* localizes to 6p25, a location that is subject to frequent deletion in a number of tumors [32,33]. *CHD5* is a well-documented tumor suppressor gene located on 1p36, a region that is commonly lost in malignancies of epithelial, neural, and hematopoietic origin [34]. Referenced data are summarized in Table S4. Thus, the 18 genes we have identified are genes that are found to be mutated in breast and colon cancer, silenced by hypermethylation in these tumors, and reside at locations subject to LOH or deletion in a number of human neoplasms.

#### Decreased Expression of Common Target Genes Predict for Poor Clinical Prognosis

One of the main hopes of comprehensively cataloging cancer mutations is that doing so may provide novel biomarkers and knowledge of genes involved in key pathways in oncogenesis. To this end, we first determined whether cancer-specific methylation of the common target genes would correlate in any way with tumor stage or grade. We determined this, first, by directly analyzing the methylation state of breast cancers of varying stages (1–4) and grades (1–3). We found that *SYNE1* and *COL7A1* are preferentially methylated in advanced tumors and *PTPRD*, *SYNE1*, and *EVL* are preferentially hypermethylated in high-grade tumors (Figure 9). For example, in stage 1 and 2 tumors, *SYNE1* is silenced 8% (1/12) of the time whereas in stage 3 and 4 tumors, the frequency of silencing is 50% (6/12). This is consistent with a role during tumor progression or during initiation of tumors predestined to evolve aggressive clinical behavior.

Given our results above, and considering that tumor stage and grade are strong prognostic determinants of disease-free survival and propensity for metastases in breast and colon cancer, we next sought to validate whether expression of the



**Figure 6.** Patterns of Mutation and Hypermethylation in Breast and Colon Cancer

(A) Frequency of cancer-specific hypermethylation of CAN genes. Data were generated from analysis of only the 18 genes demonstrating cancer-specific hypermethylation. Results for breast cancer. *p*-Value was calculated using Student's *t*-test.

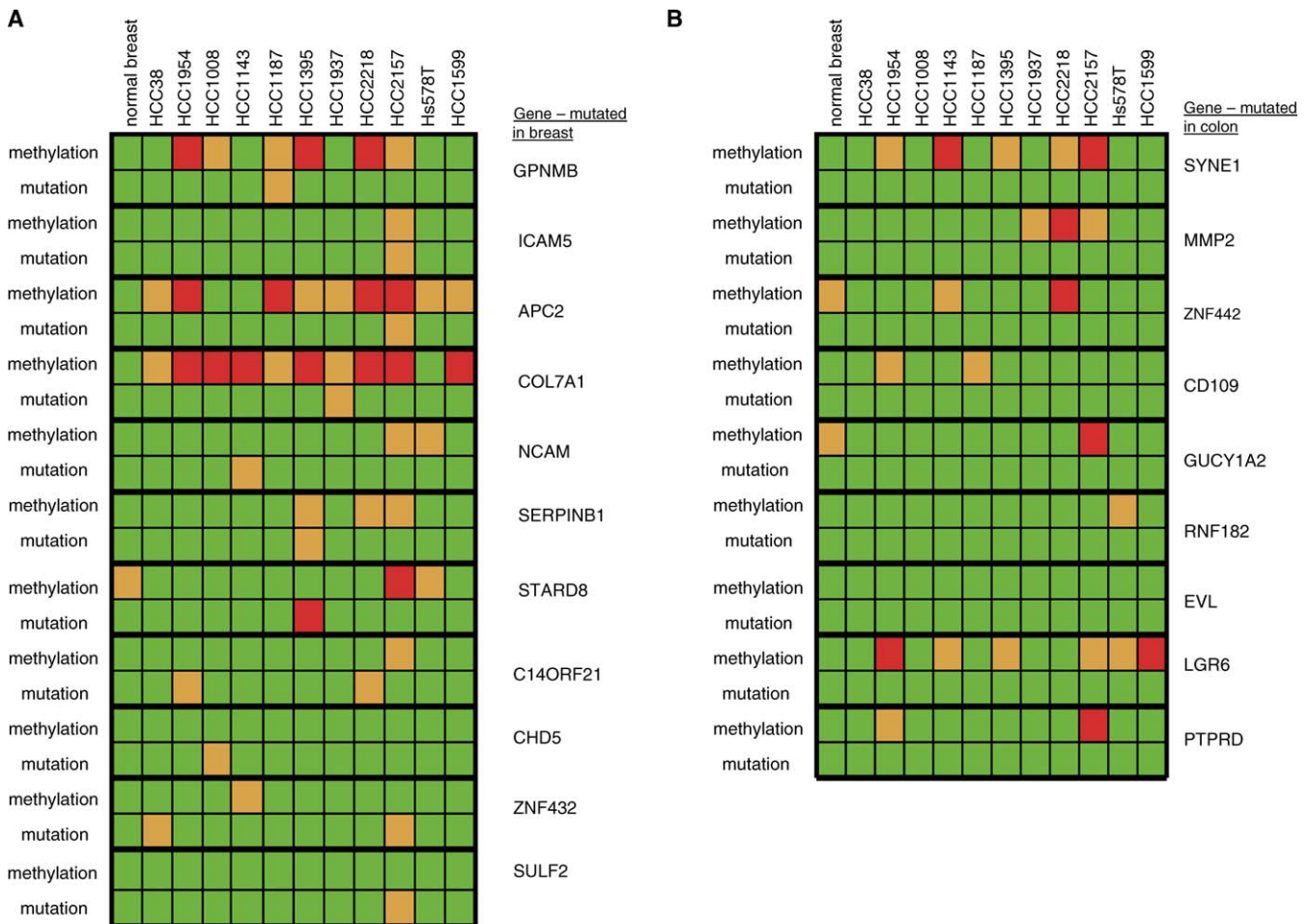
(B) Results for colon cancer.

doi:10.1371/journal.pmed.0050114.g006

genes we identified to be targets of hypermethylation and mutation affected clinical endpoints using data from external cohorts. Gene expression signatures from tumors have proven very useful for predicting clinical outcome [35,36]. To begin to address this question, we analyzed an extensive microarray database, utilizing large numbers of expression profiles on very well documented clinical samples from published expression microarray studies (Table S2). The microarray meta-analysis algorithms and statistical analysis used were as previously described [20]. These databases have been instrumental in a number of cancer gene discovery efforts [37–39].

We first verified whether we could see in the databases the key predicted relationship between DNA methylation and repressed gene expression. Unlike for gene mutations, which alone could indicate either oncogenic or tumor suppressor changes, the occurrence of hypermethylation suggests the latter in genes targeted by both mechanisms. We, thus, asked





**Figure 7.** Concomitant Mapping of Mutation and Methylation of CAN Genes in Human Breast and Colon Cancers

(A) Results for breast CAN genes. Gene names are listed at the right of the map. For each gene, the methylation and mutational status is shown for the 11 breast cell lines that were subject to large-scale sequencing described by Sjöblom et al. [7] and normal breast tissue. Green indicates no methylation or mutation detected. Red indicates complete methylation or homozygous mutation. Orange denotes heterozygous mutation or partial methylation. (B) Results for colon CAN genes. Data are presented in the same format as in (A).

doi:10.1371/journal.pmed.0050114.g007

whether genes undergoing a significant incidence of cancer-specific methylation correlated with decreased expression in tumor versus normal tissue. Genes undergoing cancer-specific methylation with low frequencies of methylation would not be predicted to have obvious gene expression correlations in the large database sets. We analyzed the following genes: *COL7A1*, *PTPRD*, *GPNMB*, *APC2*, *ICAM5*, *EVL*, *SYNE1*, and *MMP2*. All of these genes were predicted by our analysis of microarray data to have decreased overall expression in breast and/or colon cancer compared to normal tissue ( $p$ -values 0.047–2.9e<sup>-7</sup>) in the studies listed in Table S2. These in silico results are consistent with the observations we made with direct laboratory analyses.

We next examined whether decreased expression of the genes undergoing cancer-specific silencing correlated with the key clinical characteristics noted in Figure 10. The finding of decreased expression levels of seven genes is associated with unfavorable clinical characteristics in either breast, colon cancer, or both. Importantly, these genes included the four genes, *SYNE1*, *COL7A1*, *PTPRD*, and *EVL*, for which, in the studies described in Figure 9, we found relationships between stage and grade in the studies from our own tumor

samples at Johns Hopkins Hospital. Figures 11–13 show plots of normalized expression values for selected genes across multiple tumors with the indicated characteristics. Importantly, decreased expression of five of the six genes predicted for decreased disease-free or overall survival in these cancers as well as other poor prognosis features such as high grade (Figure 10). These relationships are highlighted by the fact that, when we also analyzed a number of CAN genes that we directly determined to not have altered methylation expression levels in breast or colon cancers (including *GGA1*, *PTPN14*, *ABCB8*, *OTOF*, *SIX4*, *SLCO1B3*, and *HIST1H1B*), the clinical endpoints we mentioned above were not associated with decreased expression of any of these genes.

Related to the above correlation with survival, decreased expression of five of the genes was seen in metastases when compared to primary tumors, such as *GPNMB*, *LGR6*, *EVL*, and, especially, *PTPRD*. Intriguingly, *GPNMB* encodes the glycoprotein nonmetastatic melanoma protein B, which has been shown to be differentially expressed between highly and lowly metastatic melanoma cell lines and xenografts. In these contexts, markedly lower expression levels characterize

## Functional associations of common target genes

| Gene     | Functional associations   | Location | Known cancer susceptibility loc. | Evidence of tumor suppressor function in literature |
|----------|---|----------|----------------------------------|---|
| APC2     | signal transduction, beta-catenin binding, regulator of wnt signaling                   | 19p13.3  |                                  |   |
| GPNMB    | signal transduction, negative regulator of cell proliferation, suppressor of metastasis | 7p15     |                                  |   |
| CD109    | signal transduction, decreased expression in breast and prostate vs. normal             | 6q13     |                                  |   |
| RET      | signal transduction, receptor tyrosine kinase, oncogene                                 | 10q11.2  |                                  |   |
| STARD8   | signal transduction, inhibits tumor growth, decreased expression in cancer vs. normal   | Xq13.1   |                                  |   |
| GUCY1A2  | signal transduction, guanylyl cyclase signaling   | 11q21-22 |                                  |   |
| PTPRD    | signal transduction, phosphatase  | 9p23-24  |                                  |   |
| RNF182   | signal transduction, transmembrane protein  | 6p23     |                                  |   |
| ICAM5    | cell adhesion and motility, integral to plasma membrane                                 | 19p13.2  |                                  |   |
| COL7A1   | cell adhesion and motility, basement membrane   | 3p21.1   |                                  |   |
| EVL      | cell adhesion and motility, cell polarity   | 14q32.2  |                                  |   |
| LGR6     | receptor, integral membrane protein, G-protein coupled signaling                        | 1q32.1   |                                  |   |
| NRCAM    | cell adhesion and motility  | 7q31.1   |                                  |   |
| MMP2     | cell adhesion and motility, matrix metalloproteinase-2                                  | 16q13    |                                  |   |
| CHD5     | regulation of transcription, chromodomain helicase                                      | 1p36     |                                  |   |
| ZNF432   | regulation of transcription, zinc-finger protein  | 19q13.3  |                                  |   |
| SYNE1    | nuclear membrane, intracellular trafficking   | 6q25     |                                  |   |
| C14orf21 | RNA binding, unknown function   | 14q12    |                                  |   |

All studies and data are referenced in Table S4

**Figure 8.** Functional Associations of Common Target Genes

Functional associations of genes were determined using Gene Ontology (GO) groups and available literature. Evidence for tumor suppressor function was scored positive if reports in the literature show that the gene of interest can suppress growth of cancer cells in vitro, in vivo, or in murine genetic models, or modulate known tumor suppressor function. loc., location. Known cancer susceptibility location was scored positive if evidence exists in the literature for LOH or homozygous deletion in primary human tumors at the region shown using either standard genetic mapping or comparative genomic hybridization (CGH) analysis (shaded). doi:10.1371/journal.pmed.0050114.g008

metastatic cells and overexpression of the GPNMB protein lowers metastatic potential [40].

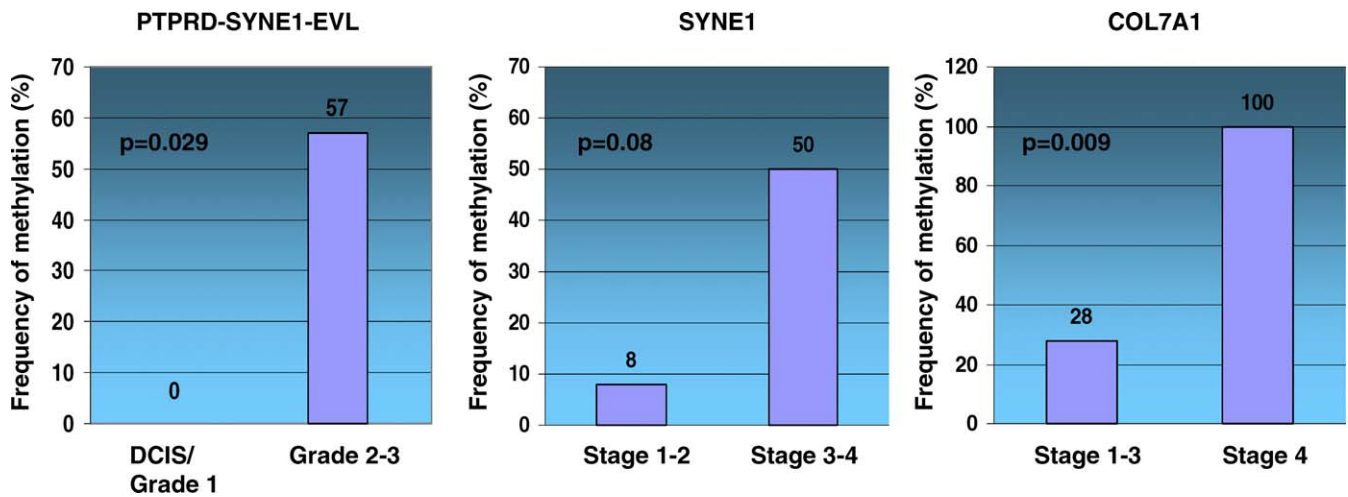
Finally, four genes are underexpressed with increasing tumor grade (*COL7A1*, *SYNE1*, *PTPRD*, and *EVL*). Since grade is a strong predictor of local recurrence and metastasis, silencing of these genes may be clinically relevant determinants of prognosis (reviewed in [41]). It is important to note that our direct analysis of tumor samples (Figure 9) is consistent with our analysis of microarray gene expression data (Figures 10–13) from these other cohorts. Figures 12 and 13 show the normalized expression levels of two of these genes, *SYNE1* and *EVL*. For each gene, expression is decreased with increasing grade in nearly all available datasets in the literature. This decreased expression parallels the greater frequency of methylation of these genes in tumors with increasing grade. *EVL* is intriguing as it is hypermethylated and silenced with increasing tumor grade and aggressiveness in primary breast tumors. However, we only observed it to be silenced in colon cell lines but not in the breast cell lines that we examined. It is possible that *EVL* methylation occurs in other breast lines we have not examined here and marks a small subset of aggressive tumors. All datasets used in the above microarray meta-analyses as well as details on the

previously published samples used are publicly available at [www.oncomine.org](http://www.oncomine.org).

## Discussion

Overall our study presents an extensive search for the presence of and interactions between both genetic and epigenetic alterations in cancer. As it currently stands, these studies do have several limitations. First, our data do not address the biological effects of the individual mutations observed in the CAN genes. Second, the data draw on only the 13,023 subset of CCDS genes that were previously sequenced, and additional genes have now been sequenced and more mutations have been discovered [42].

Despite these limitations, our study describes a valuable approach to begin to understand the biological significance of the vast amount of mutational data generated by cancer resequencing efforts. In these regards, our findings allow several important conclusions to be drawn. First, our study shows that large-scale, combined genetic and epigenetic analysis is feasible and useful for cancer gene discovery. Such combined analyses can markedly enhance links made between gene alterations and key clinical parameters for cancer. It is becoming increasingly clear that examination and interpre-



**Figure 9.** Hypermethylation of Selected “Common Target” Genes Predict for Poor Clinical Prognosis

Hypermethylation of selected common target genes by grade and stage in primary breast tumors. The first graph shows the frequency of methylation of at least one of the three genes listed in low- versus high-grade breast tumors. The second (SYNE1) and third (COL7A1) graphs show the frequency of methylation of the indicated genes in different stages of breast cancer. *p*-Values are adjusted using the Holm method. doi:10.1371/journal.pmed.0050114.g009

tation of mutations to identify cancer genes on a genome-wide scale can be significantly complicated by passenger mutations [43,44]. Furthermore, as we mentioned above in the Results, it is very unlikely that a given gene that in breast and/or colon cancer has evidence for mutations, promoter hypermethylation, reduced expression, and is localized to chromosome regions harboring frequent deletions in tumors is not important for tumor development. Consistent with this hypothesis, by beginning with a large pool of genes harboring mostly low incidence heterozygous missense mutations and then characterizing the methylation and expression status of these genes, our approach allowed us to identify genes that possess potentially prognostic value.

Our results also confirm that our microarray strategy is an effective approach to identify genes that are silenced by hypermethylation in colon and breast cancer. Other methods have been developed to identify hypermethylated genes in cancer, including restriction landmark genomic scanning, promoter CpG island microarrays, and methylation-specific

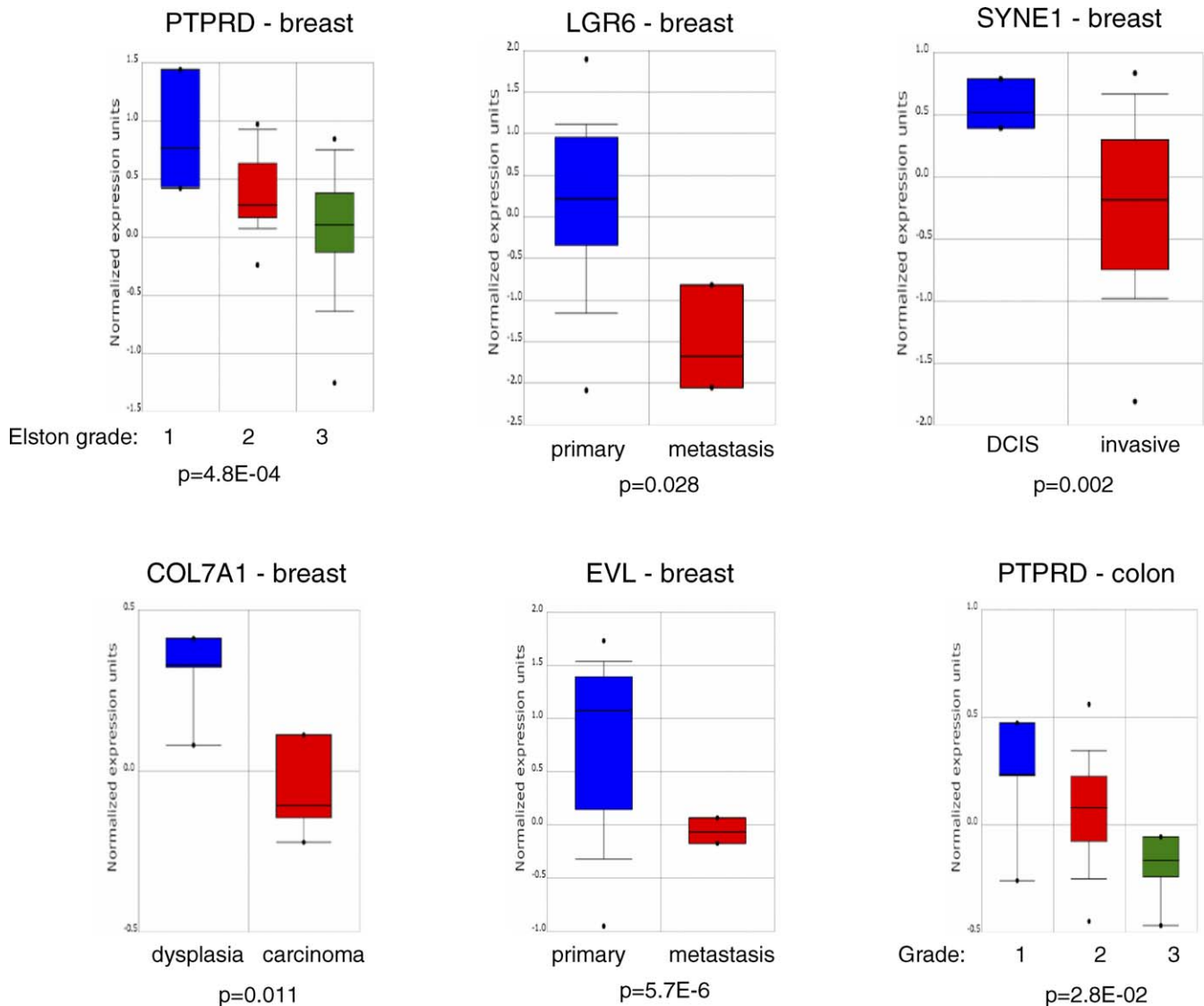
digital karyotyping [28,45]. However, the sensitivity of these techniques is restricted by the locations of methylation-sensitive restriction sites in the genome.

Several of the common target genes have been noted to undergo methylation-associated silencing in cancers by other investigators. Lund et al. noted that oncogenic *RAS* can lead to the hypermethylation of the *MMP2* gene [46]. *EVL* has been found to be hypermethylated in colon carcinoma [47]. *N-CAM* has been found to be hypermethylated in lung cancer in a survey of methylated genes described by Shames et al. [48]. The presence of methylation of the common target genes in other tumor types suggests that these genes may be targets of inactivation in a broader range of cancers, a hypothesis that warrants future investigation. In particular, it would be of value to directly compare our results with those derived by other strategies for analyzing the hypermethylome from the same as well as from other types of malignancies [48–51]. Together with these studies, our data strongly suggest that a

|        | GPNUMB             | COL7A1  | PTPRD      | SYNE1   | LGR6    | EVL       | RET     |
|--------|--------------------|---------|------------|---------|---------|-----------|---------|
| Breast | 5yr OS/DFS         |         | p=0.023    | p=0.009 | p=0.02  | p=5.2E-6  | p=8E-4  |
|        | Primary v. met     | p=0.017 | p=9.00E-03 |         | p=0.028 | p=5.7E-6  |         |
|        | Grade/invasiveness | p=0.011 | p=4.80E-04 | p=0.002 |         | p=7.1E-13 |         |
| Colon  | 5yr OS/DFS         |         |            |         |         |           |         |
|        | Primary v. met     |         | p=9.00E-03 |         |         |           | p=0.017 |
|        | Grade/invasiveness |         | p=0.028    |         |         |           |         |

**Figure 10.** Decreased Expression of Hypermethylated “Common Target” Genes Predict for Poor Clinical Prognosis

Summary of results from analysis of expression microarray data. *p*-Values from representative analyses for specific genes are shown. Blue indicates that a decrease in expression of the indicated gene is significantly associated with the following: overall survival or disease-specific survival less than 5 y, metastasis versus primary tumor, and/or increased grade or invasiveness. The *p*-value was calculated using the Student’s *t*-test with adjustment for multiple testing as described previously [21]. doi:10.1371/journal.pmed.0050114.g010



**Figure 11.** Decreased Expression of Selected “Common Target” Genes Predict for Poor Clinical Prognosis

Box-plots showing decreased expression of candidate genes correlate with unfavorable clinical features. The y-axis represents normalized expression. Shaded boxes represent the interquartile range (25th–75th percentile). Whiskers represent the 10th–90th percentile. The bars denote the median. doi:10.1371/journal.pmed.0050114.g011

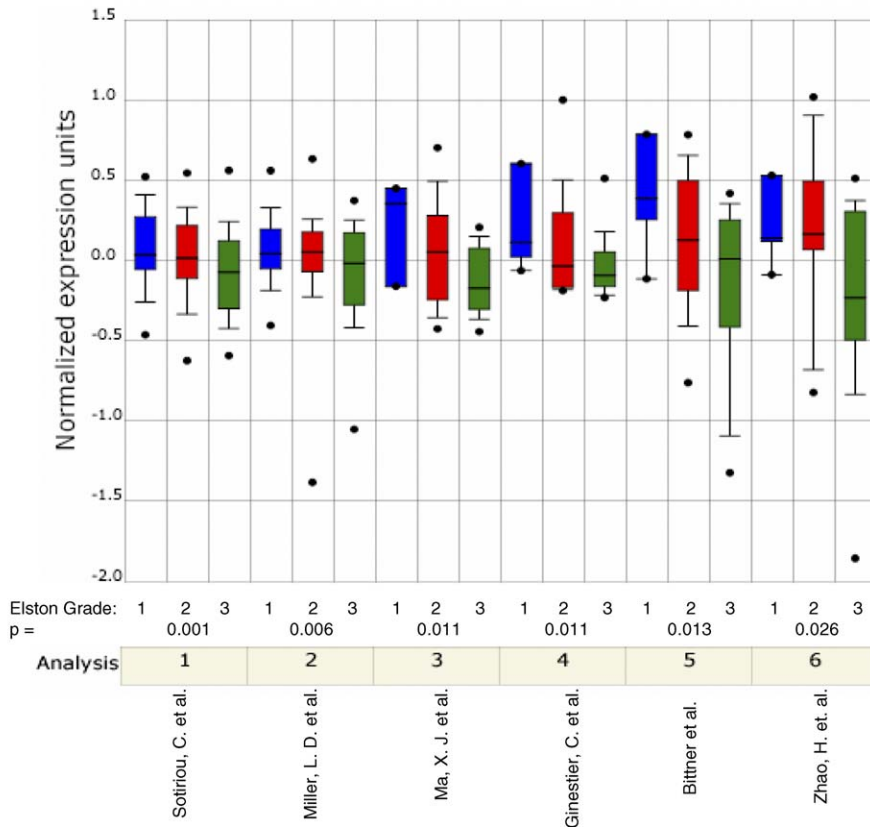
compendium of epigenetic changes underlie the progression of human cancers.

Second, our results suggest that tumors may be less biologically heterogeneous with respect to denoting key tumor suppressor pathway disruptions when consideration is given to both genetic and epigenetic changes. To our knowledge, this study represents the most comprehensive analysis of genes targeted by both mutation and hypermethylation. Prior to the present study, only a small number of genes had been found to be frequently affected by both mutations and promoter hypermethylation. Most of these genes were the initial classic tumor suppressor genes where the epigenetic event was first defined as meaningfully functional. These genes are closely linked to cancer initiation and include those for which germ-line mutations occur, such as *VHL*, *BRCA1*, and *STK11* in familial forms of renal, breast, and colon cancer, respectively [10,52,53]. These tumor suppressors are frequently hypermethylated in sporadic

forms of the corresponding tumor types [54–56]. Furthermore, methylation-associated silencing can act as a “second genetic hit” in these genes in tumors from individuals harboring germline mutations, resulting in functional LOH [57]. Our current findings now indicate that, particularly for tumor suppressor genes with a low incidence of mutations, it may be the rule rather than the exception that epigenetic inactivation is a more frequent event than genetic disruption. Tumor suppressors that are important for tumorigenesis may, then, often be targeted by multiple methods of functional inactivation.

A third important conclusion is that there may be more similarity among individual breast and colon tumors than is apparent from analysis of the mutational spectrum only, and, therefore, any comparison of biological changes between tumors may need to account for epigenetic effects in addition to genetic ones. Clearly, the same tumor suppressor genes in different cancers may undergo different modes of inactiva-

## SYNE1 - breast



**Figure 12.** *SYNE1* Expression Is Decreased with Increasing Tumor Grade in Breast Carcinoma

Representative data are shown across multiple independently published microarray studies as indicated. *p*-Values for correlation are shown below. References refer to those listed in Table S2. doi:10.1371/journal.pmed.0050114.g012

tion. This scenario is analogous to the situation that is observed for oncogenes such as *MYC*. In hematopoietic malignancies, aberrant activation of *MYC* results frequently from translocations whereas the gene is more often subject to amplifications and mutations in solid tumors [58,59]. The processes underlying these differences are fundamentally important for understanding cancer and are worthy of future study.

Finally, it is important to reiterate that our findings have allowed us to begin querying the clinical significance of genes targeted by mutation and hypermethylation. By correlating our data to expression changes in cancer microarray databases and relating these to important clinical parameters, we have identified genes that may track with disease prognosis. Indeed, previously, the discovery of hypermethylated genes such as *MGMT* have proven very useful for predicting clinical prognosis and response to therapy in diseases such as malignant glioma [60], gastric cancer [61], and lung cancer [62,63]. A recent study showed that a polycomb repression signature in metastatic prostate cancer predicts for cancer outcome [64]. Our study suggests that matching large-scale mutational and epigenetic analysis will be useful for advancing our knowledge of the biology of human cancers. These results may be useful for the development of new, more effective biomarkers and therapeutics.

## Supporting Information

**Dataset S1.** Clinical Summary of Tumor Samples Used in Methylation Analysis

All samples were processed from paraffin-embedded tissue and procured in accordance with IRB approval at Johns Hopkins Hospital. For each gene, red indicates the methylated state and green indicates the unmethylated state.

Found at doi:10.1371/journal.pmed.0050114.sd001 (27 KB XLS).

**Figure S1.** Increasing Levels of Hypermethylation Results in Decreased Gene Expression

The methylation state of *APC2*, *GPNMB*, and *COL7A1* was determined in the cell lines noted using MSP as described in the text. Real-time PCR was used to measure the expression levels of these genes in the same cell lines. Higher levels of methylation are associated with reduced gene expression.

Found at doi:10.1371/journal.pmed.0050114.sg001 (239 KB PPT).

**Table S1.** Frequency of Decreased Expression and Methylation of Genes Undergoing Cancer-Specific Hypermethylation in Breast and Colon Cell Lines

Found at doi:10.1371/journal.pmed.0050114.st001 (56 KB DOC).

**Table S2.** Microarray Reference Databases

Found at doi:10.1371/journal.pmed.0050114.st002 (57 KB DOC).

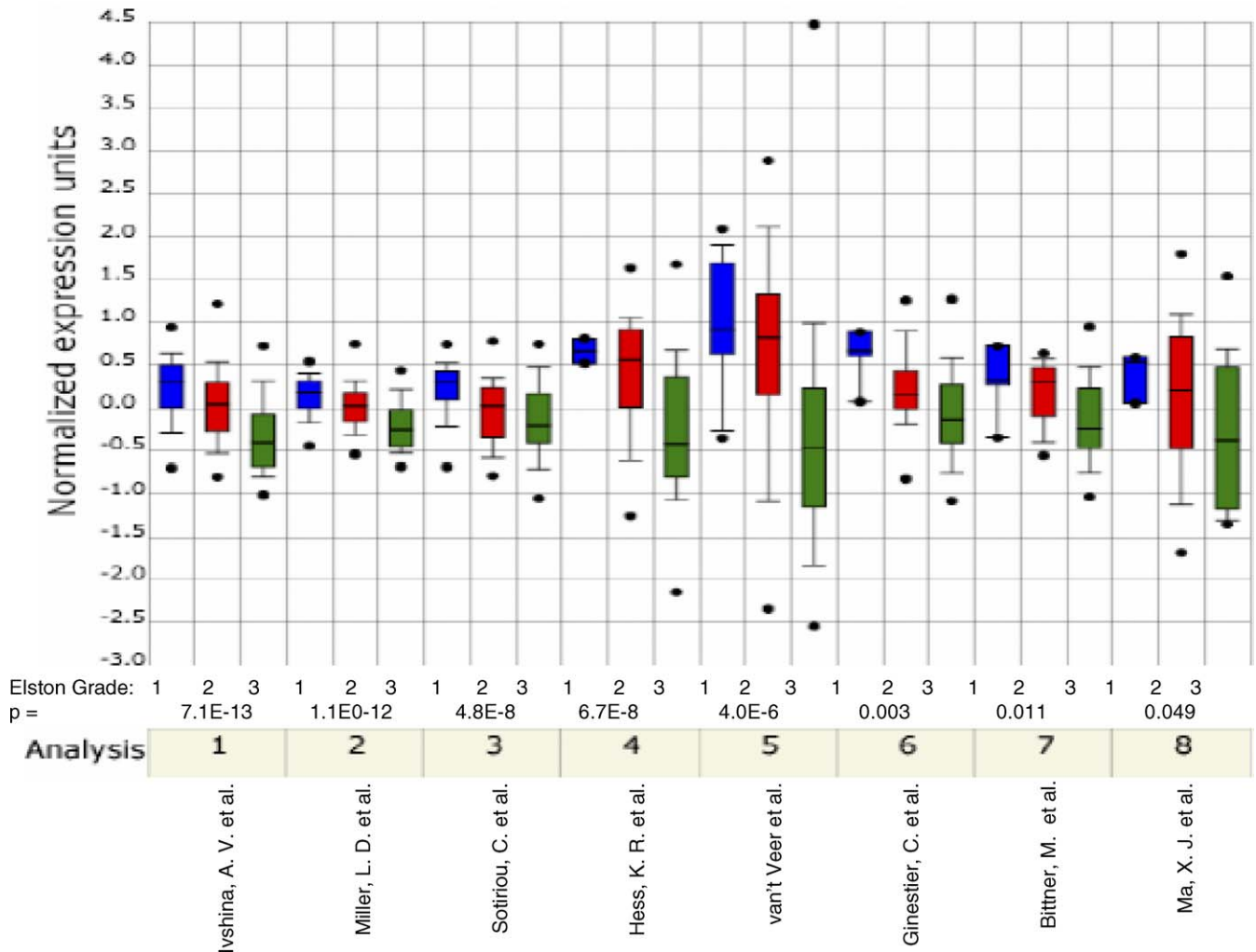
**Table S3.** Primers Used for MSP

Found at doi:10.1371/journal.pmed.0050114.st003 (94 KB DOC).

**Table S4.** Studies Referenced for Analysis in Figure 8

Found at doi:10.1371/journal.pmed.0050114.st004 (42 KB DOC).

## EVL - breast



**Figure 13.** EVL Expression Is Decreased with Increasing Tumor Grade in Breast Carcinoma

Representative data are shown across multiple independently published microarray studies as indicated.  $p$ -Values for correlation are shown below. References refer to those listed in Table S2.

doi:10.1371/journal.pmed.0050114.g013

#### Accession Numbers

Primary array data are deposited in the GEO database at the NCBI (<http://www.ncbi.nlm.nih.gov/geo/>). The accession numbers are as follows: GSM107602, GSM107603, GSM107604, GSM107605, GSM107606, GSM107607, GSM107660, GSM107662, GSM107663, GSM107664, GSM267289, GSM267290, GSM267459, GSM267460, GSM267461, GSM267462, GSM267463, GSM267464, GSM267830, GSM267831, GSM267832, GSM267833, GSM267834, GSM267835, GSM267836, GSM267837, GSM268000, GSM268001, GSM269500, GSM 269501.

The above files are linked under the composite series GSE4763 and GSE10613.

#### Acknowledgments

We wish to thank the members of the Baylin and Herman labs for helpful discussions. We thank Marco Reis for technical support. We thank Bert Vogelstein and Kenneth Kinzler for very helpful discussions and breast cancer DNA samples.

**Author contributions.** TAC and SBB designed the study. TAC and SG performed the experiments for the study. TAC, VV, KES, and SBB analyzed the data and interpreted the results. TAC, WC, LVN, LC, and

JGH contributed bioinformatics and statistical analysis to the research effort. TAC and SBB wrote the manuscript. NA enrolled patients and helped analyze the clinical outcomes data. JMY performed a portion of the bisulfite sequencing analysis of the common target.

#### References

- Vogelstein B, Kinzler KW (2004) Cancer genes and the pathways they control. *Nat Med* 10: 789–799.
- Futreal PA, Coin L, Marshall M, Down T, Hubbard T, et al. (2004) A census of human cancer genes. *Nat Rev Cancer* 4: 177–183.
- Ponder BA (2001) Cancer genetics. *Nature* 411: 336–341.
- Herman JG, Baylin SB (2003) Gene silencing in cancer in association with promoter hypermethylation. *N Engl J Med* 349: 2042–2054.
- Jones PA, Baylin SB (2007) The epigenomics of cancer. *Cell* 128: 683–692.
- Greenman C, Stephens P, Smith R, Dalgliesh GL, Hunter C, et al. (2007) Patterns of somatic mutation in human cancer genomes. *Nature* 446: 153–158.
- Sjoberg T, Jones S, Wood LD, Parsons DW, Lin J, et al. (2006) The consensus coding sequences of human breast and colorectal cancers. *Science* 314: 268–274.
- Yan H, Dobbie Z, Gruber SB, Markowitz S, Romans K, et al. (2002) Small changes in expression affect predisposition to tumorigenesis. *Nat Genet* 30: 25–26.
- Bouffler SD, Hofland N, Cox R, Fodde R (2000) Evidence for Msh2

- haploinsufficiency in mice revealed by MNU-induced sister-chromatid exchange analysis. *Br J Cancer* 83: 1291–1294.
10. Jones PA, Baylin SB (2002) The fundamental role of epigenetic events in cancer. *Nat Rev Genet* 3: 415–428.
  11. Busslinger M, Hurst J, Flavell RA (1983) DNA methylation and the regulation of globin gene expression. *Cell* 34: 197–206.
  12. Nan X, Ng HH, Johnson CA, Laherty CD, Turner BM, et al. (1998) Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* 393: 386–389.
  13. Jenuwein T (2006) The epigenetic magic of histone lysine methylation. *Febs J* 273: 3121–3135.
  14. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, et al. (2007) High-resolution profiling of histone methylations in the human genome. *Cell* 129: 823–837.
  15. Schuebel KE, Chen WE, Cope L, Glockner S, Suzuki H, et al. (2007) Comparing the DNA hypermethylation with gene mutations in human colorectal cancer. *PLoS Genet* 3: e157. doi:10.1371/journal.pgen.0030157.
  16. Ihaka R, Gentleman RC (1996) A language for data analysis and graphics. *Journal of Computational and Graphical Statistics* 5: 299–314.
  17. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, et al. (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 5: R80.
  18. Smyth GK, Speed T (2003) Normalization of cDNA microarray data. *Methods* 31: 265–273.
  19. Liu R, Wang X, Chen GY, Dalerba P, Gurney A, et al. (2007) The prognostic role of a gene signature from tumorigenic breast-cancer cells. *N Engl J Med* 356: 217–226.
  20. Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, et al. (2004) Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression. *Proc Natl Acad Sci U S A* 101: 9309–9314.
  21. Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, et al. (2004) ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia* 6: 1–6.
  22. Brandes JC, Carraway H, Herman JG (2007) Optimal primer design using the novel primer design program: MSPprimer provides accurate methylation analysis of the ATM promoter. *Oncogene* 26: 6229–6237.
  23. Herman JG, Graff JR, Myohanen S, Nelkin BD, Baylin SB (1996) Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proc Natl Acad Sci U S A* 93: 9821–9826.
  24. Suzuki H, Gabrielson E, Chen W, Anbazhagan R, van Engeland M, et al. (2002) A genomic screen for genes upregulated by demethylation and histone deacetylase inhibition in human colorectal cancer. *Nat Genet* 31: 141–149.
  25. Matsushita K, Kobayashi S, Kato M, Itoh Y, Okuyama K, et al. (1996) Reduced messenger RNA expression level of p21 CIP1 in human colorectal carcinoma tissues and its association with p53 gene mutation. *Int J Cancer* 69: 259–264.
  26. Rhee I, Bachman KE, Park BH, Jair KW, Yen RW, et al. (2002) DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. *Nature* 416: 552–556.
  27. Myohanen SK, Baylin SB, Herman JG (1998) Hypermethylation can selectively silence individual p16ink4A alleles in neoplasia. *Cancer Res* 58: 591–593.
  28. Zardo G, Tiirikainen MI, Hong C, Misra A, Feuerstein BG, et al. (2002) Integrated genomic and epigenomic analyses pinpoint biallelic gene inactivation in tumors. *Nat Genet* 32: 453–458.
  29. Santarosa M, Ashworth A (2004) Haploinsufficiency for tumour suppressor genes: when you don't need to go all the way. *Biochim Biophys Acta* 1654: 105–122.
  30. Alberici P, Jagmohan-Changur S, De Pater E, Van Der Valk M, Smits R, et al. (2006) Smad4 haploinsufficiency in mouse models for intestinal cancer. *Oncogene* 25: 1841–1851.
  31. Yang TL, Su YR, Huang CS, Yu JC, Lo YL, et al. (2004) High-resolution 19p13.2–13.3 allelotyping of breast carcinomas demonstrates frequent loss of heterozygosity. *Genes Chromosomes Cancer* 41: 250–256.
  32. Utada Y, Haga S, Kajiwara T, Kasumi F, Sakamoto G, et al. (2000) Mapping of target regions of allelic loss in primary breast cancers to 1-cM intervals on genomic contigs at 6q21 and 6q25.3. *Jpn J Cancer Res* 91: 293–300.
  33. Abdollahi A, Roberts D, Godwin AK, Schultz DC, Sonoda G, et al. (1997) Identification of a zinc-finger gene at 6q25: a chromosomal region implicated in development of many solid tumors. *Oncogene* 14: 1973–1979.
  34. Bagchi A, Papazoglu C, Wu Y, Capurso D, Brodt M, et al. (2007) CHD5 is a tumor suppressor at human 1p36. *Cell* 128: 459–475.
  35. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, et al. (2002) A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 347: 1999–2009.
  36. van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, et al. (2002) Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415: 530–536.
  37. Lu Y, Cai Z, Xiao G, Liu Y, Keller ET, et al. (2007) CCR2 expression correlates with prostate cancer progression. *J Cell Biochem* 101: 676–685.
  38. Fu Z, Kitagawa Y, Shen R, Shah R, Mehra R, et al. (2006) Metastasis suppressor gene Raf kinase inhibitor protein (RKIP) is a novel prognostic marker in prostate cancer. *Prostate* 66: 248–256.
  39. Tomlins SA, Mehra R, Rhodes DR, Smith LR, Roulston D, et al. (2006) TMPRSS2:ETV4 gene fusions define a third molecular subtype of prostate cancer. *Cancer Res* 66: 3396–3400.
  40. Weterman MA, Ajubi N, van Dinter IM, Degen WG, van Muijen GN, et al. (1995) nmb, a novel gene, is expressed in low-metastatic human melanoma cell lines and xenografts. *Int J Cancer* 60: 73–81.
  41. Soerjomataram I, Louwman MW, Ribot JG, Roukema JA, Coebergh JW (2008) An overview of prognostic factors for long-term survivors of breast cancer. *Breast Cancer Res Treat* 107: 309–330.
  42. Wood LD, Parsons DW, Jones S, Lin J, Sjoblom T, et al. (2007) The genomic landscapes of human breast and colorectal cancers. *Science* 318: 1108–1113.
  43. Forrest WF, Cavet G (2007) Comment on “The consensus coding sequences of human breast and colorectal cancers”. *Science* 317: 1500; author reply 1500.
  44. Rubin AF, Green P (2007) Comment on “The consensus coding sequences of human breast and colorectal cancers”. *Science* 317: 1500.
  45. Hu M, Yao J, Cai L, Bachman KE, van den Brule F, et al. (2005) Distinct epigenetic changes in the stromal cells of breast cancers. *Nat Genet* 37: 899–905.
  46. Lund P, Weissaupt K, Mikeska T, Jammam D, Chen X, et al. (2006) Oncogenic HRAS suppresses clusterin expression through promoter hypermethylation. *Oncogene* 25: 4890–4903.
  47. Grady WM, Parkin RK, Mitchell PS, Lee JH, Kim YH, et al. (2008) Epigenetic silencing of the intronic microRNA hsa-miR-342 and its host gene EVL in colorectal cancer. *Oncogene*. In press.
  48. Shames DS, Girard L, Gao B, Sato M, Lewis CM, et al. (2006) A genome-wide screen for promoter methylation in lung cancer identifies novel methylation markers for multiple malignancies. *PLoS Med* 3: e486. doi:10.1371/journal.pmed.0030486
  49. Jacinto FV, Ballestar E, Roperio S, Esteller M (2007) Discovery of epigenetically silenced genes by methylated DNA immunoprecipitation in colon cancer cells. *Cancer Res* 67: 11481–11486.
  50. Keshet I, Schlesinger Y, Farkash S, Rand E, Hecht M, et al. (2006) Evidence for an instructive mechanism of de novo methylation in cancer cells. *Nat Genet* 38: 149–153.
  51. Jeronimo C, Henrique R, Hoque MO, Mambo E, Ribeiro FR, et al. (2004) A quantitative promoter methylation profile of prostate cancer. *Clin Cancer Res* 10: 8472–8478.
  52. Jones PA, Laird PW (1999) Cancer epigenetics comes of age. *Nat Genet* 21: 163–167.
  53. Costello JF, Plass C (2001) Methylation matters. *J Med Genet* 38: 285–303.
  54. Herman JG, Latif F, Weng Y, Lerman MI, Zbar B, et al. (1994) Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. *Proc Natl Acad Sci U S A* 91: 9700–9704.
  55. Esteller M, Silva JM, Dominguez B, Bonilla F, Matias-Guiu X, et al. (2000) Promoter hypermethylation and BRCA1 inactivation in sporadic breast and ovarian tumors. *J Natl Cancer Inst* 92: 564–569.
  56. Esteller M, Avizienyte E, Corn PG, Lothe RA, Baylin SB, et al. (2000) Epigenetic inactivation of LKB1 in primary tumors associated with the Peutz-Jeghers syndrome. *Oncogene* 19: 164–168.
  57. Grady WM, Willis J, Guilford PJ, Dumbier AK, Toro TT, et al. (2000) Methylation of the CDH1 promoter as the second genetic hit in hereditary diffuse gastric cancer. *Nat Genet* 26: 16–17.
  58. Escot C, Theillet C, Lidereau R, Spyrtas F, Champeme MH, et al. (1986) Genetic alteration of the c-myc protooncogene (MYC) in human primary breast carcinomas. *Proc Natl Acad Sci U S A* 83: 4834–4838.
  59. Battey J, Moulding C, Taub R, Murphy W, Stewart T, et al. (1983) The human c-myc oncogene: structural consequences of translocation into the IgH locus in Burkitt lymphoma. *Cell* 34: 779–787.
  60. Esteller M, Garcia-Foncillas J, Andion E, Goodman SN, Hidalgo OF, et al. (2000) Inactivation of the DNA-repair gene MGMT and the clinical response of gliomas to alkylating agents. *N Engl J Med* 343: 1350–1354.
  61. Kato K, Iida S, Uetake H, Takagi Y, Yamashita T, et al. (2007) Methylated TMS1 and DAPK genes predict prognosis and response to chemotherapy in gastric cancer. *Int J Cancer* 122: 603–608.
  62. Nakata S, Sugio K, Uramoto H, Oyama T, Hanagiri T, et al. (2006) The methylation status and protein expression of CDH1, p16(INK4A), and fragile histidine triad in nonsmall cell lung carcinoma: epigenetic silencing, clinical features, and prognostic significance. *Cancer* 106: 2190–2199.
  63. Ramirez JL, Rosell R, Taron M, Sanchez-Ronco M, Alberola V, et al. (2005) 14-3-3sigma methylation in pretreatment serum circulating DNA of cisplatin-plus-gemcitabine-treated advanced non-small-cell lung cancer patients predicts survival: The Spanish Lung Cancer Group. *J Clin Oncol* 23: 9105–9112.
  64. Yu J, Yu J, Rhodes DR, Tomlins SA, Cao X, et al. (2007) A polycomb repression signature in metastatic prostate cancer predicts cancer outcome. *Cancer Res* 67: 10657–10663.

## Editors' Summary

**Background.** Cancer is one of the developed world's biggest killers—over half a million Americans die of cancer each year, for instance. As a result, there is great interest in understanding the genetic and environmental causes of cancer in order to improve cancer prevention, diagnosis, and treatment.

Cancer begins when cells begin to multiply out of control. DNA is the sequence of coded instructions—genes—for how to build and maintain the body. Certain “tumor suppressor” genes, for instance, help to prevent cancer by preventing tumors from developing, but changes that alter the DNA code sequence—mutations—can profoundly affect how a gene works. Modern techniques of genetic analysis have identified genes such as tumor suppressors that, when mutated, are linked to the development of certain cancers.

**Why Was This Study Done?** However, in recent years, it has become increasingly apparent that mutations are neither necessary nor sufficient to explain every case of cancer. This has led researchers to look at so-called epigenetic factors, which also alter how a gene works without altering its DNA sequence. An example of this is “methylation,” which prevents a gene from being expressed—deactivates it—by a chemical tag. Methylation of genes is part of the normal functioning of DNA, but abnormal methylation has been linked with cancer, aging, and some rare birth abnormalities.

Previous analysis of DNA from breast and colon cancer cells had revealed 189 “candidate cancer genes”—mutated genes that were linked to the development of breast and colon cancer. However, it was not clear how those mutations gave rise to cancer, and individual mutations were present in only 5% to 15% of specific tumors. The authors of this study wanted to know whether epigenetic factors such as methylation contributed to causing the cancers.

**What Did the Researchers Do and Find?** The researchers first identified 56 of the 189 candidate cancer genes as likely tumor suppressors and then determined that 36 of these genes were methylated and deactivated, often in both breast and colon (laboratory-grown) cancer cells. In nearly all cases, the methylated genes were not active but could be reactivated by being demethylated. They further showed that, in normal colon and breast tissue samples, 18 of the 36 genes were unmethylated and functioned normally, but in cells taken from breast and colon cancer tumors they were methylated.

In contrast to the genetic mutations, the 18 genes were frequently methylated across a range of tumor types, and eight genes were

methylated in both the breast and colon cancers. The authors found by reviewing the genetics and epigenetics of those 18 genes in breast and colon cancer that they were either mutated, methylated, or both. A literature review showed that at least six of the 18 genes were known to have tumor suppressor properties, and the authors determined that 16 were located in parts of DNA known to be missing from cells taken from a range of cancer tumors.

Finally, the researchers analyzed data on cancer cases to show that methylation of these 18 genes was correlated with reduced function of these genes in tumors and with a greater likelihood that a cancer will be terminal or spread to other parts of the body.

**What Do These Findings Mean?** The researchers considered only the 189 candidate cancer genes found in one previous study and not other genes identified elsewhere. They also did not consider the biological effects of the individual mutations found in those genes. Despite this, they have demonstrated that methylation of specific genes is likely to play a role in the development of breast and/or colon cancer cells either together with mutations or independently, most likely by turning off their tumor suppression function.

More broadly, however, the study adds to the evidence that future analysis of the role of genes in cancer should include epigenetic as well as genetic factors. In addition, the authors have also shown that a number of these genes may be useful for predicting clinical outcomes for a range of tumor types.

**Additional Information.** Please access these Web sites via the online version of this summary at <http://dx.doi.org/10.1371/journal.pmed.0050114>.

- A December 2006 *PLoS Medicine* Perspective article reviews the value of examining methylation as a factor in common cancers and its use for early detection
- The Web site of the American Cancer Society has a wealth of information and resources on a variety of cancers, including breast and colon cancer
- Breastcancer.org is a nonprofit organization providing information about breast cancer on the Web, including research news
- Cancer Research UK provides information on cancer research
- The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins publishes background information on the authors' research on methylation, setting out its potential for earlier diagnosis and better treatment of cancer