OPEN ACCESS Freely available online PLoS GENETICS

Comparing the DNA Hypermethylome with Gene Mutations in Human Colorectal Cancer

Kornel E. Schuebel[1☯*], Wei Chen[1,2☯], Leslie Cope[3☯], Sabine C. Glöckner[4☯], Hiromu Suzuki[5], Joo-Mi Yi[1], Timothy A. Chan[1], Leander Van Neste[6], Wim Van Criekinge[7], Sandra van den Bosch[8], Manon van Engeland[8], Angela H. Ting[1], Kamwing Jair[9], Wayne Yu[1], Minoru Toyota[5], Kohzoh Imai[5], Nita Ahuja[4], James G. Herman[1], Stephen B. Baylin[1,2*]

1 Cancer Biology Division, The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Baltimore, Maryland, United States of America, **2** Predoctoral Training Program in Human Genetics, The Johns Hopkins University, Baltimore, Maryland, United States of America, **3** Biometry and Clinical Trials Division, The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Baltimore, Maryland, United States of America, **4** Department of Surgery, The Johns Hopkins University School of Medicine, Baltimore, Maryland, United States of America, **5** First Department of Internal Medicine, Sapporo Medical University, Sapporo, Japan, **6** Department of Molecular Biotechnology, Faculty of Bioscience Engineering, Ghent University, Ghent, Belgium, **7** Oncomethylome Sciences, Liege, Belgium, **8** Department of Pathology, University of Maastricht, Maastricht, The Netherlands, **9** Bionumerik Pharmaceuticals Inc., San Antonio, Texas, United States of America

We have developed a transcriptome-wide approach to identify genes affected by promoter CpG island DNA hypermethylation and transcriptional silencing in colorectal cancer. By screening cell lines and validating tumor-specific hypermethylation in a panel of primary human colorectal cancer samples, we estimate that nearly 5% or more of all known genes may be promoter methylated in an individual tumor. When directly compared to gene mutations, we find larger numbers of genes hypermethylated in individual tumors, and a higher frequency of hypermethylation within individual genes harboring either genetic or epigenetic changes. Thus, to enumerate the full spectrum of alterations in the human cancer genome, and to facilitate the most efficacious grouping of tumors to identify cancer biomarkers and tailor therapeutic approaches, both genetic and epigenetic screens should be undertaken.

Citation: Schuebel KE, Chen W, Cope L, Glöckner SC, Suzuki H, et al. (2007) Comparing the DNA hypermethylome with gene mutations in human colorectal cancer. PLoS Genet 3(9): e157. doi:10.1371/journal.pgen.0030157

Introduction

It is now well established that loss of proper gene function in human cancer can occur through both genetic and epigenetic mechanisms [1,2]. The number of genes mutated in human tumor samples is being clarified. Recently, Sjöblom et al. [3] sequenced 13,023 genes in colorectal cancer (CRC) and breast cancer, and estimated an average of 14 significant mutations per tumor, suggesting that a relatively small number of genetic events may be sufficient to drive tumorigenesis. In contrast, the full spectrum of epigenetic alterations is not well delineated. The best-defined epigenetic alteration of cancer genes involves DNA hypermethylation of clustered CpG dinucleotides, or CpG islands, in promoter regions associated with the transcriptional inactivation of the affected genes [2]. These promoters are located proximal to nearly half of all genes [4] and are thought to remain primarily methylation free in normal somatic tissues. The exact number of such epigenetic lesions in any given tumor is not precisely known, although a growing number of screening approaches, none covering the whole genome efficiently, are identifying an increasing number of candidate genes [5–13]. Given the large number of potential target promoters present in the genome, we hypothesized that many more hypermethylated genes await discovery.

Herein, we describe a whole human transcriptome microarray screen to identify genes silenced by promoter hypermethylation in human CRC. The approach readily identifies candidate cancer genes in single tumors with a high efficiency of validation. By comparing the list of candidate hypermethylated genes with mutated genes recently identified in CRC [3], we establish key relationships between the altered tumor genome and the gene hypermethylome. Our studies provide a platform to understand how epigenetic and genetic alterations drive human tumorigenesis.

Results

Developing the Whole Transcriptome Approach

Our first step towards a global identification of hypermethylation-dependent gene expression changes was made by comparing, in a genome-wide expression array-based approach, wild-type HCT116 CRC cells with isogenic partner cells carrying individual and combinatorial genetic deletions of two major human DNA methyltransferases (Figure 1A) [14].

Editor: Jeannie T. Lee, Massachusetts General Hospital, United States of America

Received April 12, 2007; **Accepted** July 31, 2007; **Published** September 21, 2007

A previous version of this article appeared as an Early Online Release on July 31, 2007 (doi:10.1371/journal.pgen.0030157.eor).

Abbreviations: CRC, colorectal cancer; DAC, 5-aza-2′-deoxycytidine; DKO, double knockout; MSP, methylation-specific PCR; RT-PCR, reverse transcriptase PCR; TSA, trichostatin A

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

☯ These authors contributed equally to this work.

Author Summary

Loss of gene expression in association with aberrant accumulation of 5-methylcytosine in gene promoter CpG islands is a common feature of human cancer. Here, we describe a method to discover these genes that permits identification of hundreds of novel candidate cancer genes in any cancer cell line. We now estimate that as much as 5% of colon cancer genes may harbor aberrant gene hypermethylation and we term these the cancer "promoter CpG island DNA hypermethylome." Multiple mutated genes recently identified via cancer resequencing efforts are shown to be within this hypermethylome and to be more likely to undergo epigenetic inactivation than genetic alteration. Our approach allows derivation of new potential tumor biomarkers and potential pathways for therapeutic intervention. Importantly, our findings illustrate that efforts aimed at complete identification of the human cancer genome should include analyses of epigenetic, as well as genetic, changes.

Importantly, in the *DNMT1*$^{(-/-)}$*DNMT3B*$^{(-/-)}$ double knockout (DKO) HCT116 cells, which have virtually complete loss of global 5-methylcytosine, all previously individually examined hypermethylated genes lacking basal expression in wild-type cells undergo promoter demethylation with concomitant gene re-expression [10,14–16]. By stratifying genes according to altered signal intensity on a 44K Agilent Technologies array platform, we observe a unique spike of gene expression increases in the DKO cells when compared to the isogenic wildtype parental cells, or isogenic cell lines in which *DNMT1* or *DNMT3B* have been individually deleted and which harbor minimal changes in DNA methylation (Figure 1B). This minimal change in the *DNMT1*$^{(-/-)}$cells may, in part, be due to recently identified alternative transcripts arising from the *DNMT1* locus [17,18].

We tested our approach using a pharmacologic strategy based on our previous approach [10], but now markedly modified to provide whole-transcriptome coverage, to identify silenced hypermethylated genes in any cancer cell line. For densely hypermethylated and transcriptionally inactive genes, the DNA demethylating agent 5-aza-2′-deoxycytidine (DAC) has a well established capacity to induce gene re-expression [19,20]. On the other hand, for these same genes, the class I and II histone deacetylase inhibitor, trichostatin A (TSA) will not alone induce reexpression [10,21]. We now use this lack of TSA response for such genes to provide a new informatics filter to identify the majority of DNA hypermethylated genes in cancer. After treatment of HCT116 cells with either DAC or TSA (Figure 1C), we identified a zone in which gene expression did not increase with TSA (<1.4-fold) and displayed no detectable expression in mock-treated cells. Within this zone, we observed a characteristic spike of DAC-induced gene expression that virtually completely encompasses the genes with increased expression in DKO cells (compare yellow spots in Figure 1D with blue spots in Figure 1B). This gene spike is absolutely dependent upon analysis of only genes that fail to respond to histone deacetylase inhibition, underscored by a cluster analysis that shows the close relationship between genes in DKO- and DAC-treated cells with a separate grouping of gene-expression changes after TSA treatment alone or in single knockouts (Figure 1E). These data confirm previous studies covering much less of the genome, and using only treatment of cells with DAC and TSA together, in which genes with dense CpG islands that were reexpressed by TSA harbored only partial or no detectable hypermethylation [10,21].

Importantly, a similar spike of gene expression increases could be seen in five additional human CRC cell lines, SW480, CaCO2, RKO, HT29, and COLO320 (Figure 2A), as well as cell lines derived from lung, breast, ovary, kidney, and brain (unpublished data), confirming that this approach works universally in cancer cell lines and identifies overlapping gene sets (Figure 2C). However, it is important to note that—possibly because DAC incorporates into the DNA of dividing cells, and our treatments were performed for only 96 h—sensitivity for detecting the gene increases in the pharmacological approach is reduced in HCT116 cells compared to that seen in DKO cells (Figure 1D). To address the sensitivity with which our new array approach identifies CpG island hypermethylated genes, we first examined 11 genes known to be hypermethylated, completely silenced and reexpressed after DAC treatment in HCT116 cells (Figure 3A). All tested genes remained within the TSA nonresponsive zone (Figure 3B), and the direction of expression changes correlated well in DAC treated and DKO cells (Figure 3C). Importantly, for the DAC increase, five of the guide genes (45%) increased 2-fold or more and three more genes, or a total of 73%, increased 1.3-fold or more (Figure 3D). We estimate, then, that we can detect over 70% of DNA hypermethylated genes in a given cancer cell line and we test this hypothesis in studies directly below.

Figure 1. Approach for Identification of the Human Cancer Cell Hypermethylome in HCT116 CRC Cells

(A) RNA from the indicated cell lines was isolated, labeled, hybridized, scanned, and fluorescent spot intensities normalized by background subtraction and Loess transformation using Agilent Technologies 44K human microarrays. Parental wild-type HCT116 cells (WT) and isogenic knockout counterparts for DNA methyltransferase 1 (*DNMT1*$^{-/-}$) or 3b (*DNMT3B*$^{-/-}$) are compared in our study. DKO cells are doubly deficient for both *DNMT1* and *DNMT3B*.

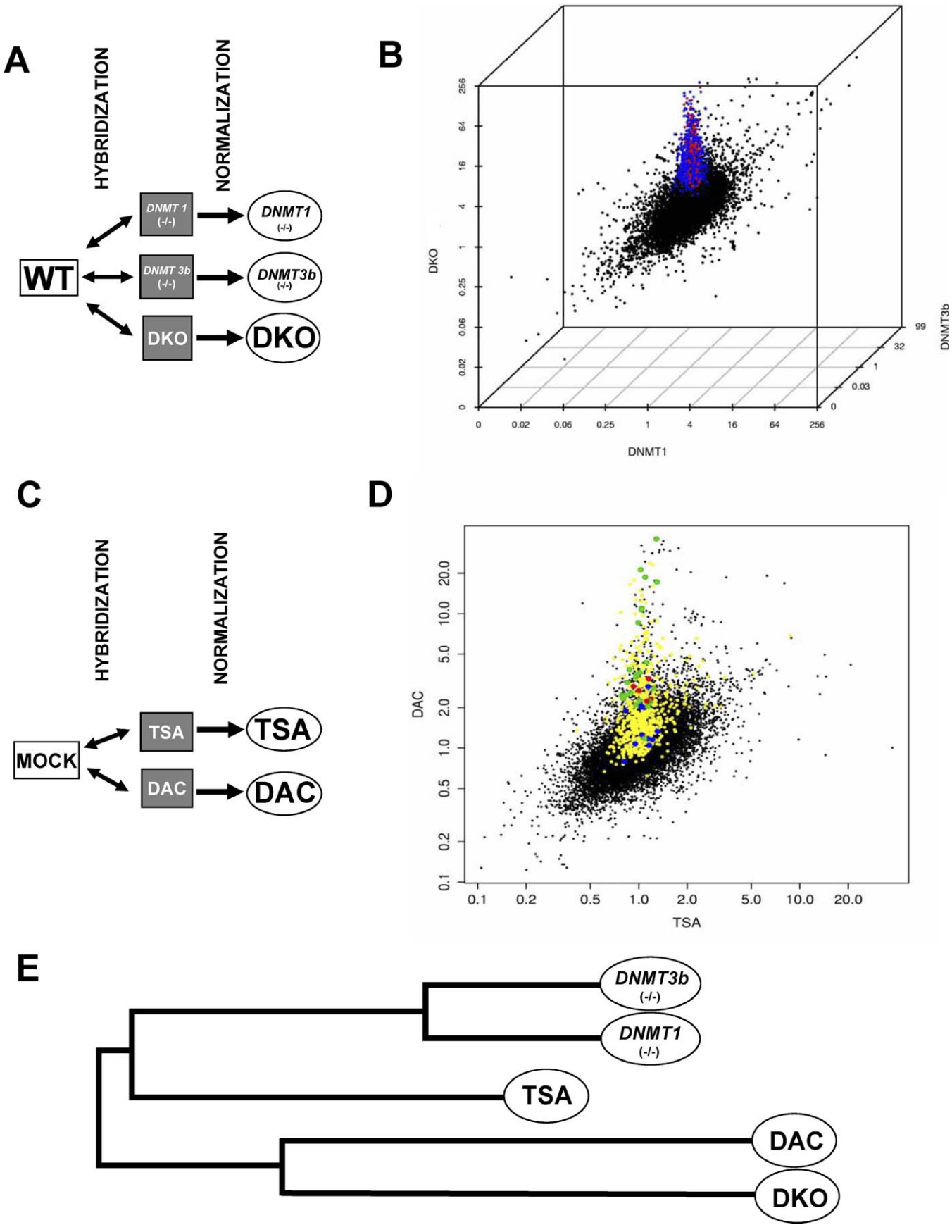
(B) Gene-expression changes in HCT116 cells with genetic disruption of various DNA methyltransferases. A 3-D scatter plot indicating the gene-expression levels in HCT 116 cells with genetic disruption of *DNMT1* (*x*-axis), *DNMT3B* (*z*-axis), and both *DNMT1* and *DNMT3B* (DKO; *y*-axis) in fold scale. Individual gene-expression changes are in black with the average for three experiments (red spots) or from an individual experiment (blue spots) for those genes in DKO cells with greater than 4-fold expression change.

(C) HCT116 cells were treated with 300 nM TSA for 18 h or 5 μM DAC for 96 h and processed as described above.

(D) Gene-expression changes for HCT116 cells treated with TSA (*x*-axis) or DAC (*y*-axis) are plotted by fold change. Yellow spots indicate genes from DKO cells with 2-fold changes and above. Notice the loss of sensitivity when compared to gene-expression increases seen in DKO cells (80% of genes greater than 4-fold in the DKO cells now becomes greater than 1.3-fold in DAC-treated cells). Green spots indicate randomly selected genes verified to have complete promoter methylation in wild-type cells, reexpression in DKO cells and after DAC treatment, while red spots indicate selected genes that were identified as false positives (See Figures 4, 6, and 7 for validation results). Blue spots indicate the location of the 11 guide genes—previously shown to be hypermethylated and completely silenced in HCT 116 cells—used in this study (see Figure 3 for description). A distinct group of genes, including five of 11 guide genes, displays increases of greater than 2-fold after DAC treatment but no increase after TSA treatment. These genes form the top tier of candidate hypermethylated genes as discussed in the text.

(E) Relatedness of whole-transcriptome expression patterns identified by dendrogram analysis. Individual single genetic disruption of *DNMT1* and *DNMT3B,* DKO and DAC treatment, and TSA treatment each form three distinct categories of gene expression changes.

doi:10.1371/journal.pgen.0030157.g001



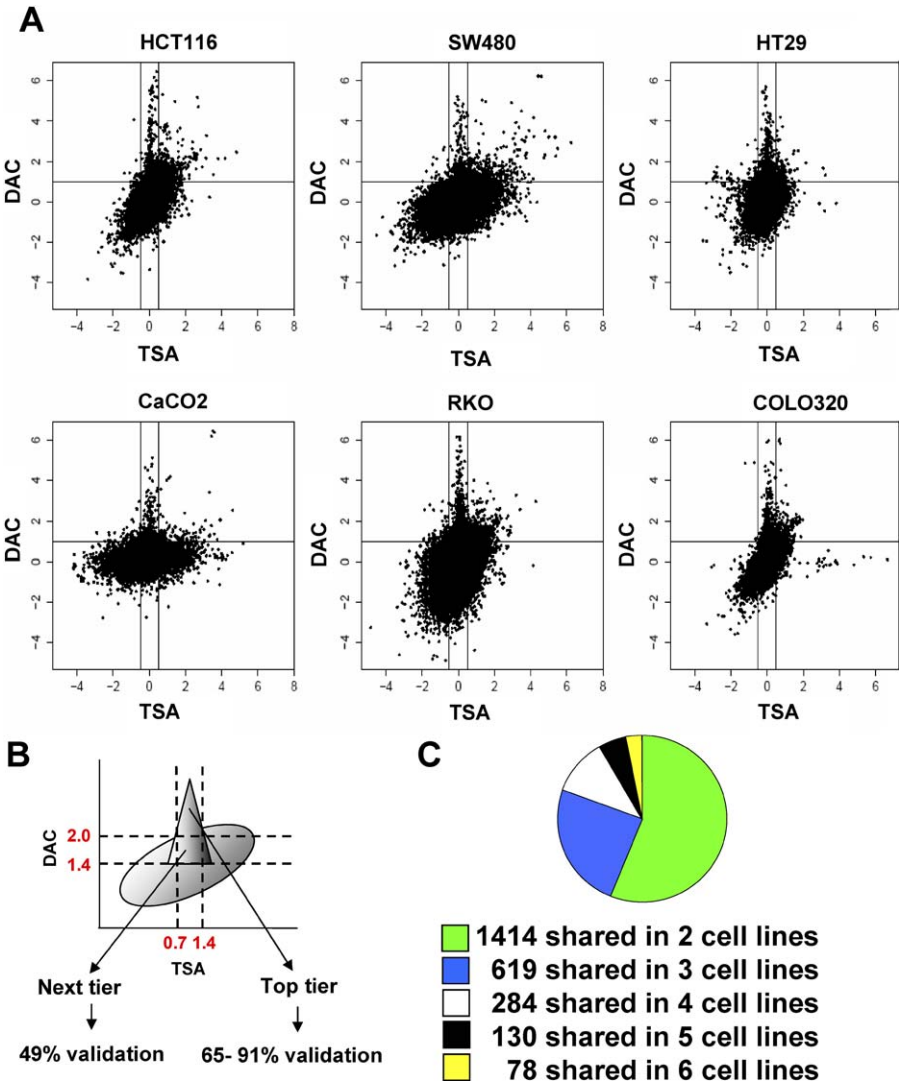


Figure 2. Characterization of the Human Cancer Cell Hypermethylome in Different Human CRC Cell Lines

(A) Gene-expression changes for the indicated cells treated with TSA (*x*-axis) or DAC (*y*-axis) are plotted by log-fold change, and individual genes are shown in black.

(B) Validation of the DNA hypermethylome. The characteristic spike of hypermethylated genes defined by treatment of cells with DAC or TSA consists of two tiers, with distinct features. The top tier of genes was identified as a zone in which gene expression did not increase with TSA (<1.4 fold) and displayed no detectable expression in wild-type cells, but increased greater than 2-fold with DAC treatment. The next tier of genes was identified as a cluster of genes for which expression changes of TSA and wild type were identical to those in the top tier, but increased between 1.4-fold and 2-fold with DAC treatment. Gene expression validation by RT-PCR and MSP indicated a validation frequency of 91% for top-tier genes in HCT116 cells, including genes that increased in DKO cells by greater than 2-fold. Next-tier genes in HCT116 cells were confirmed at a frequency of 49%, and in the SW480 top tier, with a frequency of 65%.

(C) Shared candidate hypermethylated genes in CRC cell lines. We identified a total of 5,906 unique genes in all six cell lines with expression changes falling within the criteria of top- or next-tier categories. Overlaps in gene expression changes among two, three, four, five, or six cell lines are indicated; these range from 1,414 genes shared among two cell lines to 78 genes that were shared among all six cell lines.

doi:10.1371/journal.pgen.0030157.g002

Validating the Methylation Status of Candidate Genes Derived from the Screening Approach

Based on the sensitivity differences observed between DKO- and DAC -induced gene increases (compare Figure 1B and D; also Figure 3B and 3C) and behavior of the guide genes in the array platform, we designated, within the TSA-negative zone, a top tier (2-fold increase or above) and a next tier of genes (increasing between 1.4- and 2-fold) to identify hypermethylated cancer genes (Figure 2B). Importantly, we introduced an additional filter for selecting genes from these zones based on their having no basal expression in untreated cells, since this full lack of transcription is characteristic of promoter CpG island methylated genes in cell culture. Indeed, based on these selection criteria, in HCT116 cells, 32 of 35 (91%, Figure 4) of randomly chosen CpG island–containing genes spanning the top-tier response zone of 532 genes (Figure 5), and 31 of 48 such SW480 cell genes (65%, Figure 6) from among 318 top tier genes proved to be CpG hypermethylated as measured by methylation-specific PCR (MSP) [22], and silenced in the cell line of origin as measured by reverse transcriptase PCR (RT-PCR). We also examined the efficiency of discovery for hypermethylated genes in the next

A

Gene Name	Agilent ID	Accession #	Reference
CHFR	A_23_P344037	NM_018223	16
GATA4	A_23_P384761	NM_002052	15
GATA5	A_23_P371835	NM_080473	15
INHα	A_23_P51039	NM_002191	15
SFRP1	A_23_P10127	NM_003012	48
SFRP4	A_23_P215328	NM_003014	48
SFRP5	A_23_P1352	NM_003015	48
TFF1	A_24_P322771	NM_003225	15
TFF2	A_23_P57364	NM_005423	15
TFF3	A_23_P393099	NM_003226	15
Timp3	A_23_P399078	NM_000362	21

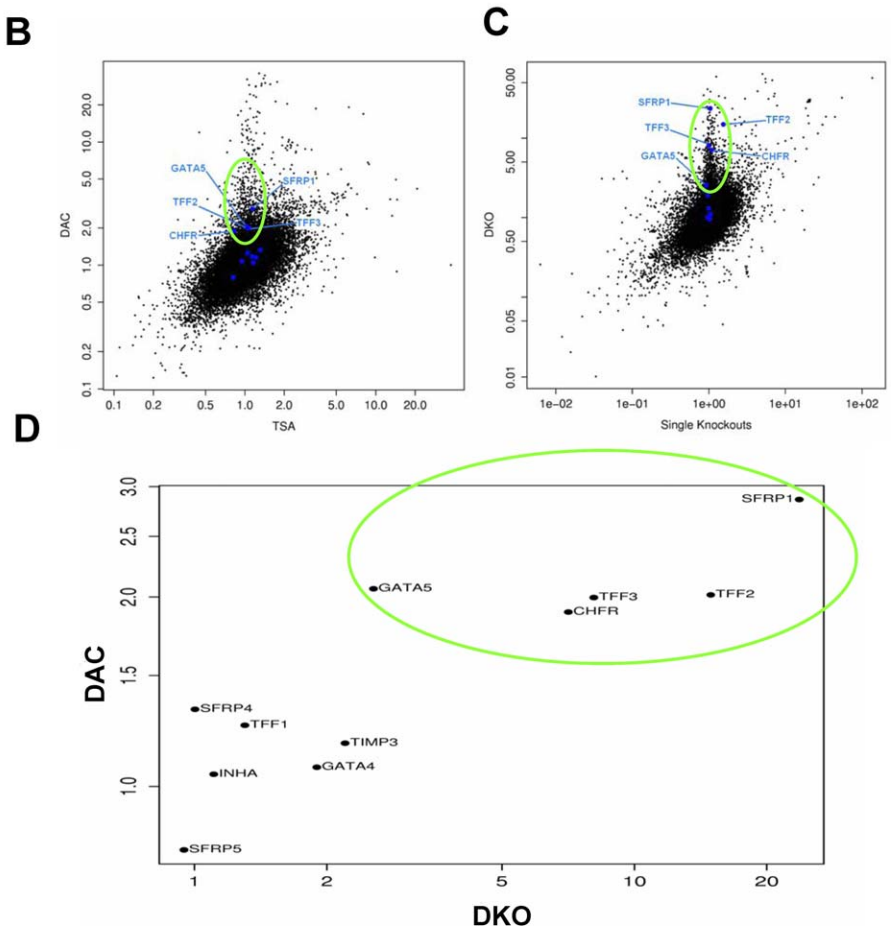


Figure 3. Guide Genes Used in This Study

(A) Gene names, Agilent Technologies probe name, Genbank accession number, and references for the 11 guide genes previously shown to be hypermethylated and completely silenced in HCT116 cells.

(B, C) Blue spots and gene names indicate the location of the 11 guide genes in a plot of TSA (*x*-axis) versus DAC (*y*-axis) gene expression changes on a log scale (B) or fold-change (C) scale. Five of 11 guide genes, circled in green, display increases of greater than 2-fold after DAC treatment but no increase after TSA treatment and these same genes have greater than 3-fold increases in DKO cells (green circle)

(D) Direct comparison of guide genes in DKO and DAC plots. A distinct group of five guide genes, indicated by a green circle, showing greater than 3-fold expression changes in DKO cells and greater than 2-fold in DAC-treated cells, define the upper tier of candidate hypermethylated genes as discussed in the text. Another three genes increased 1.3-fold, and three failed to increase with DAC treatment, allowing criteria for the next tier of gene expression to be established as described in the text.

doi:10.1371/journal.pgen.0030157.g003

tier of DAC-treated HCT116 cells. Of the 1,190 genes identified in this region, 17 of 35 (49%) randomly selected genes containing a CpG island were hypermethylated with concordant gene silencing (Figure 7). Our verification rates then demonstrate around 65% efficiency of our approach, which is close to our original estimate and which is excellent compared to previous screens for identifying new cancer hypermethylated genes [6,23]. With this level of verified hypermethylation, we calculate that the hypermethylome in HCT116 cells consists of an estimated 1,067 genes and an estimated 579 genes for the SW480 cells (See Table S1 for a detailed description of calculations). The hypermethylome would be estimated to range from 532 genes in CaCO2 to 1,389 genes in RKO cells (Table S1).

We next asked whether our top and next-tier regions truly enriched for hypermethylated genes by examining a randomly

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.