The American Journal of Pathology, Vol. 183, No. 6, December 2013

The American Journal of
PATHOLOGY
ajp.amjpathol.org

TUMORIGENESIS AND NEOPLASTIC PROGRESSION

# Whole-Genome Methylation Sequencing Reveals Distinct Impact of Differential Methylations on Gene Transcription in Prostate Cancer

Yan P. Yu,* Ying Ding,† Rui Chen,† Serena G. Liao,† Bao-Guo Ren,* Amantha Michalopoulos,* George Michalopoulos,* Joel Nelson,‡ George C. Tseng,† and Jian-Hua Luo*

*From the Departments of Pathology,* Biostatistics,† and Urology,‡ University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania*

Accepted for publication August 7, 2013.

Address correspondence to Jian-Hua Luo, M.D., Ph.D., Department of Pathology, 3550 Terrace St., Pittsburgh, PA 15261. E-mail: luoj@msx.upmc.edu.

DNA methylation is one of the most important epigenetic mechanisms in regulating gene expression. Genome hypermethylation has been proposed as a critical mechanism in human malignancies. However, whole-genome quantification of DNA methylation of human malignancies has rarely been investigated, and the significance of the genome distribution of CpG methylation is unclear. We performed whole-genome methylation sequencing to investigate the methylation profiles of 13 prostate samples: 5 prostate cancers, 4 matched benign prostate tissues adjacent to tumor, and 4 age-matched organ-donor prostate tissues. Alterations of methylation patterns occurred in prostate cancer and in benign prostate tissues adjacent to tumor, in comparison with age-matched organ-donor prostates. More than 95% alterations of genome methylation occurred in sequences outside CpG islands. Only a small fraction of the methylated CpG islands had any effect on RNA expression. Both intragene and promoter CpG island methylations negatively affected gene expression. However, suppressions of RNA expression did not correlate with levels of CpG island methylation, suggesting that CpG island methylation alone might not be sufficient to shut down gene expression. Motif analysis revealed a consensus sequence containing Sp1 binding motif significantly enriched in the effective CpG islands. *(Am J Pathol 2013, 183: 1960–1970; http://dx.doi.org/10.1016/j.ajpath.2013.08.018)*

Prostate cancer is one of the most prevalent malignancies among American men. Approximately 241,740 new cases and up to 28,170 prostate cancer deaths were estimated for the United States in 2012.[1] The mortality rate from prostate cancer is second only to lung carcinoma in the United States.[1] Although most prostate cancers are indolent and responsive to the available surgery and radiation interventions, a significant number of cases become hormone refractory and metastatic. The precise mechanism of prostate cancer progression remains elusive, despite extensive research efforts and recent advances in understanding of this disease. Comprehensive gene expression and genome analyses have suggested that a global pattern of gene expression and copy number alterations exist for prostate cancer.[2–7] The related gene products include critical molecules in signaling pathways, DNA replication, cell growth, cell-cycle checkpoints, and apoptosis.[2,8–10]

Hypermethylation of a gene promoter region is a critical epigenetic event that silences gene expression and plays important roles in normal physiology. Genome allele methylation mediates gene imprinting for inactivation of the X chromosome[11] and generates tissue-specific gene expression.[12] In pathological processes, DNA methylation inactivates tumor suppressor genes and promotes tumorigenesis.[13,14] Genome hypermethylation has been proposed as a critical mechanism in human malignancies.[13–18] Silencing of genes involved in cell-cycle control, cell survival, DNA damage repair, and signal transduction is a characteristic of cancer cells.[19–27] However, there is a lack of global quantification of CpG dinucleotide methylation and gene expression in prostate cancer. To map whole-genome CpG dinucleotide methylation leading to altered expression of hundreds of genes in prostate cancer, we

Supported by NIH grant R01-CA098249 (J.H.L.), American Cancer Society grant RSG-08-137-01-CNE (Y.P.Y.), and a grant from the University of Pittsburgh Cancer Institute.

Y.P.Y. and Y.D. contributed equally to this work.

Y.P.Y., G.C.T., and J.-H.L. contributed equally to this work as senior authors.

Copyright © 2013 American Society for Investigative Pathology.
Published by Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ajpath.2013.08.018

performed whole-genome methylation sequencing and characterized the methylation profiles in a set of prostate tumors, matched benign prostate tissues adjacent to tumor, and age-matched organ-donor prostate tissues. Unique methylation profiles distinguished the three types of samples.

## Materials and Methods

### Sample Preparation and Genomic DNA Extraction

Frozen specimens of prostate tumor (T) ($n = 5$), matched benign prostate tissues adjacent to tumor (AT) ($n = 4$), and organ-donor prostate tissues (OD) ($n = 4$) were obtained from the Tissue Bank of the University of Pittsburgh as approved by the Institutional Review Board. Clinicodemographic characteristics for the tumor specimens are listed in Table 1. Each sample was microdissected to achieve >80% purity of cancer cells or prostate epithelial cells. AT samples (>3 mm away from the cancer) were obtained using needle microdissection by board-certified pathologists (J.H.L. and G.M.). Organ donor samples were from healthy individuals free of any urological disease. Genomic DNA of each set of OD, T, and AT tissues was extracted using a commercially available tissue and blood DNA extraction kit (Qiagen, Valencia, CA).

### Sodium Bisulfite Conversion of Genomic DNA and Generation of Sequencing Library

Genomic DNA was bisulfite-converted with an EpiTect bisulfite conversion kit (Qiagen, Valencia, CA). For each sample, 1 μg of genomic DNA was incubated with sodium bisulfite solution at 60°C for 4.5 hours. This was followed by desulfonation, purified through the column according to the manufacturer's instructions. The complementary strand was synthesized by incubating bisulfite-treated DNA with random hexamer and dNTP using T4 DNA polymerase and T4 DNA ligase for 18 hours at 16°C. Double-stranded DNA was then purified with Agencourt AMPure beads (Beckman Coulter, Brea, CA). To prepare the genomic DNA libraries, 50 ng DNA was then subjected to the tagmentation reactions (ie, transposase-based DNA fragmentation and adaptor ligation) using a Nextera DNA sample preparation kit (Epicentre—Illumina, Madison, WI) for 5 minutes at 55°C. The DNA was then amplified with adaptor and sequencing primers for nine cycles of 95°C for 10 seconds, 62°C for 30 seconds, and 72°C for 3 minutes. The PCR products were purified with AMPure beads. The bisulfite-converted genomic DNA libraries were then analyzed with qPCR using Illumina (San Diego, CA) sequencing primers and visualized with an Agilent 2000 bioanalyzer (Agilent Technologies, Santa Clara, CA) or agarose gel electrophoresis.

**Table 1** Prostate Cancer Tumor Specimen Characteristics and Clinical Data

| Case | TNM | Gleason score | Age (decade) | PSA, pre-operative |
|---|---|---|---|---|
| 49T | T3bN1MX | 7 | 50s | 14.6 |
| 158T | T3aN0MX | 7 | 60s | 4.1 |
| 159T | T2cN0MX | 8 | 60s | 2.38 |
| 165T | T3bN1MX | 10 | 50s | 29.3 |
| 171T | T3bN1MX | 10 | 50s | 9.17 |

All five specimens were relapse tumors from European-origin (white) patients. Organ donor prostate specimens were from four healthy age-matched individuals (OD23 and OD25 in their 50s; OD20 and OD22 in their 60s) free of any urological disease.

### Whole-Genome Sequencing

The Illumina whole-genome sequencing system was applied to the analysis. The operation procedures strictly followed the manufacturer's instructions. In brief, the bisulfite-converted DNA libraries were hybridized to flow cells and subjected to primer extension and bridge amplification in an automated cBot process for 4 hours to generate a cluster of DNA sequencing templates. These clustered flow cells were then subjected to the sequencing analysis in an Illumina HiSeq 2000 system. Each library was sequenced with an average 27-fold genome coverage and with paired-end runs for 200-cycle analysis.

### Read Alignment and β-Value Calculation for Bisulfite Sequencing

We used Bismark software[28] version 0.6.3 to align paired-end bisulfite-treated DNA reads against the UCSC human reference genome build hg19 (*http://genome.ucsc.edu*, last accessed April 14, 2013). Bismark generated the BS-seq reference from build hg19 containing C-to-T conversion for the forward strand and G-to-A conversion for the reverse strand. Sequencing batches with more than 7% non-CpG unconverted cytosine were excluded from further analysis. Uniquely best-mapped reads were pooled with at most two mismatches and fragment size of mate pairs ranging from 400 to 1400 bp. The methylation extractor tool in Bismark called all cytosine methylation status of each CG. After methylation calling, we characterized the methylation strength of each CpG island and of 200-bp moving windows of the human whole genome in terms of the β value, calculated as β = Total number of methylated calls in region/(Total number of methylated calls in region + Total number of unmethylated calls in region).

### Read Alignment and Estimation of Expression Value for RNA-Seq

RNA-seq reads were at an average 1000× coverage. Whole-genome RNA-seq reads were aligned with TopHat software[29,30] version 1.4.1 against the UCSC reference genome build hg19, allowing two mismatches per any 100-bp read. TopHat used uniquely mapped reads (against the reference genome) and the genomic database (UCSC hg19 Gene Annotation) to build up a potential splicing sites database. Borrowing information from potential splicing sites, isoforms were determined. We then used Cufflinks software[29,31]

version 1.3.0 to calculate the fragments per kilobase of exon per million fragments mapped (FPKM) to measure the abundance of the transcripts for pair-end reads and obtained a gene-based FPKM value that summarized all isoforms in a given gene.

### Window-Based Differential Methylation Analysis

Moving windows (200 bp) on the whole human genome (build hg19) were used to study methylation differences in different regions among the different groups (T, AT, and OD). Fisher's exact test was used to assess the significance of differential methylation. A moving window was determined to be differentially methylated if the difference between the average β value of two groups was >0.4 and the adjusted *P* value from Fisher's exact test (false discovery rate controlled by Benjamini–Hochberg procedure) was <0.01. After differentially methylated moving windows were identified, we used MethylKit software[32] version 0.5.6 to characterize the genomic features of these moving windows. We defined the CpG shores as the 1000 bp upstream and downstream of a CpG island and the promoter as the 3000 bp upstream and 500 bp downstream of transcription start site.

### CpG Island Differential Methylation Analysis

A redefined CpG island annotation database was applied to define our CpG island region.[33,34] We classified the CpG islands into three types: non–gene-associated, intragene, and promoter CpG islands. The CpG gene association was defined as CpG islands located from 3000 bp upstream of an mRNA start site to the end of the transcript. A promoter CpG island was defined as a CpG island located between 3000 bp upstream and 500 bp downstream of the mRNA start site, and the intragene CpG island was defined as a CpG island located in the gene-containing genome sequence from 500 bp downstream of the transcription start site to the end of the last exon. The intergene CpG island was defined as a CpG island located in a nongene region at least 3000 bp away from an mRNA start site. We classified 66,765 CpG islands into 32,841 intragene, 9454 promoter, and 24,470 intergene CpG islands. Differentially methylated CpG islands were obtained for three pairwise comparisons: T versus OD, AT versus OD, and T versus AT. The significance of differential methylation was assessed by a two-sample *t*-test based on the β value for each sample of each specific CpG island. For each comparison group, the differentially methylated CpG islands were separated into two groups, hypermethylated and hypomethylated, using the AT or OD group as the baseline.

### Determination of Functional Effect of CpG Methylation on RNA Expression

We defined a CpG island as functional if hyper- or hypomethylation of a CpG island in one group relative to the other was associated with significantly lower or higher average RNA expression. To determine whether the differentially methylated CpG islands were functional, we combined the bisulfite-seq data with the RNA-seq data after matching the genes to the gene-associated CpG islands. We first performed quantile normalization on FPKM values from RNA-seq data across all samples. The differential expression of a given gene was determined with a preset *P*-value threshold from the *t*-test on normalized FPKM values. A differentially hypermethylated CpG island was determined to be functional if its overlapping gene was differentially down-regulated based on the RNA-seq data; conversely, a functional differentially hypomethylated CpG island was required to show up-regulation of associated gene expression. The differential expression analysis was performed through paired *t*-test on normalized FKPM values obtained from TopHat and Cufflinks. A *P*-value threshold of 0.05 was applied to both differential methylation and differential expression calling. This was combined with a requirement of opposite direction of effect size of differential methylation and differential expression. Six sets of functional differentially methylated genes were identified (by collecting all genes that had at least one functional differentially methylated CpG island): hyper- and hypomethylated genes in T versus OD, AT versus OD, and T versus AT comparisons.

### TCGA Data Preparation

Data were downloaded from the Cancer Genome Atlas (TCGA) (*http://tcga-data.nci.nih.gov/tcga*) in October 2012. Level 3 RNA-seq data were extracted from the Illumina HiSeq 2000–Agilent array platform and methylation data from an Infinium HumanMethylation450 beadchip kit (Illumina). We selected four TCGA cancer data sets that contained methylation and expression data from matched tumor and normal tissue adjacent to tumor samples: breast invasive carcinoma (BRCA), prostate adenocarcinoma (PRAD), head and neck squamous cell carcinoma (HNSC), and thyroid carcinoma (THCA). The numbers of samples available for differential methylation and gene expression integrative analyses are given in Supplemental Table S1. After gene matching between methylation data and gene expression data, paired *t*-tests for each gene were performed to calculate *P* values of differential expression and differential methylation. Because of variable sample sizes in each of the TCGA studies, the *P*-value threshold for defining functional differential methylated genes varied. We varied the *P*-value threshold to obtain a functional gene set of approximately 500 (760 in PRAD, 395 in BRCA, 422 in HNSC, and 510 in THCA).

### Motif Analysis

Paired *t*-tests for our data set and the four TCGA data sets were performed for each gene to calculate the *P* value for differential expression and differential methylation. Each gene was then ranked by the *P* values obtained from the rank-sum tests for

differential expression and differential methylation. CpG islands with increase in methylation β values ($P < 0.05$) resulting in twofold decreased expression ($P < 0.05$) of the overlapping gene were deemed concordant and effective. The noneffective groups were defined as differential methylation having weak or no effect on expression ($P > 0.5$). Motif enrichment analysis, comparing effective CpG islands versus noneffective CpG islands, was performed with MEME software[35] version 4.9.0 using the following criteria: size, 6 to 20 bp; frequency, 0 or 1 per CpG island; and present in >20 effective CpG islands. Effective groups were treated as a positive data set and noneffective groups as a negative data set. After obtaining multiple motifs from all data sets, these motifs were matched with known motifs from the TCGA database with the TOMTOM motif comparison tool.[36] Pearson correlation was set as the similarity measure with a significance threshold of $q = 0.01$ in TOMTOM. We then compared results from our data set and the four TCGA data sets to identify recurrent motifs enriched in effective CpG islands.

### Gene Set Enrichment Analysis

To investigate whether the functional CpG (200-bp moving window) methylated genes show statistically significant enrichment in certain biological pathways, we used pathways from Biocarta and KEGG databases curated by Molecular Signatures Database (MSigDB) (*http://www.broadinstitute.org/gsea/msigdb/index.jsp*, last accessed May 3, 2013; *http://www.biocarta.com/genes/index.asp*, and *http://www.genome.jp/kegg/pathway.html*, last accessed December 12, 2012).[37] One-sided Fisher's exact test (over-representation) was applied to calculate the statistical significance of pathway enrichment. The *P* values were corrected by the Benjamini–Hochberg procedure for multiple comparisons to calculate *q* values.

## Results

### Whole-Genome Methylation Analysis Suggests Partial Hypomethylation in Prostate Cancer and an Individualized Field Effect in Benign Tissues Adjacent to Tumor

To identify alterations of global genome methylation pattern in prostate cancer, microdissected genomic DNAs from the 13 prostate samples (T, AT, and age-matched OD) were bisulfite-treated and sequenced using Illumina HiSeq 2000, with an approximate coverage of $8 \times 10^{10}$ bp for each sample and 27-fold per allele. Moving windows of 200-bp width were used to scan the whole human genome (build hg19) and to aggregate the methylated and unmethylated calls from Bismark.[28,38] To ensure proper evaluation of methylation β value, each window was screened for a minimal of 50 calls on methylation status, based on 13 sequencing samples. This resulted in more than 8.8 million windows for methylation analysis. The majority (76%) of windows of human genome sequences appeared to be methylated (β = 0.5 to 1.0) (Figure 1A). Genome sequences with high methylation (average $\beta > 0.9$) or without methylation (average $\beta < 0.1$) from OD, AT, and T samples were identified. More than 1.8 million windows were identified as highly methylated in T samples (Figure 1B). In contrast, more than 2.6 million sequences were found with high methylation in OD samples. The AT samples had 2.27 million highly methylated windows. Interestingly, OD tissues also had a larger number of nonmethylated windows (70,837) ($\beta < 0.1$), compared with T (59,904) (Figure 1B). Approximately 8% (146,452) of highly methylated windows were unique to the T samples, but 27% (703,895) were unique to OD tissue. These results suggest that partial loss of hypermethylation may be an important epigenomic alteration in prostate cancer.

A total of 216,299 sequence windows were differentially methylated in T samples, compared with OD ($\Delta\beta > 0.4$, $q < 0.01$) (Figure 1C). Of these sequences, 173,297 (>80%) were differentially hypomethylated in T samples. In contrast, only 58,290 sequences were differentially methylated between T and AT samples. More than 93% (54,329) of these sequences were hypomethylated. Significant differential methylation was also found between AT and OD samples, with 36,839 sequences differentially methylated in AT, of which 31,336 (85.1%) were hypomethylated. Interestingly, only 1% of these sequences were located in CpG islands, and 2% to 4% in shore regions. Of these differentially methylated sequences, 2% to 4% were located in promoter regions of T compared with OD, or T compared with AT, or AT compared with OD (Figure 1D). A majority of the differentially methylated sequences were located in intergenic regions with no known function, suggesting that methylation of these sequences serves a function unrelated to gene transcription.

To investigate the relatedness of methylation patterns of these samples, a hierarchical clustering analysis of 13 samples based on β values of a filtered set of windows was performed. Three main groups of clustering were identified (Figure 1E). All four OD samples segregated into a branch, unrelated to either AT or T; however, three of the four AT samples clustered with the matched T sample. These results suggest that methylation in benign prostate adjacent to tumor is closely related to cancer rather than to normal prostate tissues, and indicates an individualized field effect.

To investigate the effect of methylation on gene expression, RNA sequencing was performed for each sample. Differentially methylated sequences ($\Delta\beta > 0.4$, $q < 0.01$) overlapping with known genes were selected to correlate with expression suppression of the overlapping genes (>0.5-fold down-regulation, $P < 0.05$). Methylations of 10,710 sequence windows (1418 genes) correlated with gene expression alteration for T versus OD samples, whereas 3591 sequences (828 genes) correlated for T versus AT samples (Figure 2). Also, 2978 eight sequences (908 genes) were differentially methylated between AT and OD samples.

Pathway analyses indicated differential methylation association with several important signaling pathways in T

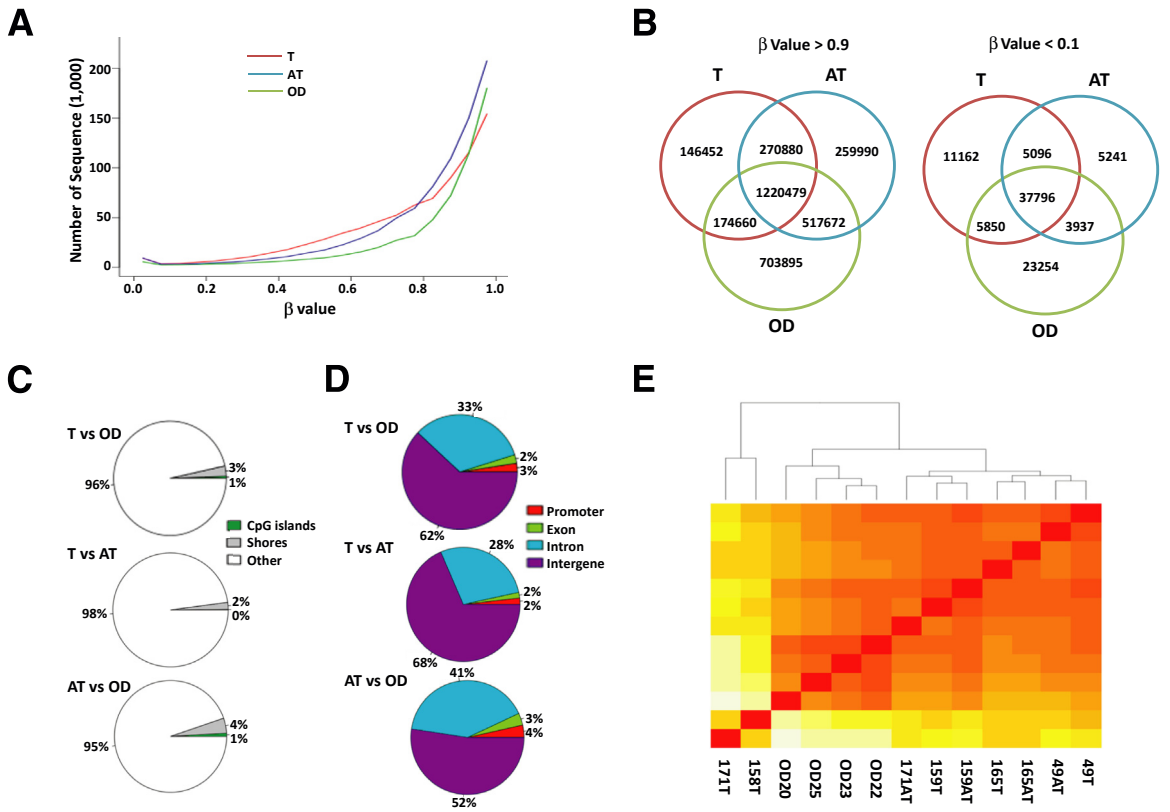


**Figure 1** Methylation of human genomes of T, AT, and OD prostate tissue. **A:** Methylation status of approximately 8.8 million genomic sequences (200-bp window) of T, AT, and OD. **B:** Distribution of hypermethylated ($\beta > 0.9$) and hypomethylated ($\beta < 0.1$) sequences in T, AT, and OD. **C:** Characteristics of differentially methylated sequences for T versus OD, T versus AT, and AT versus OD. **D:** Location distribution of differentially methylated sequences for the same three comparisons. **E:** Phylogenetic dendrogram of correlation of methylation β values among the 13 samples. AT, matched benign prostate tissues adjacent to cancer; OD, organ-donor prostate tissues; T, frozen specimens of prostate cancer.

samples, compared with OD samples, including hypermethylation and down-regulation of several tumor suppressor genes (eg, *RXRG*, *FH*, *CCDC6*, *RARA*, and *GPX3*) in the T samples. In addition, significant hypermethylation and down-regulation of genes in extracellular matrix, adhesion, and integrin signaling pathways were found in T samples (Supplemental Table S2), whereas several oncogenes [eg, *NCOA4* (alias *PTC3*), *BIRC2*, *FGF6*, and *HGF*] were hypomethylated and up-regulated (Supplemental Table S3). Significant numbers of calcium signaling pathway genes were down-regulated and hypermethylated in T samples, compared with AT, whereas the ubiquitin-mediated proteolysis pathway was hypomethylated (Supplemental Tables S4 and S5). These results suggest a critical role of DNA methylation in prostate cancer. Interestingly, significant demethylation and up-regulation were found in RAS and MAPK pathways in AT samples, compared with OD (Supplemental Tables S6 and S7), suggesting a precursor alteration that may lead to prostate cancer.

## CpG Island Methylation Analysis Indicates a Strong Field Effect of Prostate Cancer

CpG island methylation is widely credited for gene expression suppression. Our CpG island methylation-sequencing analyses showed that 34,493 to 36,530 of 66,765 (51.7% to 54.7%) CpG islands had at least one allele methylated across all samples. There was no significant statistical difference in the total number of CpG islands methylated among the different groups. Approximately 40% of the CpG islands had both alleles methylated ($\beta > 0.8$) in all three groups, with little variation among samples (Figure 3A and Supplemental Figure S1). When the CpG islands were categorized into promoter, intragene, and intergene CpG islands, 53.4% of the intragene CpG islands and 63.9% of the intergene islands were methylated in at least one allele. In contrast, only 16.2% of the CpG islands in the promoter region were methylated (Figure 3, B–D). When samples were segregated into OD, AT, and T types, a significant increase in one-allele CpG island methylation ($\beta = 0.5$ to 0.8) in T versus OD ($P = 0.04$) was found, whereas there was no significant difference between two-allele methylation ($\beta > 0.8$) across all groups. The frequencies of one-allele methylation of the intergene ($P = 0.017$) and promoter CpG ($P = 0.0017$) CPG islands were increased in the T samples, compared with the age-matched OD samples. Interestingly, similar differential methylations were also found in the AT versus OD comparison for either intergene ($P = 0.031$) or promoter ($P = 0.026$), albeit in lower frequencies.

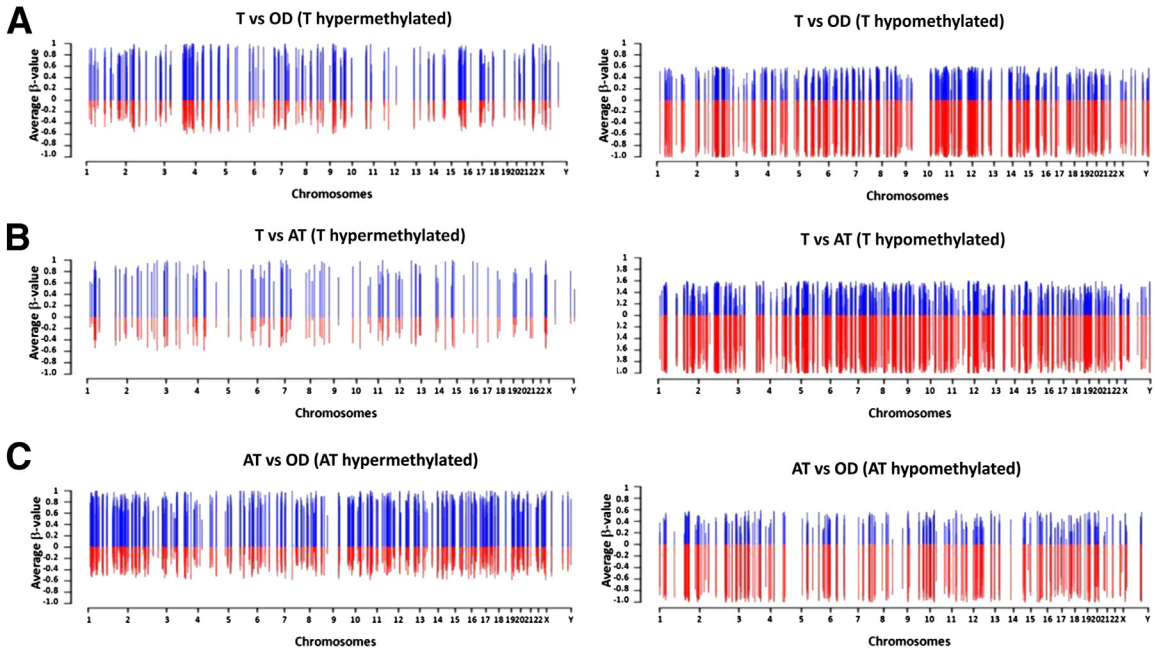


**Figure 2** Differential methylation of gene-associated differentially methylated sequences among T, AT, and OD tissues. Differential methylation of a sequence (200-bp window) is defined as a >0.5 difference in average β value, with $q < 0.01$, between T and OD, T and AT, or AT and OD.

To investigate whether specific gene-associated CpG methylations are associated with the development of prostate cancer, intragene and promoter CpG islands in the T samples were screened for a >0.2 difference in average β value ($P < 0.05$), compared with the corresponding CpG islands in the OD samples. The screening identified 2258 gene-associated CpG islands hypermethylated in T samples, relative to OD samples (Figure 4). Only 343 CpG islands had reduced methylation in T samples. Relative to the AT samples, however, only 41 CpG islands (12 promoters and 29 intragenes) were hypermethylated in the T samples, whereas 76 had reduced methylation. In contrast, 1035 gene-associated CpG islands were hypermethylated in the AT samples, relative to the OD samples; 774 (75%) of these CpG islands overlapped with those in the T samples, and 70 had reduced methylation. These results suggest that significant alterations in methylation patterns occur in prostate tissues adjacent to tumor and that these alterations are quite similar to those in prostate cancer.

## A Small Fraction of Methylated CpG Islands Correlates with Gene Expression Suppression

To evaluate the effect of CpG island methylation on gene expression, the top and bottom quantile of expressed genes were selected to determine the distribution of promoter CpG island methylation. Approximately 3.5-fold more genes were methylated in at least one allele (β = 0.6 to 1.0) in the promoter in the low-expressing versus top-expressing genes in T samples (Figure 5). Similar results were also found in OD and AT samples. To a lesser extent, such results were also found in intragene CpG island methylation distribution (Supplemental Figure S2).

To investigate whether all methylated CpG islands are functional in suppressing gene expression, differential expression of genes was determined among the T, AT, and OD groups according to the criteria of greater than twofold alteration and $P < 0.05$. Gene-associated CpG islands were deemed effective when differentially increased methylation ($P < 0.05$) resulted in down-regulation of mRNA transcripts in matched samples. Compared with the OD samples, in the T samples only 416/4901 (8.5%) CpG island methylations were effective in reducing the transcription of the corresponding genes (Figure 6A and Supplemental Figure S3). Interestingly, 354 (85%) of these CpG islands were located within the introns or internal exons of the genes, and only 62 (15%) CpG islands were located in the promoter/exon 1 regions. With the AT samples used as baseline values, 100/1172 (8.5%) CpG islands were determined to be effective in suppressing gene expression in the T samples. The majority of these CpG islands [88/100 (88%)] were located in intron or internal exon regions, and 12 were considered classical promoter CpG islands. Similar findings were obtained in comparison of AT versus OD samples: 184/2512 (7.3%) CpG islands were found to be effective in the AT samples, and 164 (89%) of these CpG islands were intragene islands. These results suggest that intragene CpG islands play a critical role in regulating gene expression. To rule out a data-specific effect, concordance analyses were performed of the four TCGA cancer data sets. Less than 15% of CpG island differential methylation correlated with down-regulation of RNA expression in these data sets (Supplemental Figure S4). This largely confirmed our findings of an extensive presence of nonfunctional CpG islands in the human genome.

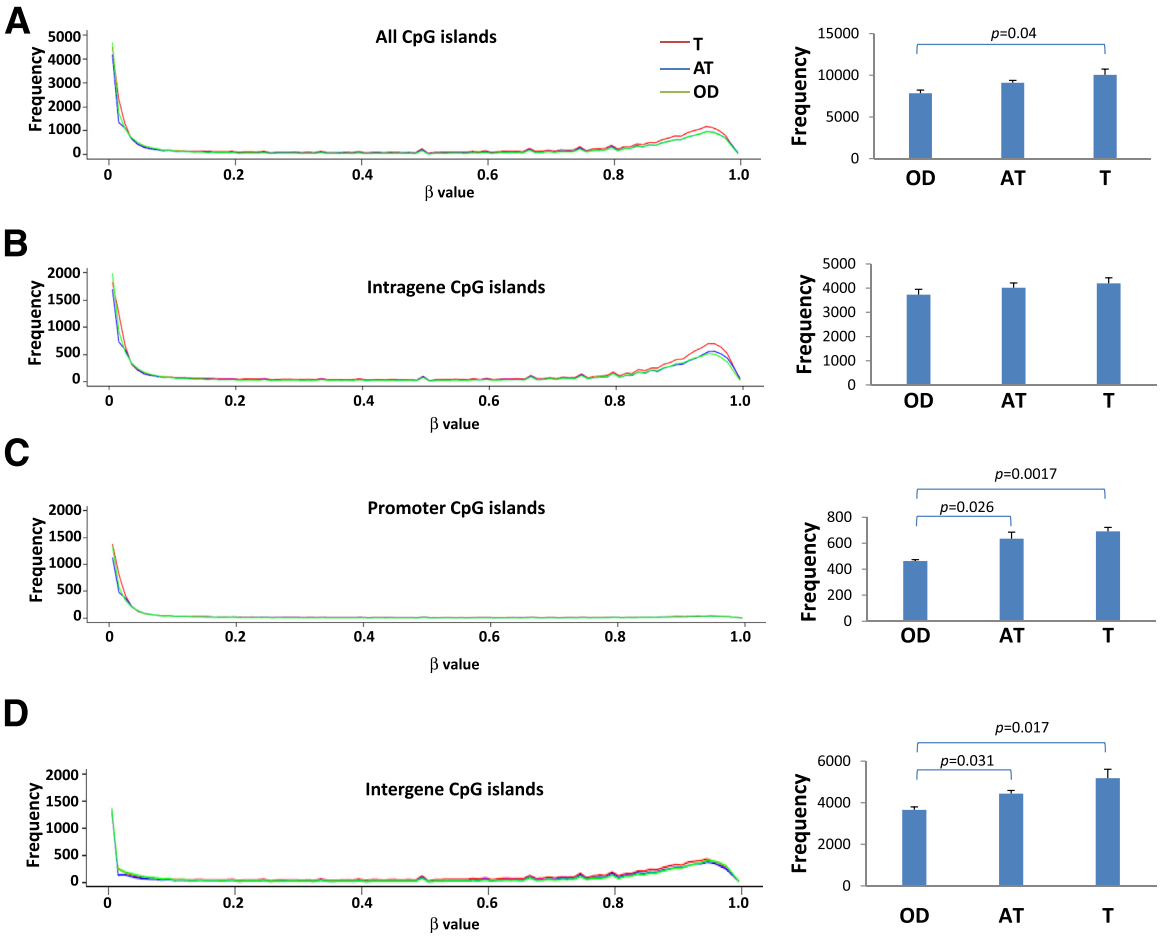


**Figure 3** Distribution of methylation of CpG islands in OD, T, and AT tissues expressed as the average β value of each CpG island and the average frequency of one-allele methylation (β = 0.5 to 0.8) for all CpG islands (**A**) and for the intragene (**B**), promoter (**C**), and intergene (**D**) categories.

To investigate whether the level of CpG methylation in a given CpG island correlates with the expression suppression activity, the differences in CpG island methylation were correlated with the differences in RNA transcription levels. Surprisingly, we found poor correlation between the quantity of methylated CpGs and suppression of RNA transcription (Pearson coefficient = 0.1 to 0.2). The levels of transcription suppression by intragene or promoter CpG islands were not significantly different (Figure 6B). Analyzing the length, GC content, and CpG count of the CpG islands (normalized to CpG island length) and quantifying the levels of CpG methylation in the TATA box regions yielded largely insignificant results between effective and noneffective CpG islands (Supplemental Table S8). These results suggest that the density of CpG dinucleotide methylation, the GC contents of CpG islands, and the distribution of methylated CpGs are not the critical factors in gene suppression by DNA methylation.

To investigate whether a sequence element in a CpG island contributes to the effectiveness of a CpG island in suppressing gene transcription when methylated, sequence motifs enriched in effective CpG islands (differential methylation and differential expression at $P < 0.05$) were compared with noneffective ones (differential methylation and differential expression at $P > 0.5$) using the MEME motif analysis tool.[35] A CCCC[GT]CCCCC motif was significantly enriched in the effective CpG islands. Similar motif analysis was then applied to the effective intragene T group, as well as to four TCGA cancer data sets. The top enriched motifs for each data set were identified, and significant sequence homology among these motifs was found (Figure 7). Subsequently, these top enriched motifs from each data set were matched against the curated online TCGA database using the tool TOMTOM.[36] The results indicate that a sequence motif rich in guanine nucleotides significantly close to a known binding motif for Sp1 was found in all the data sets except HNSC (Figure 7). The wide presence of this motif in effective CpG islands suggests a possible critical role of this sequence in enabling the transcription suppression of methylated CpG islands.

To investigate whether differential methylation of intergene CpG islands also occurs among the OD, AT, and T groups, the samples were screened for CpG islands with a >0.2 difference in methylation β value ($P < 0.05$). Relative to OD samples, 1689 intergene CpG islands were hypermethylated and 341 were hypomethylated in the

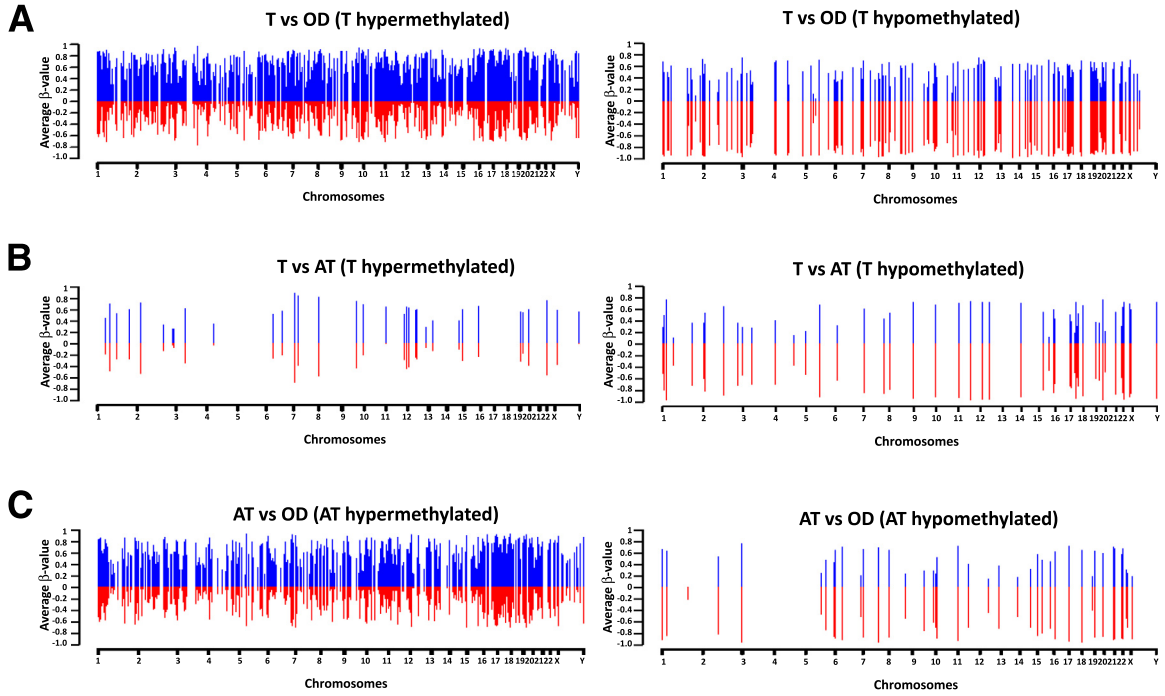


**Figure 4** Differential methylation of gene-associated CpG islands among T, AT, and OD tissue. Differential methylation is defined as a >0.2 difference of average β values and $P < 0.05$ for T versus OD (**A**), T versus AT (**B**), or AT versus OD (**C**).

T samples (Supplemental Figure S5). Relative to AT samples, only 43 intergene CpG islands were hypermethylated and 71 were hypomethylated in the T samples. Significant differential methylations were also found in the AT group, relative to the age-matched OD group: 835 hypermethylated and 65 hypomethylated CpG islands. The functions of these CpG islands have yet to be characterized. The differential methylation status of these islands in our analyses, however, suggests that they may play a role in carcinogenesis.

## Discussion

Whole-genome methylation sequencing is a new and important tool with which to gain insight into the development of cancer. Compared with array-based or methylation CpG pull-down sequencing, bisulfite sequencing provides a substantial advantage in both resolution and accuracy. First, bisulfite whole-genome sequencing determines the methylation status of CpG dinucleotide across the entire genome and mitochondria DNA, whether within or outside of a CpG island. Second, bisulfite sequencing quantifies the level of methylation in an accurate manner such that we can determine minute methylation variation in any CpG dinucleotide. Because of the highly quantitative feature, this method enables detection of subtle differences in methylation between groups. This is unlikely to be achieved with other methodologies. To our knowledge, the present study is the first application of whole-genome bisulfite sequencing in prostate cancer.

Our analysis suggests that the majority of hypermethylated sequences were located outside the boundary of CpG islands. More than 95% of alterations of CpG methylation in cancer occur in sequences unrelated to CpG islands. The results of genomic differential methylation among T, AT, and OD samples were robust and significant when the analysis was performed using the entire genomic sequence. The numbers of differential methylation sequences were dramatically reduced, however, when the analysis was focused only on CpG islands. These findings suggest that potential functionally significant methylation information lies outside CpG

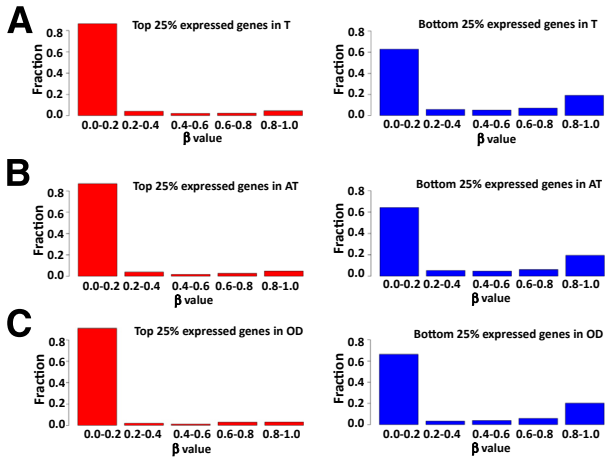


**Figure 5** Low-level gene expression is associated with higher frequency of promoter CpG island hypermethylation. Promoter CpG island methylation β value distribution of top and bottom quantile of expressed genes in T (**A**), AT (**B**), and OD (**C**) tissue.

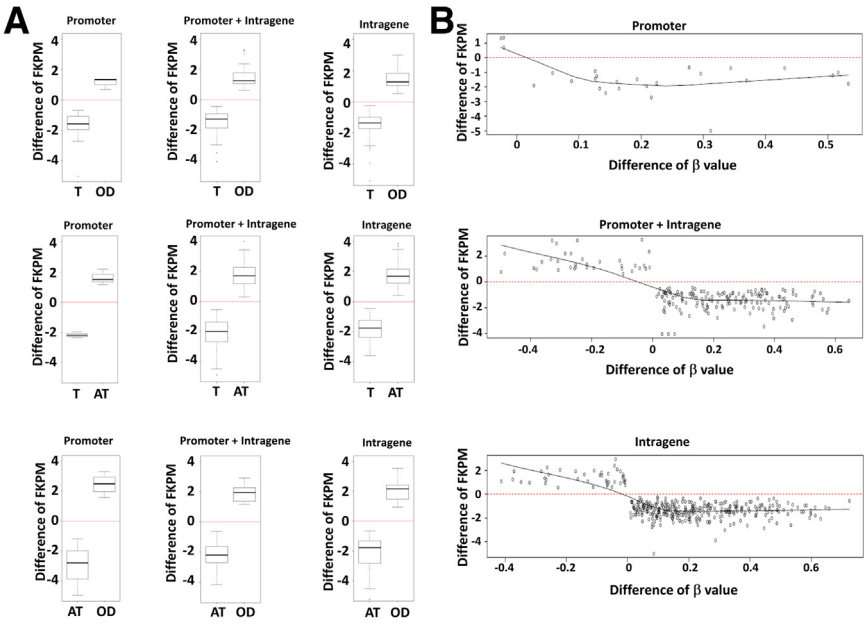


**Figure 6** Effect of CpG island methylation on RNA expression. **A**: Correlation of CpG island methylation on RNA expression. Effective CpG islands of promoter and/or intragene were screened by the difference at $P < 0.05$ in β values between T and OD, A and AT, or AT and OD and RNA expression with greater than twofold down-regulation at $P < 0.05$. **B**: Concordance of CpG island differential methylation and suppression of RNA expression. Differences in methylation levels (β values) were plotted against differences in RNA expression (FKPM). Each data point represents an average value from each CpG island that is deemed effective in suppressing RNA expression.

islands. Large numbers of differentially methylated sequences (both CpG island and non-CpG island) exist in the intergene regions, suggesting that the function of methylation of genome sequences probably serves more than just regulation of gene transcription.

A surprising finding was that more than 90% of differential CpG island methylations had minimal effect on gene expression when matched RNA sequencing was performed to quantify the transcription level. The number of effective differential CpG island methylations was barely greater than 400 when T samples were compared with ODs. There was a lack of correlation between the effectiveness of CpG methylation-mediated suppression and levels of methylation, size of CpG islands, CpG counts, or TATA box region methylation. This suggests that CpG island methylation is probably not the overriding factor in suppressing gene transcription. Our motif analyses suggest a significant enrichment of transcription binding complex in most of the effective CpG islands in the promoter regions. Similar sequences were also found enriched in the effective groups from three of the four TCGA cancer data sets (PRAD, BRCA, and THCA). All these motif sequences share the Sp1 binding sites. Thus, transcription factor binding is implicated for methylation-induced gene transcription suppression. This argument is strengthened by the finding that most of the intragene CpG islands effective in suppressing gene transcription also contain the similar sequence. Our findings suggest that the transcription suppression activity of CpG islands is quite independent of the location relative to the RNA polymerase binding site (TATA), whether in the promoter or intragene region.

Previous studies have suggested a significant field effect of prostate cancer through methylation genome array analyses.[39–42] Our whole-genome methylation sequencing analyses affirm a significant field effect of prostate cancer in benign tissues adjacent to tumor. This is evidenced by the close similarity in methylation patterns between the T and AT tissues in both whole-genome CpG and CpG island methylation analyses. The numbers of CpG islands differentially methylated between age-matched OD tissues and T tissues were more than 20-fold higher than between T and AT tissues, even if these tissues were located in different quadrants or different sides from the cancers.

The present findings suggest that epigenomic changes occur in morphologically nonmalignant cells. These changes

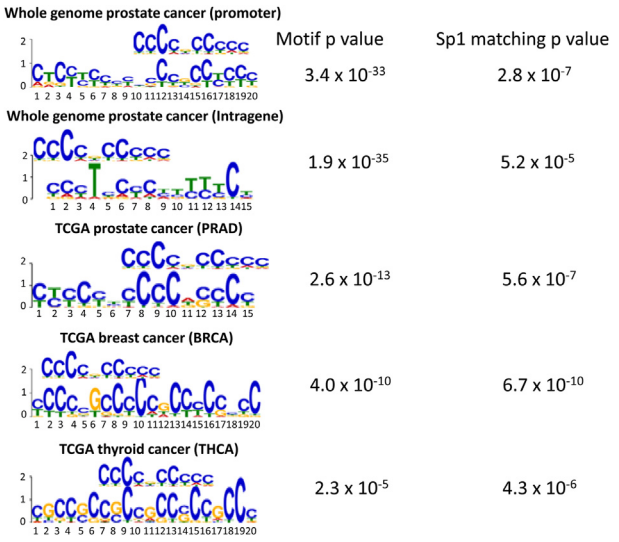


**Figure 7** Motif enrichment in effective CpG islands. Motif identification using MEME programs among CpG islands with an increase at $P < 0.05$ of methylation resulted in greater than twofold decrease of RNA expression at $P < 0.05$ in the prostate cancer whole-genome methylation sequencing set and in the PRAD, BRCA, and THCA data sets from TCGA.

likely form the basis of a cancerous epigenome that propels the benign tissue to develop into morphological cancer. Two good examples of OD-to-T transition in methylation patterns are the *LDLRAD2* and *FGF2* genes (Supplemental Figure S3). The effective CpG island of *LDLRAD2* is intronic, whereas the CpG island of *FGF2* is in promoter/exon 1 regions. OD samples had low levels of methylations in both genes; however, significant elevation of methylation of CpGs was found in AT samples. The highest levels of methylation were found in T samples. This methylation negatively affected the expression of these two genes (Supplemental Figure S3). The significant methylation pattern differences in benign prostate tissues in a prostate gland containing cancer (the AT group), relative to normal prostate without cancer (the OD group), have clinical significance, because a finding of such a pattern signifies the presence of prostate cancer in a nearby location. Indeed, recent studies suggest that methylation of some field-effect genes is useful in detecting nearby prostate cancer.[43,44] A mis-hit prostate biopsy with an AT methylation pattern warrants additional examination for the presence of cancer in the prostate gland.

## Supplemental Data

Supplemental material for this article can be found at *http://dx.doi.org/10.1016/j.ajpath.2013.08.018*.

## References


1. Siegel R, Naishadham D, Jemal A: Cancer statistics, 2012. CA Cancer J Clin 2012;(62):10–29
2. Luo JH: Gene expression alterations in human prostate cancer. Drugs Today (Barc) 2002, 38:713–719
3. Luo JH, Yu YP, Cieply K, Lin F, Deflavia P, Dhir R, Finkelstein S, Michalopoulos G, Becich M: Gene expression analysis of prostate cancers. Mol Carcinog 2002, 33:25–35
4. Yu YP, Landsittel D, Jing L, Nelson J, Ren B, Liu L, McDonald C, Thomas R, Dhir R, Finkelstein S, Michalopoulos G, Becich M, Luo JH: Gene expression alterations in prostate cancer predicting tumor aggression and preceding development of malignancy. J Clin Oncol 2004, 22:2790–2799
5. Yu YP, Song C, Tseng G, Ren BG, LaFramboise W, Michalopoulos G, Nelson J, Luo JH: Genome abnormalities precede prostate cancer and predict clinical relapse. Am J Pathol 2012, 180:2240–2248
6. Dhanasekaran SM, Barrette TR, Ghosh D, Shah R, Varambally S, Kurachi K, Pienta KJ, Rubin MA, Chinnaiyan AM: Delineation of prognostic biomarkers in prostate cancer. Nature 2001, 412:822–826
7. Kim JH, Dhanasekaran SM, Mehra R, Tomlins SA, Gu W, Yu J, Kumar-Sinha C, Cao X, Dash A, Wang L, Ghosh D, Shedden K, Montie JE, Rubin MA, Pienta KJ, Shah RB, Chinnaiyan AM: Integrative analysis of genomic aberrations associated with prostate cancer progression. Cancer Res 2007, 67:8229–8239
8. Magee JA, Araki T, Patil S, Ehrig T, True L, Humphrey PA, Catalona WJ, Watson MA, Milbrandt J: Expression profiling reveals hepsin overexpression in prostate cancer. Cancer Res 2001, 61:5692–5696
9. Ernst T, Hergenhahn M, Kenzelmann M, Cohen CD, Ikinger U, Kretzler M, Hollstein M, Gröne HJ: Genexpressions-Analyse des Prostatacarcinomes [Gene expression profiling in prostatic cancer]. German. Verh Dtsch Ges Pathol 2002, 86:165–175
10. Glinsky GV, Glinskii AB, Stephenson AJ, Hoffman RM, Gerald WL: Gene expression profiling predicts clinical outcome of prostate cancer. J Clin Invest 2004, 113:913–923
11. Monk M, Grant M: Preferential X-chromosome inactivation, DNA methylation and imprinting. Dev Suppl 1990:55–62
12. Hu JF, Vu TH, Hoffman AR: Promoter-specific modulation of insulin-like growth factor II genomic imprinting by inhibitors of DNA methylation. J Biol Chem 1996, 271:18253–18262
13. Yu G, Tseng GC, Yu YP, Gavel T, Nelson J, Wells A, Michalopoulos G, Kokkinakis D, Luo JH: CSR1 suppresses tumor growth and metastasis of prostate cancer. Am J Pathol 2006, 168:597–607
14. Yu YP, Yu G, Tseng G, Cieply K, Nelson J, Defrances M, Zarnegar R, Michalopoulos G, Luo JH: Glutathione peroxidase 3, deleted or methylated in prostate cancer, suppresses prostate cancer growth and metastasis. Cancer Res 2007, 67:8043–8050
15. Brooks JD, Weinstein M, Lin X, Sun Y, Pin SS, Bova GS, Epstein JI, Isaacs WB, Nelson WG: CG island methylation changes near the GSTP1 gene in prostatic intraepithelial neoplasia. Cancer Epidemiol Biomarkers Prev 1998, 7:531–536
16. Lou W, Krill D, Dhir R, Becich MJ, Dong JT, Frierson HF Jr., Isaacs WB, Isaacs JT, Gao AC: Methylation of the CD44 metastasis suppressor gene in human prostate cancer. Cancer Res 1999, 59:2329–2331
17. Curradi M, Izzo A, Badaracco G, Landsberger N: Molecular mechanisms of gene silencing mediated by DNA methylation. Mol Cell Biol 2002, 22:3157–3173
18. Liu L, Yoon JH, Dammann R, Pfeifer GP: Frequent hypermethylation of the RASSF1A gene in prostate cancer. Oncogene 2002, 21:6835–6840
19. Graff JR, Herman JG, Lapidus RG, Chopra H, Xu R, Jarrard DF, Isaacs WB, Pitha PM, Davidson NE, Baylin SB: E-cadherin expression is silenced by DNA hypermethylation in human breast and prostate carcinomas. Cancer Res 1995, 55:5195–5199
20. Herman JG, Merlo A, Mao L, Lapidus RG, Issa JP, Davidson NE, Sidransky D, Baylin SB: Inactivation of the CDKN2/p16/MTS1 gene is frequently associated with aberrant DNA methylation in all common human cancers. Cancer Res 1995, 55:4525–4530
21. Kito H, Suzuki H, Ichikawa T, Sekita N, Kamiya N, Akakura K, Igarashi T, Nakayama T, Watanabe M, Harigaya K, Ito H: Hypermethylation of the CD44 gene is associated with progression and metastasis of human prostate cancer. Prostate 2001, 49:110–115
22. Vanaja DK, Cheville JC, Iturria SJ, Young CY: Transcriptional silencing of zinc finger protein 185 identified by expression profiling is associated with prostate cancer progression. Cancer Res 2003, 63:3877–3882
23. Woodson K, Hayes R, Wideroff L, Villaruz L, Tangrea J: Hypermethylation of GSTP1, CD44, and E-cadherin genes in prostate cancer among US blacks and whites. Prostate 2003, 55:199–205
24. Yoshiura K, Kanai Y, Ochiai A, Shimoyama Y, Sugimura T, Hirohashi S: Silencing of the E-cadherin invasion-suppressor gene by CpG methylation in human carcinomas. Proc Natl Acad Sci USA 1995, 92:7416–7419
25. Bastian PJ, Yegnasubramanian S, Palapattu GS, Rogers CG, Lin X, De Marzo AM, Nelson WG: Molecular biomarker in prostate cancer: the role of CpG island hypermethylation. Eur Urol 2004, 46:698–708
26. Nakayama M, Gonzalgo ML, Yegnasubramanian S, Lin X, De Marzo AM, Nelson WG: GSTP1 CpG island hypermethylation as a molecular biomarker for prostate cancer. J Cell Biochem 2004, 91:540–552
27. Kang GH, Lee S, Lee HJ, Hwang KS: Aberrant CpG island hypermethylation of multiple genes in prostate cancer and prostatic intraepithelial neoplasia. J Pathol 2004, 202:233–240
28. Krueger F, Andrews SR: Bismark: a flexible aligner and methylation caller for bisulfite-seq applications. Bioinformatics 2011, 27:1571–1572

29. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L: Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 2012, 7:562−578
30. Trapnell C, Pachter L, Salzberg SL: TopHat: discovering splice junctions with RNA-seq. Bioinformatics 2009, 25:1105−1111
31. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L: Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 2010, 28: 511−515
32. Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, Mason CE: methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. Genome Biol 2012, 13:R87
33. Jones PA, Takai D: The role of DNA methylation in mammalian epigenetics. Science 2001, 293:1068−1070
34. Takai D, Jones PA: The CpG island searcher: a new WWW resource. In. Silico Biol 2003, 3:235−240
35. Bailey TL, Elkan C: Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol 1994, 2:28−36
36. Gupta S, Stamatoyannopolous J, Bailey T, Noble W: Quantifying similarity between motifs. Genome Biol 2007, 8:R24
37. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP: Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA 2005, 102:15545−15550
38. Wu H, Caffo B, Jaffee HA, Irizarry RA, Feinberg AP: Redefining CpG islands using hidden Markov models. Biostatistics 2010, 11:499−514
39. Kobayashi Y, Absher DM, Gulzar ZG, Young SR, McKenney JK, Peehl DM, Brooks JD, Myers RM, Sherlock G: DNA methylation profiling reveals novel biomarkers and important roles for DNA methyltransferases in prostate cancer. Genome Res 2011, 21: 1017−1027
40. Kim JH, Dhanasekaran SM, Prensner JR, Cao X, Robinson D, Kalyana-Sundaram S, Huang C, Shankar S, Jing X, Iyer M, Hu M, Sam L, Grasso C, Maher CA, Palanisamy N, Mehra R, Kominsky HD, Siddiqui J, Yu J, Qin ZS, Chinnaiyan AM: Deep sequencing reveals distinct patterns of DNA methylation in prostate cancer. Genome Res 2011, 21:1028−1041
41. Yu YP, Paranjpe S, Nelson J, Finkelstein S, Ren B, Kokkinakis D, Michalopoulos G, Luo JH: High throughput screening of methylation status of genes in prostate cancer using an oligonucleotide methylation array. Carcinogenesis 2005, 26:471−479
42. Luo JH, Ding Y, Chen R, Michalopoulos G, Nelson J, Tseng G, Yu YP: Genome-wide methylation analysis of prostate tissues reveals global methylation patterns of prostate cancer. Am J Pathol 2013, 182: 2028−2036
43. Truong M, Yang B, Livermore A, Wagner J, Weeratunga P, Huang W, Dhir R, Nelson J, Lin DW, Jarrard DF: Using the epigenetic field defect to detect prostate cancer in biopsy negative patients. J Urol 2013, 189:2335−2341
44. Yang B, Bhusari S, Kueck J, Weeratunga P, Wagner J, Leverson G, Huang W, Jarrard DF: Methylation profiling defines an extensive field defect in histologically normal prostate tissues associated with prostate cancer. Neoplasia 2013, 15:399−408