NEOPLASIA
www.neoplasia.com

Volume 15 Number 4 April 2013 pp. 373–383 373

Epigenomic Alterations in Localized and Advanced Prostate Cancer[1,2]

Pei-Chun Lin*,†, Eugenia G. Giannopoulou‡,§, Kyung Park*, Juan Miguel Mosquera*, Andrea Sboner*,§, Ashutosh K. Tewari¶, Levi A. Garraway#,, Himisha Beltran††, Mark A. Rubin*,†,3 and Olivier Elemento‡,§,3**

*Department of Pathology and Laboratory Medicine, Weill Cornell Medical College, New York, NY; †Graduate Program in Biochemistry and Structural Biology, Cell and Developmental Biology, and Molecular Biology, Graduate School of Medical Sciences, Weill Cornell Medical College, New York, NY; ‡Department of Physiology and Biophysics, Weill Cornell Medical College, New York, NY; §HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine, Weill Cornell Medical College, New York, NY; ¶Department of Urology and the LeFrak Center for Robotic Surgery, Weill Cornell Medical College, New York, NY; #Broad Institute of Harvard and MIT, Cambridge, MA; **Department of Medical Oncology, Dana Farber Cancer Institute, Boston, MA; ††Department of Medicine, Hematology Oncology Division, Weill Cornell Medical College, New York, NY

Abstract

Although prostate cancer (PCa) is the second leading cause of cancer death among men worldwide, not all men diagnosed with PCa will die from the disease. A critical challenge, therefore, is to distinguish indolent PCa from more advanced forms to guide appropriate treatment decisions. We used Enhanced Reduced Representation Bisulfite Sequencing, a genome-wide high-coverage single-base resolution DNA methylation method to profile seven localized PCa samples, seven matched benign prostate tissues, and six aggressive castration-resistant prostate cancer (CRPC) samples. We integrated these data with RNA-seq and whole-genome DNA-seq data to comprehensively characterize the PCa methylome, detect changes associated with disease progression, and identify novel candidate prognostic biomarkers. Our analyses revealed the correlation of cytosine guanine dinucleotide island (CGI)–specific hypermethylation with disease severity and association of certain breakpoints (deletion, tandem duplications, and interchromosomal translocations) with DNA methylation. Furthermore, integrative analysis of methylation and single-nucleotide polymorphisms (SNPs) uncovered widespread allele-specific methylation (ASM) for the first time in PCa. We found that most DNA methylation changes occurred in the context of ASM, suggesting that variations

Abbreviations: ASM, allele-specific methylation; CGI, cytosine guanine dinucleotide (CpG) island; CRPC, castration-resistant prostate cancer; ERRBS, Enhanced Reduced Representation Bisulfite Sequencing; PCa, prostate cancer

Address all correspondence to: Olivier Elemento, PhD, Department of Physiology and Biophysics, HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine, Weill Cornell Medical College, 1305 York Avenue, New York, NY 10065. E-mail: ole2001@med.cornell.edu or Mark A. Rubin, MD, Department of Pathology and Laboratory Medicine, Weill Cornell Medical College, 1300 York Avenue C-410A, New York, NY 10065. E-mail: rubinma@med.cornell.edu

[1]This study was funded by the Early Detection Research Network grant U01 CA 11275-07 (M.A.R.), the Prostate Cancer Foundation Young Investigator Award (H.B.), and the National Science Foundation CAREER grant (O.E.).

[2]This article refers to supplementary materials, which are designated by Tables W1 to W4 and Figures W1 to W4 and are available online at www.neoplasia.com.

[3]These authors share the senior authorship.

Received 20 December 2012; Revised 28 January 2013; Accepted 29 January 2013

DOI 10.1593/neo.122146

in tumor epigenetic landscape of individuals are partly mediated by genetic differences, which may affect PCa disease progression. We further selected a panel of 13 CGIs demonstrating increased DNA methylation with disease progression and validated this panel in an independent cohort of 20 benign prostate tissues, 16 PCa, and 8 aggressive CRPCs. These results warrant clinical evaluation in larger cohorts to help distinguish indolent PCa from advanced disease.

Neoplasia (2013) 15, 373–383

Introduction

Prostate cancer (PCa) is the most common cancer in men worldwide [1]. Although 1 in 6 men will likely be diagnosed with PCa during his lifetime, only 1 in 36 will ultimately die from the disease, reflecting a 5-year survival rate close to 100% [2]. Overtreatment is a major concern in PCa, as many patients diagnosed with clinically localized PCa do not require definitive treatment [3,4]. Patients presenting with advanced forms, including metastatic and castration-resistant prostate cancer (CRPC), however, have much worse outcomes; therefore, distinguishing indolent from advanced PCa has been an imperative task in PCa research.

In this study, we sought to investigate how genome-wide DNA methylation patterns might help differentiate between indolent and advanced PCa. We also reasoned that single-nucleotide resolution analysis of DNA methylation might provide mechanistic insights into the evolution of PCa toward more advanced forms. DNA methylation primarily arises at cytosine guanine dinucleotide (CpG) sites and is associated with epigenetic regulation of gene expression [5]. Perturbed DNA methylation patterns have been shown to arise during PCa tumorigenesis and have been implicated in PCa etiology and disease progression [6]. DNA methylation profiling studies using microarrays, MethylPlex–next-generation sequencing, and MeDIP-Seq have indeed identified a large number of DNA methylation changes in PCa [7–11]. However, few studies have investigated DNA methylation changes in CRPC, in part due to limited availability of tissues. DNA hypomethylation has been found to occur during PCa progression by evaluating 5-methylcytosine content in genomic DNA, while hypermethylation at certain CpG island (CGI) promoter genes were found in both localized PCa and metastatic PCa [10,12–14]. An array-based DNA methylation profiling study of CRPC recently found that alterations in DNA methylation arose more frequently than mutations or copy number changes [15], further strengthening the rationale for interrogating metastatic PCa to identify biomarkers of DNA methylation related to PCa disease progression. To our knowledge, the present study is the first to perform broad coverage single-nucleotide resolution analysis of CRPC.

We previously used next-generation sequencing to characterize the whole genome and transcriptome of seven clinically localized PCa together with matched benign adjacent prostate tissues as well as of seven CRPC cases [16,17]. In an effort to understand how aberrant DNA methylation may contribute to PCa and to the metastatic phenotype, we profiled the global DNA methylation patterns of these same cases using Enhanced Reduced Representation Bisulfite Sequencing (ERRBS). ERRBS with single-base resolution gives a broader genome-wide coverage by extending to CGI shores, compared to Reduced Representation Bisulfite Sequencing [18]. Since bisulfite sequencing provides single-base resolution, it allowed us to investigate how frequently allele-specific methylation (ASM) occurs in genetically and epigenetically unstable PCa and CRPC samples. ASM is mainly studied in the context of genomic imprinting, whose role is to ensure and maintain parent-of-origin effects on gene expression in a small set of genes [19,20]; its occurrence in non-imprinted genomic regions of cancer cells is less clear. In this study, we integrated data of methylation and SNPs to delineate and understand ASM in clinical PCa cases.

We report novel insights into DNA methylation patterns in PCa progression, association between DNA methylation and breakpoints, ASM changes, and the identification of a panel of CGIs highly related to PCa progression. We propose that differential methylation changes could be used not only as early detection biomarkers but also as new prognostic biomarkers in assisting treatment management decisions.

Materials and Methods

Prostate Tissue Samples

All tissue samples (benign prostate tissues and PCa and CRPC samples) were collected as part of an Institutional Review Board–approved protocol at Weill Cornell Medical College (WCMC). PCa (hormone naïve PCa) and CRPC were obtained, and DNA extraction was performed as previously described [16,17]. None of the PCa samples were from patients who received preoperative hormonal therapy, chemotherapy, or radiotherapy. Hematoxylin and eosin (H&E) slides of each prostate tissue sample were evaluated by the study's pathologists (M.A.R., K.P., and J.M.M.) to determine areas of benign and cancer; tumors were selected for high-density cancer foci (>90% tumor tissue). Biopsy cores (1.5 mm) from frozen tissue blocks (ERRBS) or formalin-fixed, paraffin-embedded (FFPE) blocks (validation) were prepared for DNA extraction. Table W1 shows Gleason grade, age, preoperative serum prostate-specific antigen levels, and staging of all of the patients from whom prostate tissues were obtained for this study. Seven separately collected benign prostatic tissue samples matched to the cancer tissues (obtained from the same patients) were used as normal controls. Samples for the validation cohort consisted of 20 benign prostate tissues and 16 PCa and 8 CPRC samples. The benign prostate tissues also came from the matched PCa and CRPC patients. All CRPC samples were from patients with variable degrees of neuroendocrine differentiation, and three of the eight cases were pure neuroendocrine histology and lacked protein androgen receptor (AR) expression. The PCa samples used in ERRBS were from patients without any recurrence following radical prostatectomy for 5 to 6 years and 3 to 7 years for the PCa patients in the validation cohort (surgeries performed between 2006 and 2010). Benign prostate tissues were benign glands from a cancer-free region of the matched PCa or CRPC patients, taken from a separate frozen or FFPE tissue block without a cancer region as determined by the study's pathologists. DNA from FFPE blocks was extracted using the DNeasy Blood and Tissue Kit (Qiagen, Germantown, MD) according to the manufacturer's instructions and then evaluated

by the FFPE Quality Control Kit (Illumina, San Diego, CA). Samples that passed the quality control (QC) test were used in the validation. Clinical parameters of samples used in ERRBS and the validation study are listed in Tables W1 and W2.

Enhanced Reduced Representation Bisulfite Sequencing

Sample preparations were performed at the WCMC Epigenomics Core Facility as previously described [18]. In brief, the preparation steps included: 1) MspI enzyme digestion; 2) end repair of digested DNA; 3) adenylation; 4) adapter ligation, with pre-annealed 5-methylcytosine–containing Illumina adapters; 5) isolation of library fragments of 150 to 400 bp from a 1.5% agarose gel (using low-range ultra-agarose from Bio-Rad, Des Plaines, IL); 6) bisulfite conversion using the EZ DNA Methylation Kit (Zymo Research, Irvine, CA) with the following changes: i) C_T conversion incubation in a thermocycler (Eppendorf, Hauppauge, NY) with the condition: 30 seconds at 95°C followed by 15 minutes at 50°C for 55 cycles, and ii) elution into nuclease-free water; 7) polymerase chain reaction (PCR) amplification; each library was prepared with FastStart High Fidelity DNA Polymerase (Roche, Indianapolis, IN) and Illumina PCR primers PE1.0 and 2.0. The thermocycler conditions were given as follows: 5 minutes at 94°C, 18 cycles of 20 seconds at 94°C, 30 seconds at 65°C, 1 minute at 72°C, followed by 3 minutes at 72°C. PCR products were isolated using Agencourt AMPure XP (Beckman Coulter, Brea, CA) beads per manufacturer's recommended protocol (Agencourt) All amplified libraries were evaluated using a Qubit 1.0 fluorometer and Quant-iT dsDNA HS Assay Kit (Invitrogen, Grand Island, NY) for quantitation and bioanalyzer visualization (Agilent 2100 Bioanalyzer; Agilent, Santa Clara, CA). The WCMC Computational Genomics Core Facility supported ERRBS data analysis.

Quantitative DNA Methylation Analysis by MassARRAY EpiTYPER

Measurement of DNA methylation levels was performed at the WCMC Epigenomics Core Facility with matrix-assisted laser desorption ionization/time-of-flight mass spectrometry using MassARRAY EpiTYPER assays (Sequenom, San Diego, CA) as previously described [21]. Genomic DNA was bisulfite-converted using the EZ DNA Methylation Kit according to the manufacturer's protocol (Zymo Research). The spectra's methylation ratios were then calculated using EpiTYPER software v1.0 (Sequenom). The EpiTYPER primers were designed through the Sequenom EpiDesigner web site at http://www.epidesigner.com/start3.html. The reverse primers were tagged with a T7 promoter sequence: cagtaatacgactcactatagggagaaggct. EpiTYPER primer sequences are listed in Table W4.

Computational Approaches

Bisulfite-treated read alignment and methylation calls were performed as previously described [18]. Briefly, bisulfite reads were aligned to the bisulfite-converted hg19 reference genome using Bismark [22]. All samples had bisulfite conversion rates >99.5%. Differentially methylated CpGs were identified using the Fisher exact test with correction for Benjamini-Hochberg multiple testing. Differentially methylated regions (DMRs) were defined as regions containing at least five differentially methylated CpGs [false discovery rate (FDR) = 10%] and whose total methylation difference was more than 10%. DMRs were annotated using ChIPseeqerAnnotate from the ChIPseeqer package [23]. Motif and pathway analyses were performed with default parameters using FIRE [24] and iPAGE [25], respectively. When comparing methylation at CGIs between groups of samples, we used the LIMMA approach on log-transformed methylation values. Methylation of a specific region was calculated by averaging the methylation levels of all covered CpGs in that region. Clonality rates were calculated by counting how many times rare non-converted cytosine among reads mapped at the same genomic location and orientation, as described in Gertz et al. [26]. Random Forests [27] and 10-fold cross-validation were used to build classification models based on the samples in the validation cohort. Both classification and the estimation of sensitivity, specificity, and area under the receiver operating characteristic (ROC) curve [area under curve (AUC)] were performed in R using the Caret package (at http://caret.r-forge.r-project.org/).

Data Deposition Statement

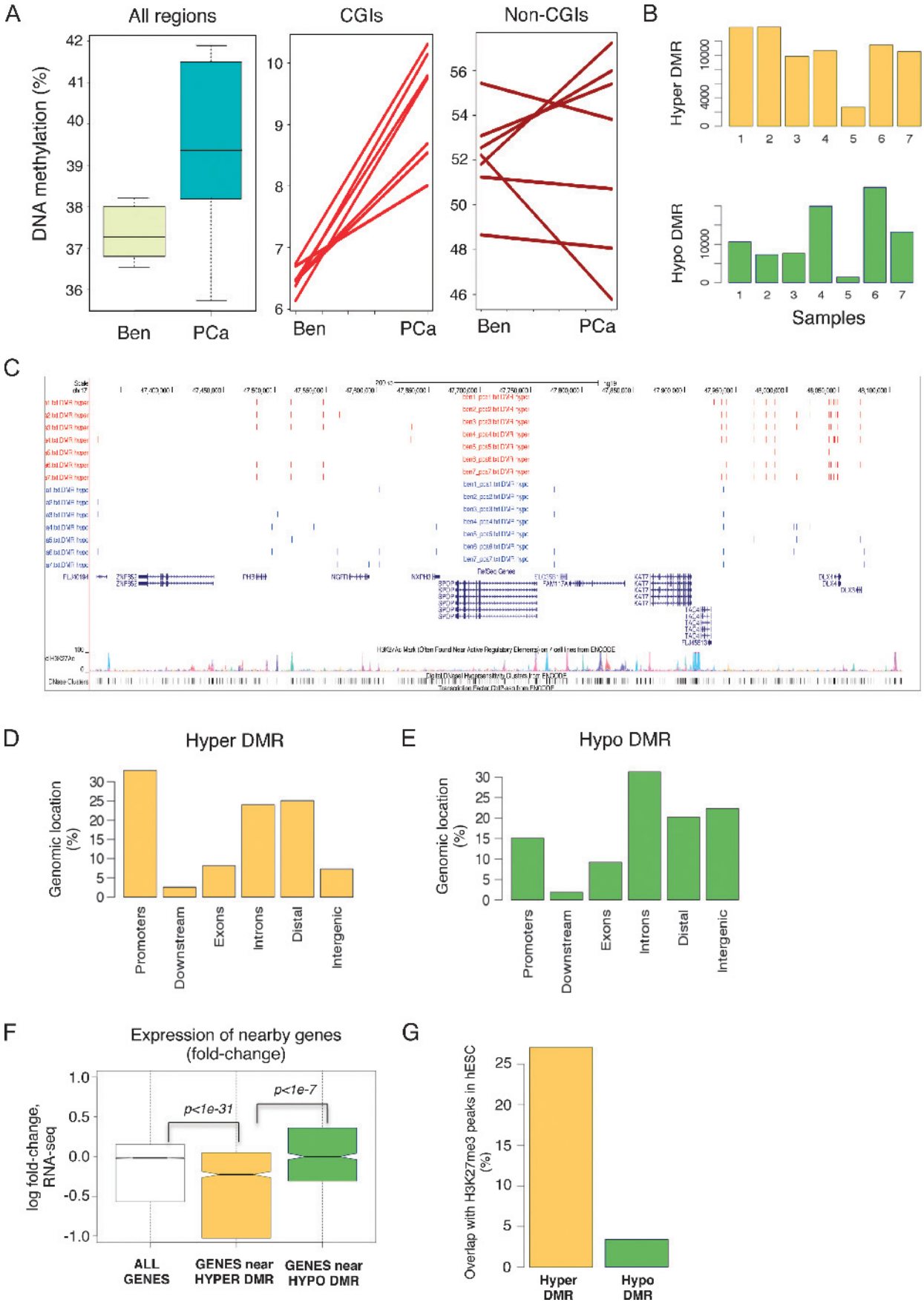
The ERRBS data have been deposited for public access in the National Center for Biotechnology Information's (NCBI) gene expression omnibus (GEO) database and is accessible through GEO Series accession number GSE41701. DNA-seq and RNA-seq data are available as described in Beltran et al. and Berger et al. [16,17].

Results

Profound and Widespread Changes in DNA Methylation in Prostate Adenocarcinoma

We first interrogated seven pairs of localized PCa and their matched benign prostate tissues. ERRBS provided more than 10-fold sequencing coverage on >2.5 million CpG sites genome-wide for each sample. By combining methylation levels from all interrogated CpG sites, we observed an overall increase in DNA methylation in PCa (Figure 1*A*), driven in large part by a sharp and systematic methylation increase at CGIs (P = .0002). The changes at non-CGI regions (defined as at least 10 kb away from known CGIs) were more diverse and less systematic among PCa patients (P = .85); there was a trend toward global hypomethylation in four of the seven cases. This analysis also indicated that global methylation levels were significantly more heterogeneous in PCa compared to benign prostate tissues. We confirmed this heterogeneity when we determined DMRs within each pair and detected differing numbers of hypermethylated and hypomethylated regions in each case (Figure 1*B*), with some cases showing more hypomethylated DMRs while others showed more hypermethylated DMRs. Despite this heterogeneity and the different genetic background of each patient, we found that many hypermethylated and hypomethylated DMRs resided at the same locations between individuals, suggesting that aberrant DNA methylation in PCa follows a predetermined path (Figure 1*C*). When we examined DNA methylation patterns across different genomic regions, we found that hypermethylated DMRs were mostly enriched at promoters [defined as ±2 kb windows centered on RefSeq transcription start sites (TSS); Figure 1*D*]. However, hypomethylated DMRs were mostly located at introns and intergenic regions (Figure 1*E*).

We then sought to integrate RNA-seq data performed on the same samples to examine the relationship between aberrant methylation and gene expression changes between PCa and matched benign tissue from the same patient. We found that overall genes located near hypermethylated DMRs had decreased expression (P < 1e–31), while genes near hypomethylated DMRs had increased expression (Figure 1*F*; data shown for patient 1). This indicates that, within controlled isogenic conditions, the frequently assumed but broadly elusive negative correlation between DNA methylation and gene expression tends to hold.


A
All regions
CGIs
Non-CGIs
DNA methylation (%)
Ben
PCa
B
Hyper DMR
Hypo DMR
Samples
C
D
Hyper DMR
Genomic location (%)
Promoters
Downstream
Exons
Introns
Distal
Intergenic
E
Hypo DMR
F
Expression of nearby genes (fold-change)
log fold-change, RNA-seq
p<1e-31
p<1e-7
ALL GENES
GENES near HYPER DMR
GENES near HYPO DMR
G
Overlap with H3K27me3 peaks in hESC (%)
Hyper DMR
Hypo DMR

We also found that hypermethylated DMRs overlapped strongly with regions enriched for H3K27me3 in human embryonic stem cells [28] (Figure 1*G*), confirming previous reports of such concurrent epigenomic changes in PCa cell lines and other cancer types [29–32].

De novo DNA motif analysis was carried out to examine the sequences of DMRs. We focused the analysis on differentially methylated CGIs, since these regions behaved in a consistent way across cases (Figure 1*A*). We used non-differentially methylated CGIs as control, thus controlling for CpG content. This analysis identified three DNA motifs preferentially enriched with hypermethylated DMRs, including one that resembled the motif bound by the transcription factor early growth response gene 1 (EGR1; Figure W1*A*). When examining RNA-seq data, we found that EGR1 expression levels were lower in PCa compared to benign prostate tissues (Figure W1*B*). This down-regulation was confirmed by immunohistochemistry staining (Figure W1*C*). Although more extensive work is required to investigate the role of EGR1 in PCa, we speculate that EGR1 binding to CGIs may prevent DNA methylation, while loss of EGR1 expression may lead to increased DNA methylation.

Structural Variations in PCa Are Correlated with DNA Methylation Patterns in Benign Tissues

The previously described genomic characterization of our seven PCa samples provided a unique opportunity to integrate genomic and epigenomic information and better understand the events that lead to progression from benign lesion to PCa [17]. To evaluate whether DNA methylation patterns in normal tissues may influence the formation of genomic alterations that lead to PCa, we examined structural variation breakpoints that have occurred in PCa and determined their DNA methylation levels in matched benign prostate tissues. For this analysis, we examined DNA methylation in 10-kb windows centered on each breakpoint. We found that in benign tissues, deletion and tandem duplication breakpoints that occurred in PCa were less methylated than control random regions ($P = .004$ and $P = .026$, respectively); however, interchromosomal translocation breakpoints were more methylated than expected ($P = .008$; Figure 2*A*). Thus, certain structural variation breakpoints have non-random methylation levels in benign prostate tissues, hinting at a potential role of the epigenome in mediating these alterations. Moreover, Berger et al. found that CpG sites had significantly higher mutation rate (more than 10-fold) than all other genomic positions did [17]. When we examined CpGs that were mutated in PCa samples, we found that these CpGs were highly methylated in their matched benign tissues (Figure 2*B*). By comparison, a random selection of non-mutated CpGs showed much lower DNA methylation in the benign tissues. Most mutations occurring at CpGs were C-to-T transition mutations; however, we also found the less frequent C-to-A and C-to-G transversion mutations. The C-to-T transition may come from spontaneous deamination of methylated cytosine to thymidine. While many of these mutations may be passenger mutations that have been deaminated in the initial malignant clone, our results altogether suggest that many genetic alterations in PCa have a non-random methylation signature in normal prostate cells. Moreover, our results suggest that non–C-to-T mutations at methylated cytosines occur at a non-negligible rate and that highly methylated CpG sites are inherently fragile and mutation prone, while non-methylated CpGs are protected from alterations.

Widespread Allele-Specific DNA Methylation in PCa

Taking advantage of single CpG resolution and high coverage provided by ERRBS, we sought to investigate whether certain regions of the methylome showed ASM and whether ASM levels differ between matched tumors and benign tissues. We therefore mined the ERRBS data and looked for statistical associations between SNPs (defined here as mismatches between the samples and the reference genome) and CpG methylation within the same ERRBS reads (Figure 3*A*). Comparing CpGs covered in both benign tissues and matched tumors, we identified between 60,000 and 150,000 CpGs that coincided with a heterozygous SNP within the same read and for which we could perform a statistical test of association (Fisher exact test) [17]. After correction for multiple testing (Benjamini-Hochberg, FDR = 20%), we identified several thousand CpGs with ASM in both benign tissues and PCa. When we compared the fraction of CpGs with ASM, we observed a significant increase in PCa compared to benign tissues ($P = .0009$, *t* test; Figure 3*B*). Importantly, we found that unequal read clonality (PCR duplicates) could not explain these differences, as clonality levels were approximately equal between PCa and matched benign tissues (Figure W2).

We speculated that allele-specific changes in methylation could be responsible for a significant fraction of all changes in DNA methylation (that is, one allele undergoes methylation changes, while the other does not). Accordingly, we found that ASM preferentially occurred in differentially methylated CpGs and DMRs, for both hypermethylated and hypomethylated (Figure 3, *C* and *D*). Altogether, these results suggest that a large fraction of DNA methylation changes occur on a single-allele basis. We also sought to determine whether ASM occurred in the context of allele-specific expression. While we found overlaps (i.e., genes whose promoter showed ASM and with different allelic usage between DNA-seq and RNA-seq; not shown), the coverage differences between ERRBS, DNA-seq, and RNA-seq data limited the analysis and precluded a definitive conclusion.

DNA Methylation at CGIs Increases with Disease Severity

To examine how DNA methylation patterns may have evolved in more advanced PCa, we profiled the DNA methylome of six CRPC samples from patients who failed endocrine therapy and/or developed a predominantly androgen-independent PCa associated with lack of AR expression and extensive neuroendocrine differentiation [16]. Examining DNA methylation levels at CGIs, we found that CRPC

Figure 1. Overall increased DNA methylation in PCa, consistent across cases. (A) Percentages of DNA methylation in all regions, CGIs, and non-CGIs of matched benign prostate tissues (Ben) and PCa. (B) Numbers of hypermethylated or hypomethylated DMRs of individual PCa samples. (C) Example of hypermethylated and hypomethylated DMRs, many of which occur independently in different patients. (D) Percentage of hypermethylated regions occurring within indicated genomic locations. (E) Percentage of hypomethylated regions occurring within indicated genomic locations. (F) Expression fold changes of nearby genes in regard of hypermethylated or hypomethylated DMRs (data shown for case 1). (G) Percentage of hypermethylated or hypomethylated DMRs overlapped with H3K27me3 peaks in human embryonic stem cells (hESC).

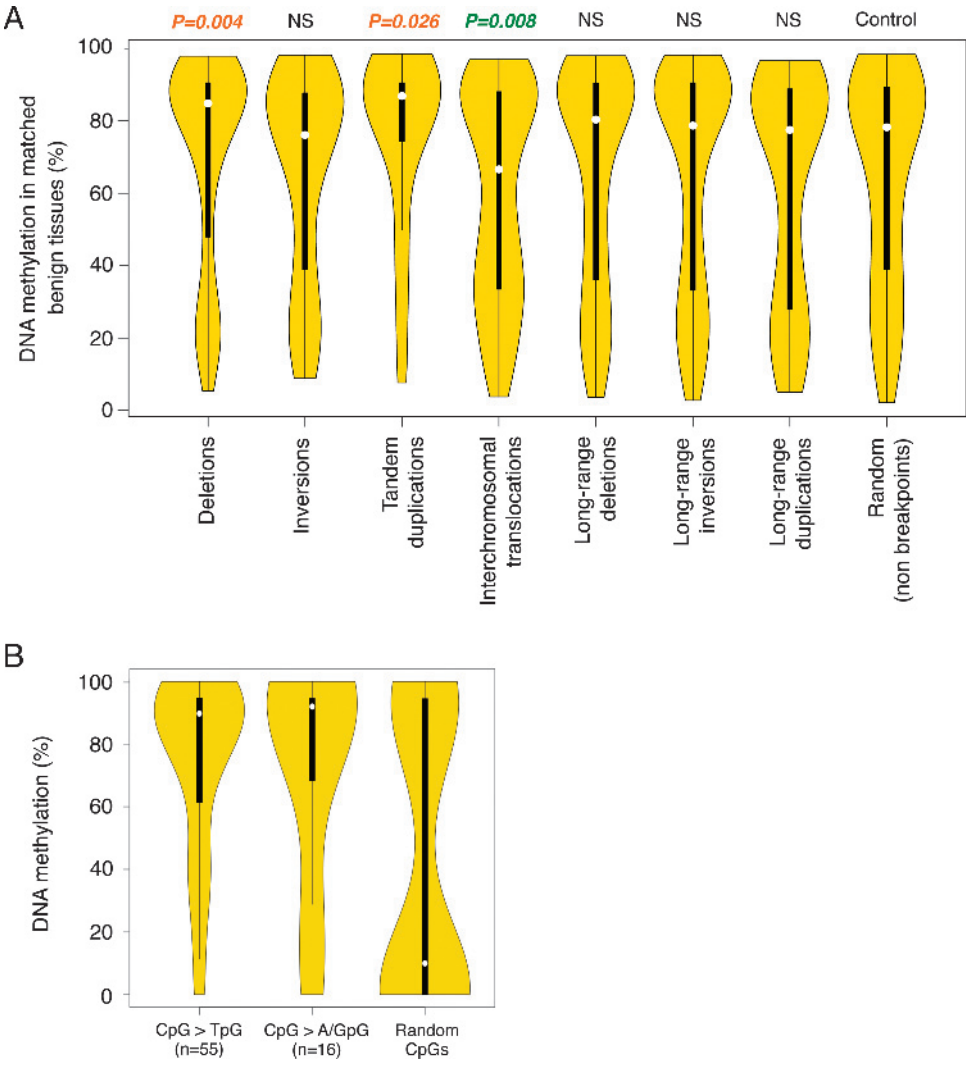


Figure 2. DNA methylation patterns in benign prostate tissues are associated with certain types of mutations in PCa. (A) Percentage of DNA methylation in matched benign tissues for various breakpoints. The width of the violin plots correlates with the fraction of mutations with corresponding methylation levels (*Y* axis). NS, not significant. (B) Percentage of DNA methylation of CpG sites in benign prostate tissues, where C>T, C>A/G mutations occur.

had higher levels compared to PCa (Figure 4*A*). Taken together, our analyses indicate that DNA methylation levels at CGIs increase with disease severity from benign tissue to PCa onto CRPC. There was also a less systematic and less pronounced average trend toward DNA methylation increase with disease severity in non-CGI regions (Figure 4*A*). Nevertheless, DNA methylation levels in CRPC were highly variable among cases. To further investigate the global DNA methylation increase, we focused on CGIs and performed gene ontology analysis of the genes near CGIs that were differentially methylated in PCa and CRPC using iPAGE [25] (LIMMA analysis, FDR = 5%; Figure W3*A*). Interestingly, the genes associated with hypermethylated CGIs had specific enrichments related to G-protein signaling, cAMP-mediated signaling, and hormone activity. This suggests thatDNA methylation increases during the transition between indolent PCa and highly aggressive CRPC are at least partially functionally coherent. Wealso found that EGR1 mRNA levels are further decreased in CRPC compared to PCa, implicating the potential role of EGR1 down-regulation in enabling CGI hypermethylation (Figure W3*B*).

We then sought to identify a subset of these CGIs whose methylation increases with disease severity. We determined CGIs whose methylation levels increased systematically between benign prostate tissues and PCa (LIMMA analysis, FDR = 5%, n = 692). We also performed the same analysis for CGIs whose methylation levels increased consistently and reproducibly between PCa and CRPC (LIMMA analysis, FDR = 5%, n = 1229). We found that the overlap between two groups of hypermethylated CGIs was higher than expected (P < 1e−104, hypergeometric P value; Figure 4*B*). This indicates that, to a large extent, the same genomic regions gain methylation in the transition from benign prostate tissue to PCa, and this trend continues onto CRPC. There were 228 CGIs in this overlap and the DNA methylation heat map of the 228 CGIs provided a visual demonstration of the positive association between PCa disease progression and increasing DNA methylation levels (Figure W3*C*). Cluster analysis based on CGI methylation not only further confirmed the strong DNA methylation signature uncovered here but also showed that some PCa samples clustered with CRPC samples, suggesting a propensity of these cases to

progress toward a more advanced disease (Figure W3*D*). As a control, we performed the same analysis for hypomethylated CGIs; in that analysis, hypomethylated CGIs between benign tissues and PCa and between PCa and CRPC did not overlap more frequently than expected by chance (Figure W3*E*). These results indicate that unlike hypermethylation, hypomethylation may occur randomly during progression to advanced PCa. Likewise, when we analyzed ASM in CRPC, we found that ASM occurred at the same levels as benign prostate tissues (Figure W3*F*).

We then sought to validate these observations on an independent patient cohort. We integrated gene expression from the RNA-seq data with the 228 CGIs and selected a panel of 13 CGIs for validation (Table W3). The 13 CGIs showed persistently increasing hypermethylation from benign prostate tissues to PCa, and from PCa to CRPC, while genes associated with these CGIs also demonstrated differential expression (LIMMA, FDR = 5%, fold change > 1.5). Among the panel of 13 CGIs, three associated genes, *general receptor for phosphoinositides* (*GRASP*), *glutathione S-transferase π 1* (*GSTP1*), and *tropomyosin 4* (*TPM4*), have been reported as differentially methylated in PCa *versus* normal tissues in Mahapatra et al. [8], thus providing independent validation of our analysis. Tissues from FFPE blocks of an independent cohort of 20 benign prostate tissues and 16 PCa and 8 CRPC samples were used to examine the panel of 13 CGIs. We used MassARRAY EpiTYPER assay to directly interrogate methylation of the 13 CGIs, thus seeking not only biologic validation in an independent cohort but also technical validation of the methylation measurements provided by ERRBS. As shown in Figure 4, *C* and *D*, we observed the same increase in DNA methylation correlating with disease severity in FFPE samples using MassARRAY EpiTYPER assay as that revealed by ERRBS on frozen tissues. We found that all 13 CGIs had increased methylation levels between benign and primary PCa cases (FDR = 5%); 8 of the 13 CGIs had increased methylation levels between primary cases and CRPC. Intriguingly, four advanced cases (FF18, 55, 57, and 59) that had no previous treatment showed high levels of DNA

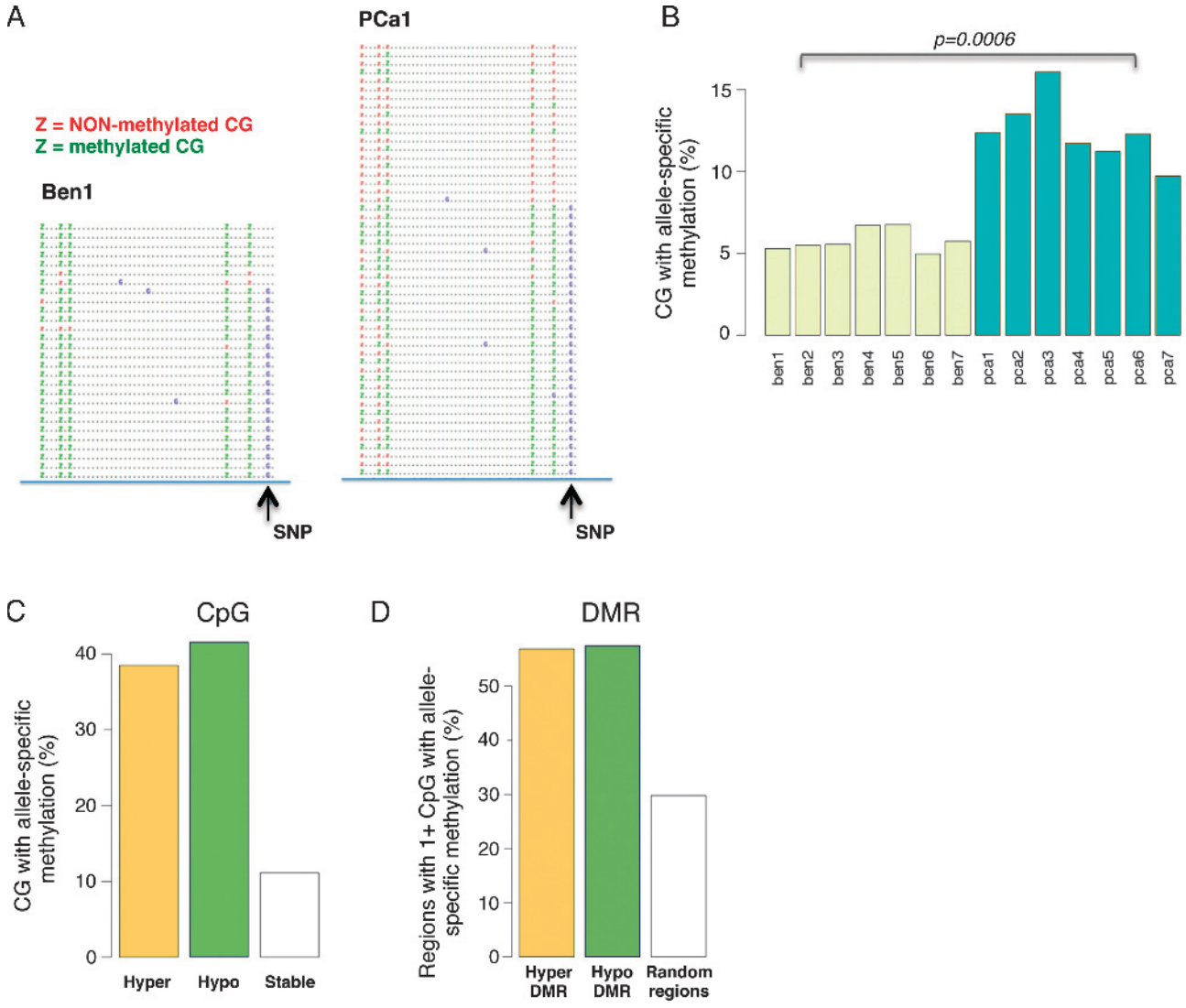


Figure 3. Allele-specific DNA methylation is highly frequent in PCa. (A) Methylated and non-methylated CpGs correlate with the presence or absence of a nearby SNP in the first pair of benign prostate tissue and PCa. (B) Fraction of CpGs associated with ASM in benign prostate tissue and PCa. Only CpGs with a nearby SNP are considered in this analysis. (C) Fraction of CpGs with ASM in hypermethylated or hypomethylated CGIs. (D) Fraction of regions containing at least one CpG with ASM in hypermethylated or hypomethylated DMRs.

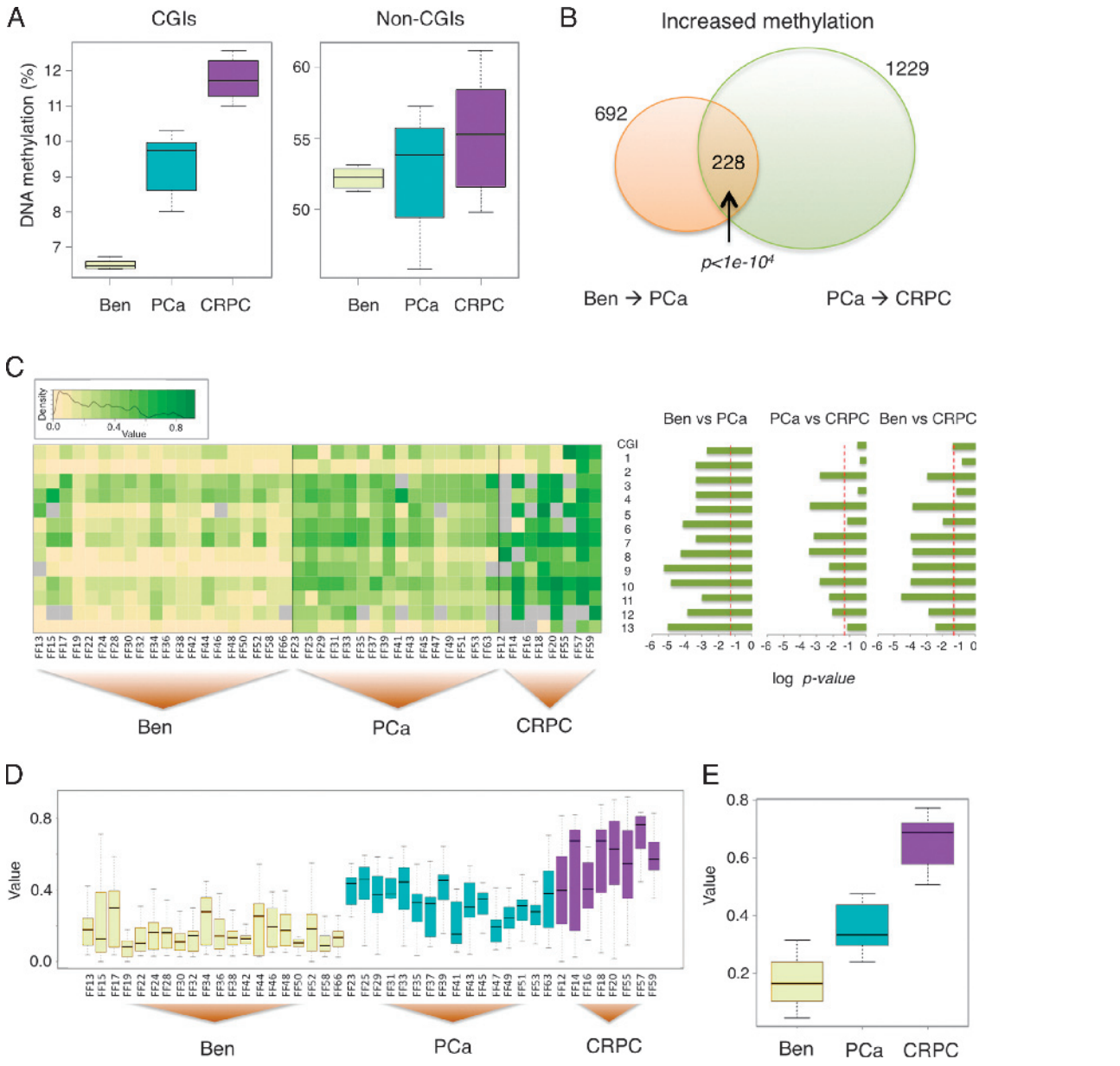


Figure 4. DNA methylation increases with and may predict disease severity. (A) Percentage of DNA methylation in CGIs and non-CGIs of benign prostate tissues (Ben), PCa, and CRPC. (B) Venn diagram displays the methylation overlap observed between two sections: increased methylation from benign prostate tissues to PCa and increased methylation from PCa to CRPC; the inset number represents the coverage in each section. Two hundred twenty-eight CGIs appear in both sections with $P < 1e{-}104$. (C) Left: Heat map of DNA methylation levels of the panel of 13 CGIs in the three groups, 20 benign prostate tissues and 16 PCa and 8 CRPC samples (value range, 0 to 1). Right: Comparison between the three groups with adjusted log P values (Benjamini-Hochberg correction FDR controlled at 0.05). (D) Boxplots of the average DNA methylation levels of the panel for individual samples. (E) Boxplots of the average DNA methylation levels of the panel for the three groups. The AUC, sensitivity, and specificity for the comparison of benign and PCa are 0.9375, 100%, and 75%, and those for the comparison of PCa and CRPC are 0.975, 95%, and 95%.

methylation at these selected CGIs. Using Random Forests [27], we built two binary classification models to discriminate benign prostate tissues from PCa, as well as PCa from CRPC, based on the methylation levels of the 13 CGIs. In 10-fold cross-validation, the AUC scores for the predictive models were high, 0.9375 and 0.975, respectively (Figure 4*E*), indicating a high sensitivity and specificity. Taken together, our data suggest that both models provide excellent discriminatory power and the panel may be useful in predicting disease aggressiveness at diagnosis.

Discussion

In this study, we performed global DNA methylation profiling in primary PCa samples, matched benign prostate tissues, and CRPC and integrated DNA methylation patterns with whole-genome and transcriptome sequencing data. We used ERRBS to profile genome-wide DNA methylation status, thus interrogating >2.5 M single CpGs in each sample. ERRBS has recently been developed to capture more broadly genome-wide DNA methylation at single-base resolution, including CGI shores [18]. ERRBS yields quantitative comparisons

of methylation levels and enhances methylation detection at various genomic regions. By providing deep coverage of regulatory regions, especially CGIs, and low coverage of repeat regions (e.g., LINE1 elements), ERRBS enables a range of new analyses, including ASM analyses.

The analyses described in this study led to a number of new findings. First, by examining global methylation patterns in benign prostate tissue *versus* PCa *versus* CRPC, we found increasing levels of DNA methylation at CGIs correlated with increasing severity of the disease. As most CRPCs still retain AR expression, the methylation patterns observed in this study may not be representative of most CRPCs. Our motif analysis indicates that EGR1 DNA motifs are overrepresented at hypermethylated CGIs in PCa and CRPC. Previous studies using genome-wide chromatin immunoprecipitation (ChIP) assays in other cell types have indeed shown that EGR1 binds heavily to CGIs [33]. Our analyses also showed that EGR1 mRNA and protein levels decreased with disease severity. These data suggest that down-regulation of EGR1 may enable DNA methylation to accumulate at CGIs, perhaps through the recruitment of CGI-binding protein complexes that include DNA methyltransferase; however, a broader evaluation of EGR1 expression and a further investigation of the role of EGR1 in PCa disease progression are warranted. Of the 228 CGIs that showed increasing methylation during disease progression, we selected a panel of 13 CGIs for further validation, as the expression of their associated genes also exhibited significant changes and they may play important roles in PCa progression, especially toward a late stage and lethal neuroendocrine phenotype. We propose examination of the methylation status of this panel of 13 CGIs as putative diagnostic and prognostic biomarkers, which may aid in directing treatment decisions.

According to RNA-seq data, the majority of genes on this panel were downregulated (70%, 9 of the 13 CGIs). *GSTP1*, a well-known DNA methylation biomarker for PCa, was among the associated genes shown to have a continuous decline in gene expression during PCa disease progression. Its silencing may render prostate cells more sensitive to environmental insults and consequently more susceptible to genomic damage [6]. *Capping protein* (*actin filament*), *gelsolin-like* (*CAPG*), a proposed tumor suppressor in certain cancers [34], has been shown to modulate cancer cell invasion in breast and PCa cell lines [35]. *Retinoic acid receptor responder 2* (*RARRES2*), also known as chemerin or tazarotene-induced gene 2, was found to attract dendritic cells and macrophages to the site of inflammation [36]. As such, down-regulation of RARRES2 may affect the dynamics of the PCa microenvironment. TPM4 has been connected to human cancer, even though its role is less clear. Invasive squamous cervical cancer showed lower expression of TPM4, while lymph node metastasis in breast cancer has a significant association with TPM4 up-regulation [37,38]. *Family with sequence similarity 107, member A* (*FAM107A*), also known as *DRR1* or *TU3A*, was previously reported to be downregulated and deleted in renal cell carcinoma, and the expression of FAM107A inhibited cell growth [39]. *Filamin A interacting protein 1–like* (*FILIP1L*) has also been shown downregulated in ovarian cancer, and its down-regulation is inversely correlated with promoter methylation and advanced phenotype [40].

Furthermore, the association of gene-body methylation with active gene expression has been found in human cells and other organisms [41,42]. Similarly, *T-box containing transcription factor* (*TBX1*) was found to have a hypermethylated CGI around its second exon and elevated gene expression in our cohort. A functional link has previously been established between TBX1 and retinoic acid (RA) signaling where overexpressed TBX1 inhibits RA signaling, in part, due to down-regulation of ALDH1a2 and PCNA [43,44]. As down-regulation of ALDH1a2 was detected in human PCa and in the TRAMP mouse model of PCa [45,46], this study suggests a regulation of ALDH1a2 and RA signaling through up-regulation of TBX1 by DNA methylation during PCa progression. Likewise, the CGI that overlapped with the second exon of *potassium voltage-gated channel subfamily C* (*KCNC2*) was also hypermethylated in PCa, with concurrent up-regulation of the gene. KCNC2 encodes a membrane protein for voltage-gated potassium channel. Its function in PCa is unclear and may be related to the neuroendocrine phenotype of CRPC in this cohort. Intriguingly, both RARRES2 and TBX1 were identified in the panel and related to aberrant regulation of RA signaling during PCa progression. Even though we found that their hypermethylation resulted in opposite directions of gene expression (i.e., down-regulation in RARRES2 and up-regulation in TBX1), the overall outcome may down-regulate the RA signaling.

Another important finding in our study is that certain breakpoints, like deletion, tandem duplications, or interchromosomal translocations, are associated with distinct DNA methylation patterns in premalignant prostate tissue. We also found that highly methylated CpG sites are likely more fragile and more frequently targeted for somatic base mutations during PCa tumorigenesis compared to unmethylated CpGs. However, given that the benign tissues are benign glands taken from a cancer-free region of the matched PCa patient, we cannot rule out that epigenomic changes may have arisen in these morphologically normal tissues. Thus, the methylation patterns detected in benign tissues and associated with PCa breakpoints may not be present in the truly normal tissues (e.g., from autopsy) taken from other individuals. Altogether, our study points to an important role of the epigenome in mediating or facilitating a broad range of genomic alterations in PCa. Whether this role is more central in PCa than in other malignancies is currently unclear.

Our study revealed for the first time (to the best of our knowledge) frequent ASM changes in PCa. It also showed that a vast fraction of DNA methylation changes might occur on a single-allele basis. The mechanisms behind these single-allele methylation changes are unclear and may involve differential binding of transcription factors to each allele due to the presence of SNPs or different nuclear localization patterns. ASM was recently found at the H3K27me3-bound regions in a study of bisulfite sequencing of chromatin-immunoprecipitated DNA using PCa cell lines [30]. Thus, it is possible that a complex interplay exists between DNA methylation and H3K27me3, and this interplay precludes changes from occurring on certain alleles marked with either epigenetic mark.

Recent large-scale genome-wide DNA methylation profiling studies have discovered large numbers of DMRs. Our analysis on paired samples further identified a consistent trend in many DMRs located similarly across the entire genome, suggesting that DNA methylation changes may follow certain genetic cues during PCa pathogenesis. Moreover, we noted that within paired samples hypermethylation was significantly correlated with down-regulation of gene expression, while hypomethylation associates with up-regulation. Although this concept is broadly used, recent genome-wide studies have observed more complex regulations [18,42].

Both hypermethylation and hypomethylation take place in PCa as well as other types of cancer. On the basis of the analysis of 5-methylcytosine content of genomic DNA, previous studies have suggested that hypomethylation occurs later in course of metastatic PCa [13,14]. In contrast, through the use of sequencing-based DNA

methylation analysis, the results from this study and that of Kim et al. did not find decreased CpG methylation levels of metastatic PCa [10]. However, our approach is biased toward regions of high guanine-cytosine (GC) content, and LINE-1 and other repetitive regions shown to be hypomethylated during PCa progression are underrepresented by ERRBS; it stands to reason that a significant amount of hypomethylation may occur later at low or intermediate GC content regions, which are not adequately represented by ERRBS or M-NGS.

Gene fusions between members of the *ETS* family and the androgen-regulated *TMPRSS2* are frequently detected in PCa [47]. Differences in global methylation patterns or LINE-1 methylation were reported when categorizing PCa according to ETS gene fusion status [10,11]. Consistent with the study of Kim et al., we observed lower global methylation in *TMPRSS2:ERG* fusion-negative PCa (Figure W4), in contrast to the study of Borno et al., which showed increased levels of global methylation in the fusion-negative PCa. This discrepancy may be explained by the different approaches that were employed in these studies, as the levels of GC content coverage varied considerably: both ERRBS and M-NGS focus on regions of high GC content with minimal coverage of repeat elements, while MeDIP-seq has a broader coverage over CpGs in repetitive elements [48].

From a clinical point of view, the robust observation that DNA methylation increases with disease severity, especially at CGIs, provides new opportunities for detecting and diagnosing advanced PCa earlier than currently possible and for assisting treatment decisions. As a first step toward this goal, we provide a panel of 13 CGIs whose methylation reliably and reproducibly increases with disease severity, including association with a late stage and lethal neuroendocrine phenotype. Classification models based on the DNA methylation levels of these 13 CGIs and using Random Forest analysis deliver great discriminatory power, pointing toward a potential clinical use of the panel.

Larger studies with long-term follow-up, and inclusion of various stages of disease progression including CRPCs with retained AR expression, will be required to assess the true clinical value of these 13 CGIs and of other epigenetic alterations in the diagnosis, prognosis, and treatment of PCa.

Acknowledgments

We acknowledge the important contribution of the WCMC Epigenomics Core Facility and the WCMC Computational Genomics Core Facility and the technical assistance provided by Naoki Kitabayashi, Theresa McDonald, and Yifang Liu. We appreciate Ari M. Melnick and Daniel Di Bartolo for helpful suggestions.

References

[1] Bray F, Jemal A, Grey N, Ferlay J, and Forman D (2012). Global cancer transitions according to the Human Development Index (2008–2030): a population-based study. *Lancet Oncol* **13**, 790–801.

[2] Siegel R, Naishadham D, and Jemal A (2012). Cancer statistics, 2012. *CA Cancer J Clin* **62**, 10–29.

[3] Prensner JR, Rubin MA, Wei JT, and Chinnaiyan AM (2012). Beyond PSA: the next generation of prostate cancer biomarkers. *Sci Transl Med* **4**, 127rv123.

[4] Wilt TJ, Brawer MK, Jones KM, Barry MJ, Aronson WJ, Fox S, Gingrich JR, Wei JT, Gilhooly P, Grob BM, et al. (2012). Radical prostatectomy versus observation for localized prostate cancer. *N Engl J Med* **367**, 203–213.

[5] Bernstein BE, Meissner A, and Lander ES (2007). The mammalian epigenome. *Cell* **128**, 669–681.

[6] Nelson WG, De Marzo AM, and Yegnasubramanian S (2009). Epigenetic alterations in human prostate cancers. *Endocrinology* **150**, 3991–4002.

[7] Kobayashi Y, Absher DM, Gulzar ZG, Young SR, McKenney JK, Peehl DM, Brooks JD, Myers RM, and Sherlock G (2011). DNA methylation profiling reveals novel biomarkers and important roles for DNA methyltransferases in prostate cancer. *Genome Res* **21**, 1017–1027.

[8] Mahapatra S, Klee EW, Young CY, Sun Z, Jimenez RE, Klee GG, Tindall DJ, and Donkena KV (2012). Global methylation profiling for risk prediction of prostate cancer. *Clin Cancer Res* **18**, 2882–2895.

[9] Kron K, Pethe V, Briollais L, Sadikovic B, Ozcelik H, Sunderji A, Venkateswaran V, Pinthus J, Fleshner N, van der Kwast T, et al. (2009). Discovery of novel hypermethylated genes in prostate cancer using genomic CpG island microarrays. *PLoS One* **4**, e4830.

[10] Kim JH, Dhanasekaran SM, Prensner JR, Cao X, Robinson D, Kalyana-Sundaram S, Huang C, Shankar S, Jing X, Iyer M, et al. (2011). Deep sequencing reveals distinct patterns of DNA methylation in prostate cancer. *Genome Res* **21**, 1028–1041.

[11] Borno ST, Fischer A, Kerick M, Falth M, Laible M, Brase JC, Kuner R, Dahl A, Grimm C, Sayanjali B, et al. (2012). Genome-wide DNA methylation events in TMPRSS2:ERG fusion negative prostate cancers implicate an EZH2 dependent mechanism with miRNA-26a hypermethylation. *Cancer Discov* **2**(11), 1024–1035.

[12] Yegnasubramanian S, Kowalski J, Gonzalgo ML, Zahurak M, Piantadosi S, Walsh PC, Bova GS, De Marzo AM, Isaacs WB, and Nelson WG (2004). Hypermethylation of CpG islands in primary and metastatic human prostate cancer. *Cancer Res* **64**, 1975–1986.

[13] Bedford MT and van Helden PD (1987). Hypomethylation of DNA in pathological conditions of the human prostate. *Cancer Res* **47**, 5274–5276.

[14] Yegnasubramanian S, Haffner MC, Zhang Y, Gurel B, Cornish TC, Wu Z, Irizarry RA, Morgan J, Hicks J, DeWeese TL, et al. (2008). DNA hypomethylation arises later in prostate cancer progression than CpG island hypermethylation and contributes to metastatic tumor heterogeneity. *Cancer Res* **68**, 8954–8967.

[15] Friedlander TW, Roy R, Tomlins SA, Ngo VT, Kobayashi Y, Azameera A, Rubin MA, Pienta KJ, Chinnaiyan A, Ittmann MM, et al. (2012). Common structural and epigenetic changes in the genome of castration-resistant prostate cancer. *Cancer Res* **72**, 616–625.

[16] Beltran H, Rickman DS, Park K, Chae SS, Sboner A, MacDonald TY, Wang Y, Sheikh KL, Terry S, Tagawa ST, et al. (2011). Molecular characterization of neuroendocrine prostate cancer and identification of new drug targets. *Cancer Discov* **1**, 487–495.

[17] Berger MF, Lawrence MS, Demichelis F, Drier Y, Cibulskis K, Sivachenko AY, Sboner A, Esgueva R, Pflueger D, Sougnez C, et al. (2011). The genomic complexity of primary human prostate cancer. *Nature* **470**, 214–220.

[18] Akalin A, Garrett-Bakelman FE, Kormaksson M, Busuttil J, Zhang L, Khrebtukova I, Milne TA, Huang Y, Biswas D, Hess JL, et al. (2012). Base-pair resolution DNA methylation sequencing reveals profoundly divergent epigenetic landscapes in acute myeloid leukemia. *PLoS Genet* **8**, e1002781.

[19] Reik W and Walter J (2001). Genomic imprinting: parental influence on the genome. *Nat Rev Genet* **2**, 21–32.

[20] Kerkel K, Spadola A, Yuan E, Kosek J, Jiang L, Hod E, Li K, Murty VV, Schupf N, Vilain E, et al. (2008). Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation. *Nat Genet* **40**, 904–908.

[21] Ehrich M, Nelson MR, Stanssens P, Zabeau M, Liloglou T, Xinarianos G, Cantor CR, Field JK, and van den Boom D (2005). Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. *Proc Natl Acad Sci USA* **102**, 15785–15790.

[22] Krueger F and Andrews SR (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572.

[23] Giannopoulou EG and Elemento O (2011). An integrated ChIP-seq analysis platform with customizable workflows. *BMC Bioinformatics* **12**, 277.

[24] Elemento O, Slonim N, and Tavazoie S (2007). A universal framework for regulatory element discovery across all genomes and data types. *Mol Cell* **28**, 337–350.

[25] Goodarzi H, Elemento O, and Tavazoie S (2009). Revealing global regulatory perturbations across human cancers. *Mol Cell* **36**, 900–911.

[26] Gertz J, Varley KE, Reddy TE, Bowling KM, Pauli F, Parker SL, Kucera KS, Willard HF, and Myers RM (2011). Analysis of DNA methylation in a three-generation family reveals widespread genetic influence on epigenetic regulation. *PLoS Genet* **7**, e1002228.

[27] Breiman L (2001). Random forests. *Mach Learn* **45**, 5–32.

[28] Ku M, Koche RP, Rheinbay E, Mendenhall EM, Endoh M, Mikkelsen TS, Presser A, Nusbaum C, Xie X, Chi AS, et al. (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet* **4**, e1000242.

[29] Gal-Yam EN, Egger G, Iniguez L, Holster H, Einarsson S, Zhang X, Lin JC, Liang G, Jones PA, and Tanay A (2008). Frequent switching of Polycomb repressive marks and DNA hypermethylation in the PC3 prostate cancer cell line. *Proc Natl Acad Sci USA* **105**, 12979–12984.

[30] Statham AL, Robinson MD, Song JZ, Coolen MW, Stirzaker C, and Clark SJ (2012). Bisulfite sequencing of chromatin immunoprecipitated DNA (BisChIP-seq) directly informs methylation status of histone-modified DNA. *Genome Res* **22**, 1120–1127.

[31] Schlesinger Y, Straussman R, Keshet I, Farkash S, Hecht M, Zimmerman J, Eden E, Yakhini Z, Ben-Shushan E, Reubinoff BE, et al. (2007). Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for *de novo* methylation in cancer. *Nat Genet* **39**, 232–236.

[32] Velichutina I, Shaknovich R, Geng H, Johnson NA, Gascoyne RD, Melnick AM, and Elemento O (2010). EZH2-mediated epigenetic silencing in germinal center B cells contributes to proliferation and lymphomagenesis. *Blood* **116**, 5247–5255.

[33] Kubosaki A, Tomaru Y, Tagami M, Arner E, Miura H, Suzuki T, Suzuki M, Suzuki H, and Hayashizaki Y (2009). Genome-wide investigation of *in vivo* EGR-1 binding sites in monocytic differentiation. *Genome Biol* **10**, R41.

[34] Watari A, Takaki K, Higashiyama S, Li Y, Satomi Y, Takao T, Tanemura A, Yamaguchi Y, Katayama I, Shimakage M, et al. (2006). Suppression of tumorigenicity, but not anchorage independence, of human cancer cells by new candidate tumor suppressor gene CapG. *Oncogene* **25**, 7373–7380.

[35] Van den Abbeele A, De Corte V, Van Impe K, Bruyneel E, Boucherie C, Bracke M, Vandekerckhove J, and Gettemans J (2007). Downregulation of gelsolin family proteins counteracts cancer cell invasion *in vitro*. *Cancer Lett* **255**, 57–70.

[36] Wittamer V, Franssen JD, Vulcano M, Mirjolet JF, Le Poul E, Migeotte I, Brezillon S, Tyldesley R, Blanpain C, Detheux M, et al. (2003). Specific recruitment of antigen-presenting cells by chemerin, a novel processed ligand from human inflammatory fluids. *J Exp Med* **198**, 977–985.

[37] Li DQ, Wang L, Fei F, Hou YF, Luo JM, Zeng R, Wu J, Lu JS, Di GH, Ou ZL, et al. (2006). Identification of breast cancer metastasis-associated proteins in an isogenic tumor metastasis model using two-dimensional gel electrophoresis and liquid chromatography-ion trap-mass spectrometry. *Proteomics* **6**, 3352–3368.

[38] Lomnytska MI, Becker S, Bodin I, Olsson A, Hellman K, Hellstrom AC, Mints M, Hellman U, Auer G, and Andersson S (2011). Differential expression of ANXA6, HSP27, PRDX2, NCF2, and TPM4 during uterine cervix carcinogenesis: diagnostic and prognostic value. *Br J Cancer* **104**, 110–119.

[39] Wang L, Darling J, Zhang JS, Liu W, Qian J, Bostwick D, Hartmann L, Jenkins R, Bardenhauer W, Schutte J, et al. (2000). Loss of expression of the *DRR 1* gene at chromosomal segment 3p21.1 in renal cell carcinoma. *Genes Chromosomes Cancer* **27**, 1–10.

[40] Burton ER, Gaffar A, Lee SJ, Adeshuko F, Whitney KD, Chung JY, Hewitt SM, Huang GS, Goldberg GL, Libutti SK, et al. (2011). Downregulation of Filamin A interacting protein 1-like is associated with promoter methylation and induces an invasive phenotype in ovarian cancer. *Mol Cancer Res* **9**, 1126–1138.

[41] Suzuki MM and Bird A (2008). DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet* **9**, 465–476.

[42] Ball MP, Li JB, Gao Y, Lee JH, LeProust EM, Park IH, Xie B, Daley GQ, and Church GM (2009). Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat Biotechnol* **27**, 361–368.

[43] Caterino M, Ruoppolo M, Fulcoli G, Huynth T, Orru S, Baldini A, and Salvatore F (2009). Transcription factor TBX1 overexpression induces downregulation of proteins involved in retinoic acid metabolism: a comparative proteomic analysis. *J Proteome Res* **8**, 1515–1526.

[44] Ryckebusch L, Bertrand N, Mesbah K, Bajolle F, Niederreither K, Kelly RG, and Zaffran S (2010). Decreased levels of embryonic retinoic acid synthesis accelerate recovery from arterial growth delay in a mouse model of DiGeorge syndrome. *Circ Res* **106**, 686–694.

[45] Touma SE, Perner S, Rubin MA, Nanus DM, and Gudas LJ (2009). Retinoid metabolism and ALDH1A2 (RALDH2) expression are altered in the transgenic adenocarcinoma mouse prostate model. *Biochem Pharmacol* **78**, 1127–1138.

[46] Kim H, Lapointe J, Kaygusuz G, Ong DE, Li C, van de Rijn M, Brooks JD, and Pollack JR (2005). The retinoic acid synthesis gene ALDH1a2 is a candidate tumor suppressor in prostate cancer. *Cancer Res* **65**, 8118–8124.

[47] Rubin MA, Maher CA, and Chinnaiyan AM (2011). Common gene rearrangements in prostate cancer. *J Clin Oncol* **29**, 3659–3668.

[48] Harris RA, Wang T, Coarfa C, Nagarajan RP, Hong C, Downey SL, Johnson BE, Fouse SD, Delaney A, Zhao Y, et al. (2010). Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. *Nat Biotechnol* **28**, 1097–1105.

Table W1. Clinical Parameters of Samples Used in ERRBS.

PCa	STID No.	Serum PSA (ng/ml)	Age at Diagnosis	Gleason Score (H&E)	Pathological Stage	EST-Fusion Status	Year of Radical Prostatectomy
1	508	7.8	57	3 + 4 = 7	T2c	NEG	2007
2	581	9.2	69	4 + 5 = 9	T3b	POS	2007
3	1701	2.1	62	3 + 4 = 7	T3a	POS	2007
4	1783	9.8	66	4 + 4 = 8	T2c	NEG	2008
5	2832	6.6	66	3 + 4 = 7	T2c	POS	2007
6	3027	10.2	66	4 + 4 = 8	T3b	NEG	2008
7	3043	7.2	69	3 + 4 = 7	T2c	NEG	2008

CRPC	STID No.	Serum PSA (ng/ml)	Age at Diagnosis	Metastatic	AR Status	Overall Survival	EST-Fusion Status	Tissue Type
1	4240	1.93	74	–	(–)	13 months	NEG	NEPC
2	7520	0.06	56	Yes	(–)	Alive	POS	NEPC
3	7820	–	–	–	(–)	9 months	NEG	NEPC
4	7821	–	–	–	(–)	9 months	NEG	NEPC
5	8220	–	–	–	(–)	–	NEG	NEPC
6	8740	48.5	59	Yes	(–)	3 months	NEG	NEPC

Table W2. Clinical Parameters of FFPE Samples of PCa and CRPC Used in the Validation Cohort.

PCa	STID No.	Serum PSA (ng/ml)	Age at Diagnosis	Gleason Score (H&E)	Pathological Stage	Benign in the Cohort	Year of Radical Prostatectomy
FF23	S10-25464	15	74	4 + 3 = 7	T2c	FF22	2010
FF25	S10-17814	5.94	70	3 + 4 = 7	T2a	FF24	2010
FF29	S10-29475	7.4	69	4 + 5 = 9	T3a	FF28	2010
FF31	S09-21599	5.6	62	3 + 4 = 7	T2c	FF30	2009
FF33	S08-6573	–	65	3 + 4 = 7	T2c	FF32	2008
FF35	S08-25678	4.98	68	4 + 3 = 7	T2c	FF34	2008
FF37	S08-15334	–	58	4 + 3 = 7	T3a	FF36	2008
FF39	S07-30717	8.8	63	3 + 4 = 7	T3b	FF38	2007
FF41	S07-29284	5.4	61	3 + 4 = 7	T2c	n/a	2007
FF43	S07-30675	–	50	3 + 4 = 7	T2c	FF42	2007
FF45	S07-28150	–	58	3 + 4 = 7	T2c	FF44	2007
FF47	S07-28191	–	57	3 + 4 = 7	T2c	FF46	2007
FF49	S07-30522	–	63	3 + 4 = 7	T2c	FF48	2007
FF51	S07-31303	–	60	3 + 4 = 7	T2c	FF50	2007
FF53	S07-29250	–	55	3 + 3 = 6	T2a	FF52	2007
FF63	S06-1447	3.15	55	5 + 4 = 9	T4	n/a	2006

CRPC	STID No.	Tissue Type	Serum PSA (ng/ml)	Age at Diagnosis	Metastatic	Prior Chemotherapy	AR Status	Benign in the Cohort
FF12	1633/04	NEPC	–	–	–	–	(+)	n/a
FF14	10586	NEPC	–	–	–	–	(–)	FF13
FF16	21771/07	NEPC	–	–	–	–	(+)	FF15
FF18	S05-31294	NEPC	5.9	63	–	No	(+)	FF17
FF20	S10-5780	NEPC	–	69	yes	Yes	(–)	FF19
FF55	S97-1161	NEPC	–	78	–	No	(+)	n/a
FF57	S02-1928	NEPC	–	79	yes	No	(+)	n/a
FF59	S12-2609	NEPC	0.7	65	yes	No	(–)	FF58

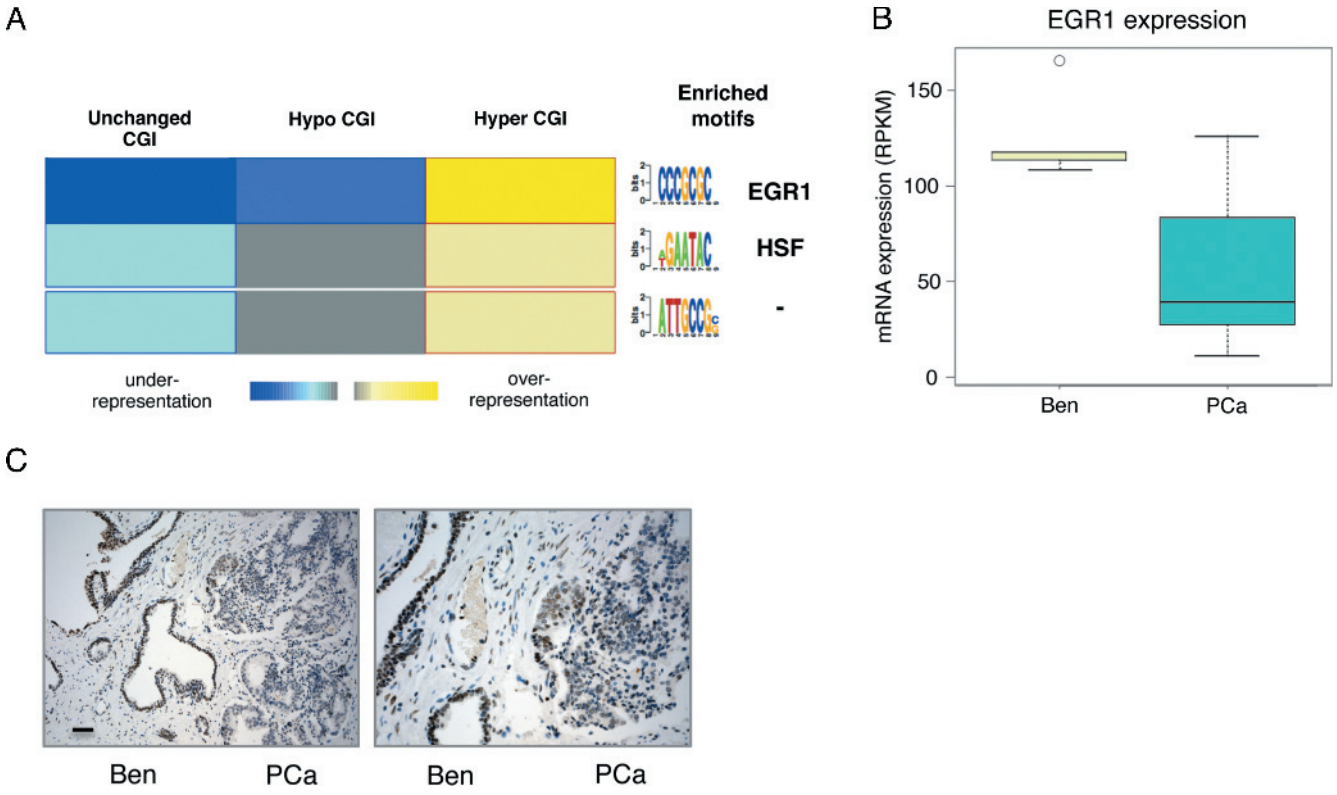


Figure W1. Possible involvement of EGR1 in DNA hypermethylation. (A) Motif analysis of hypermethylated and hypomethylated CGIs. (B) EGR1 mRNA expression between benign prostate tissues and PCa. (C) Immunohistochemistry staining of EGR1 in patient 4 is shown here. Left: ×20; scale bar, 100 μm; right: ×40.

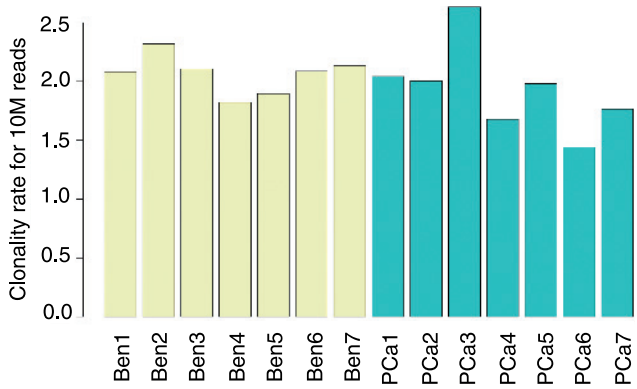


Figure W2. PCR duplicates and clonal reads do not explain differences in ASM. Clonality rate for 10 million reads in benign prostate tissue and PCa.

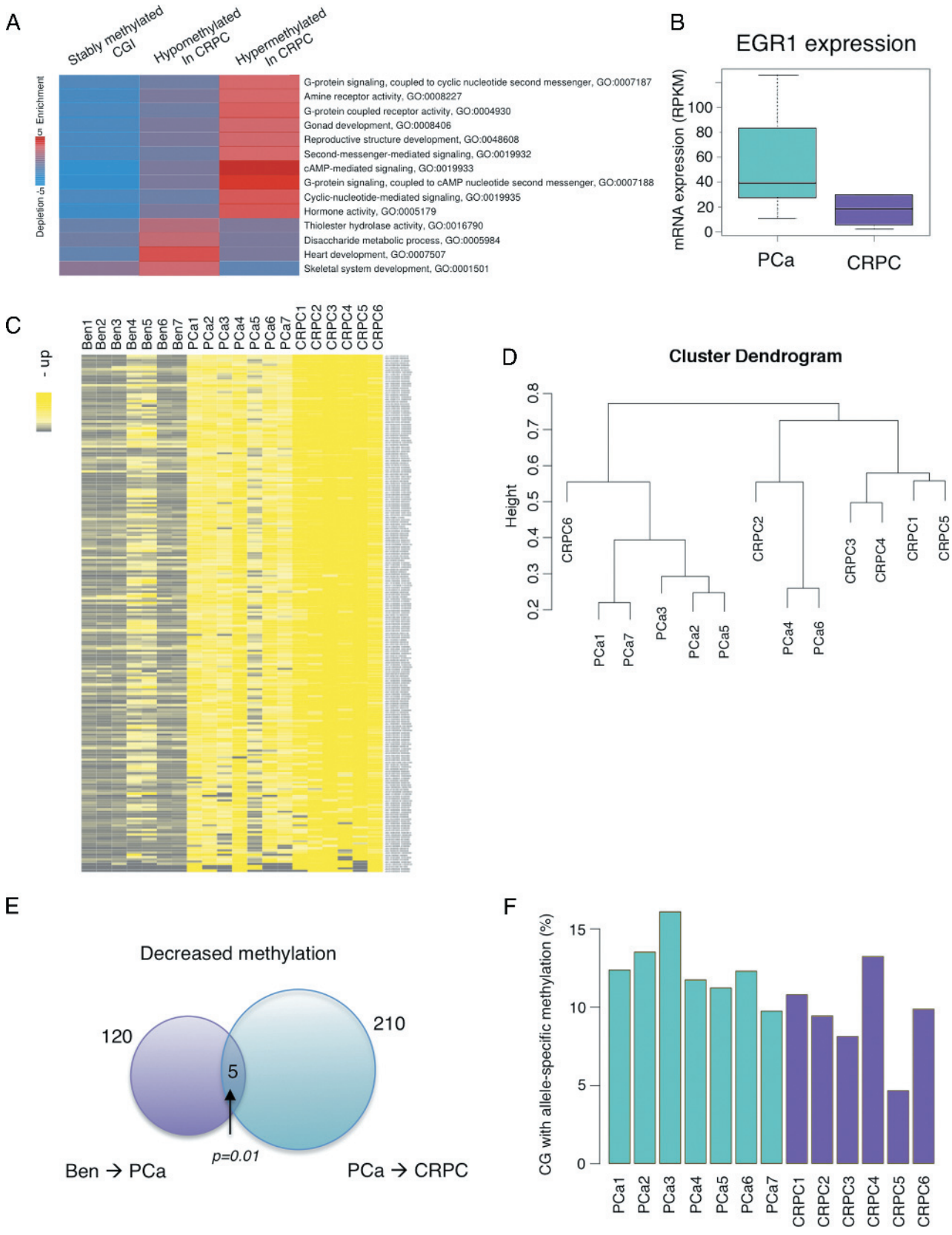


Figure W3. Pathway analysis and comparison between PCa and CRPC. (A) The enriched pathways associated with DNA methylation status in CRPC; red, enrichment; blue, depletion. (B) EGR1 mRNA expression between PCa and CRPC. (C) Heat map of the 228 CGIs. (D) Cluster dendrogram of PCa and CRPC using the 228 CGIs. (E) Venn diagram displays the methylation overlap observed between two sections: decreased methylation from benign prostate tissues to PCa and decreased methylation from PCa to CRPC; the inset number represents the coverage in each section. Five CGIs appear in both sections with $P = .01$. (F) Percentage of CpG with ASM in PCa and CRPC. Only CpG with nearby SNPs are used for this analysis.

Table W3. The Panel of the 13 CGIs and the Associated Genes.

No.	Chr	Start	End	Gene	Description	DIST	EXP	Benign	PCa	CRPC
1	chr12	75601081	75601752	*KCNC2*	Potassium voltage-gated channel subfamily C	2095	Up	4.644	6.035	11.677
2	chr16	67427284	67428950	*ZDHHC1*	Probable palmitoyltransferase ZDHHC1	204	Down	8.21	7.266	5.733
3	chr22	19746924	19747141	*TBX1*	T-box transcription factor TBX1 isoform C	2808	Up	2.078	3.865	9.733
4	chr2	85640969	85641259	*CAPG*	Macrophage-capping protein	83	Down	22.958	11.654	8.677
5	chr7	150037459	150039031	*RARRES2*	Retinoic acid receptor responder protein 2	518	Down	77.08	48.548	15.319
6	chr12	52400467	52401696	*GRASP*	General receptor for phosphoinositides	335	Down	9.953	4.452	2.018
7	chr11	64815040	64815722	*SAC3D1*	SAC3 domain-containing protein 1	3081	Up	4.907	6.564	19.878
8	chr19	16186789	16188275	*TPM4*	Tropomyosin α-4 chain isoform 2	398	Down	42.185	23.212	14.138
9	chr11	67350928	67351953	*GSTP1*	Glutathione *S*-transferase P	376	Down	168.097	62.937	42.57
10	chr14	36991594	36992488	*NKX2-1*	Homeobox protein Nkx-2.1 isoform 2	2611	Up	0.014	1.817	66.708
11	chr3	58572478	58572903	*FAM107A*	Downregulated in renal cell carcinoma	9200	Down	18.921	3.797	1.016
12	chr20	45279781	45280169	*SLC13A3*	Solute carrier family 13 member 3 isoform a	125	Down	6.766	6.334	3.648
13	chr3	99594969	99595215	*FILIP1L*	Filamin A interacting protein 1–like isoform 1	46	Down	22.005	5.101	1.525

Table W4. MassARRAY EpiTYPER Primers of the 13 CGIs.

No.	Chr	Start	End	Gene	DIST	Forward	Reverse
Epi1	chr12	75601081	75601752	*KCNC2*	2095	aggaagagagGTAGTATTTTTAAGATTTTGTTTGGAAT	cagtaatacgactcactatagggagaaggctATACCCCAAAAAACCAACTCCT
Epi2	chr16	67427284	67428950	*ZDHHC1*	204	aggaagagagGGTTTTTTGAGGAAATAGTTTTTTA	cagtaatacgactcactatagggagaaggctTACTCCAAACTCTAAATCCTACCCA
Epi3	chr22	19746924	19747141	*TBX1*	2808	aggaagagagAAGAGGGTTTTGTATTTTTAGGGTG	cagtaatacgactcactatagggagaaggctCTAAAAACTCACCTTCCATATCCCT
Epi4	chr2	85640969	85641259	*CAPG*	83	aggaagagagGGTTGTTATTAGTTTTAAGTGGGGG	cagtaatacgactcactatagggagaaggctACCTAAAAACTCACCTTCCATATCC
Epi5	chr7	150037459	150039031	*RARRES2*	518	aggaagagagGGAGATTTAGGGAGAGATATAATGGG	cagtaatacgactcactatagggagaaggctATAAAATTCCTCCAAAACCACCTAC
Epi6	chr12	52400467	52401696	*GRASP*	335	aggaagagagAGGGATAGAGATAGTTTTAGGTAAGTTGA	cagtaatacgactcactatagggagaaggctAAAAATCCAAAAAACAATAACCCTC
Epi7	chr11	64815040	64815722	*SAC3D1*	3081	aggaagagagTTTTTTTATTTTTTTGGTTGTAGAGAAG	cagtaatacgactcactatagggagaaggctAAAACCCAAATAATCTTTCTCCCCT
Epi8	chr19	16186789	16188275	*TPM4*	398	aggaagagagGAGGGAAAGATGTGAAAATTTTATT	cagtaatacgactcactatagggagaaggctCCACAACTACTAAAAAATACCCCTT
Epi9	chr11	67350928	67351953	*GSTP1*	376	aggaagagagTTTTGTTGTTTGTTTATTTTTTAGG	cagtaatacgactcactatagggagaaggctTACTAAAAACTCTAAACCCCATCCC
Epi10	chr14	36991594	36992488	*NKX2-1*	2611	aggaagagagGGGAAGAGAAGGATATTTGTATTTTT	cagtaatacgactcactatagggagaaggctTTAAACCCTACCCTACCCTAACC
Epi11	chr3	58572478	58572903	*FAM107A*	9200	aggaagagagGTTTTTTATTGTAGAGTTAGTATTGTTGG	cagtaatacgactcactatagggagaaggctACACCTACCCTAACCCCTACCC
Epi12	chr20	45279781	45280169	*SLC13A3*	125	aggaagagagGATTTAGAAGGTAGGGTGGGATTTA	cagtaatacgactcactatagggagaaggctCTTTTCCTCCAAAAAAAACAAAAAA
Epi13	chr3	99594969	99595215	*FILIP1L*	46	aggaagagagGATTTGTTATTTGGTTGTTTGATAG	cagtaatacgactcactatagggagaaggctTACCTCCCAATAACTTTATTAACCC

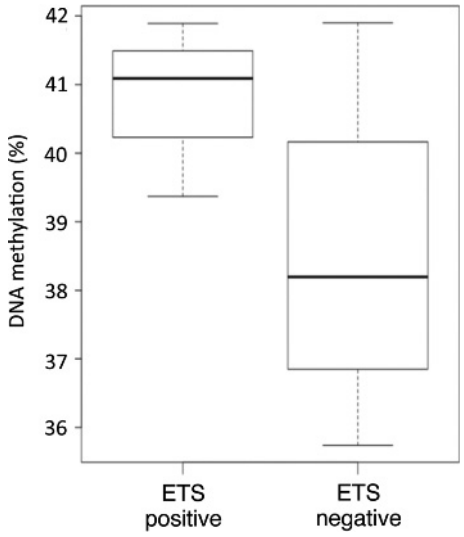


Figure W4. DNA methylation levels in *TMPRSS2:ERG* fusion-positive and fusion-negative samples.