

4-23-2018

Genomic DNA Methylation-Derived Algorithm Enables Accurate Detection of Malignant Prostate Tissues

Erfan Aref-Eshghi
Western University

Laila C Schenkel
Western University

Peter Ainsworth
Western University

Hanxin Lin
Western University

David I Rodenhiser
Western University

See next page for additional authors

Follow this and additional works at: <https://ir.lib.uwo.ca/paedpub>



Part of the [Pediatrics Commons](#)

Citation of this paper:

Aref-Eshghi, Erfan; Schenkel, Laila C; Ainsworth, Peter; Lin, Hanxin; Rodenhiser, David I; Cutz, Jean-Claude; and Sadikovic, Bekim, "Genomic DNA Methylation-Derived Algorithm Enables Accurate Detection of Malignant Prostate Tissues" (2018). *Paediatrics Publications*. 309.
<https://ir.lib.uwo.ca/paedpub/309>

Authors

Erfan Aref-Eshghi, Laila C Schenkel, Peter Ainsworth, Hanxin Lin, David I Rodenhiser, Jean-Claude Cutz, and Bekim Sadikovic



Genomic DNA Methylation-Derived Algorithm Enables Accurate Detection of Malignant Prostate Tissues

Erfan Aref-Eshghi^{1,2}, Laila C. Schenkel^{1,2}, Peter Ainsworth^{1,2}, Hanxin Lin^{1,2}, David I. Rodenhiser^{3,4,5}, Jean-Claude Cutz⁶ and Bekim Sadikovic^{1,2*}

¹ Department of Pathology and Laboratory Medicine, Western University, London, ON, Canada, ² Molecular Genetics Laboratory, Molecular Diagnostics Division, London Health Sciences, London, ON, Canada, ³ Department of Pediatrics, Western University and Children's Health Research Institute, London, ON, Canada, ⁴ Department of Biochemistry, Western University and Children's Health Research Institute, London, ON, Canada, ⁵ Department of Oncology, Western University and Children's Health Research Institute, London, ON, Canada, ⁶ Department of Pathology and Laboratory Medicine, McMaster University, Hamilton, ON, Canada

OPEN ACCESS

Edited by:

Fabio Grizzi,
Humanitas Research Hospital, Italy

Reviewed by:

Hung-Ming Lam,
University of Washington,
United States
Felix K. H. Chun,
Universitätsklinikum Hamburg-
Eppendorf, Germany

*Correspondence:

Bekim Sadikovic
bekim.sadikovic@lhsc.on.ca

Specialty section:

This article was submitted to
Genitourinary Oncology,
a section of the journal
Frontiers in Oncology

Received: 17 January 2018

Accepted: 21 March 2018

Published: 23 April 2018

Citation:

Aref-Eshghi E, Schenkel LC,
Ainsworth P, Lin H, Rodenhiser DI,
Cutz J-C and Sadikovic B (2018)
Genomic DNA Methylation-Derived
Algorithm Enables Accurate
Detection of Malignant Prostate
Tissues.
Front. Oncol. 8:100.
doi: 10.3389/fonc.2018.00100

Introduction: The current methodology involving diagnosis of prostate cancer (PCa) relies on the pathology examination of prostate needle biopsies, a method with high false negative rates partly due to temporospatial, molecular, and morphological heterogeneity of prostate adenocarcinoma. It is postulated that molecular markers have a potential to assign diagnosis to a considerable portion of undetected prostate tumors. This study examines the genome-wide DNA methylation changes in PCa in search of genomic markers for the development of a diagnostic algorithm for PCa screening.

Methods: Archival PCa and normal tissues were assessed using genomic DNA methylation arrays. Differentially methylated sites and regions (DMRs) were used for functional assessment, gene-set enrichment and protein interaction analyses, and examination of transcription factor-binding patterns. Raw signal intensity data were used for identification of recurrent copy number variations (CNVs). Non-redundant fully differentiating cytosine-phosphate-guanine sites (CpGs), which did not overlap CNV segments, were used in an L1 regularized logistic regression model (LASSO) to train a classification algorithm. Validation of this algorithm was performed using a large external cohort of benign and tumor prostate arrays.

Results: Approximately 6,000 probes and 600 genomic regions showed significant DNA methylation changes, primarily involving hypermethylation. Gene-set enrichment and protein interaction analyses found an overrepresentation of genes related to cell communications, neurogenesis, and proliferation. Motif enrichment analysis demonstrated enrichment of tumor suppressor-binding sites nearby DMRs. Several of these regions were also found to contain copy number amplifications. Using four non-redundant fully differentiating CpGs, we trained a classification model with 100% accuracy in discriminating tumors from benign samples. Validation of this algorithm using an external cohort of 234 tumors and 92 benign samples yielded 96% sensitivity and 98% specificity. The model was found to be highly sensitive to detect metastatic lesions in bone, lymph node, and soft tissue, while being specific enough to differentiate the benign hyperplasia of prostate from tumor.

Conclusion: A considerable component of PCa DNA methylation profile represent driver events potentially established/maintained by disruption of tumor suppressor activity. As few as four CpGs from this profile can be used for screening of PCa.

Keywords: prostate cancer, DNA methylation, protein interaction, transcription factor binding, copy number variation, differentially methylated regions, machine learning, LASSO

INTRODUCTION

Prostate cancer (PCa) is the most common malignancy in men, representing the third greatest cause of mortality in the general population after lung and colorectal cancer (1). Current screening protocols for PCa rely on digital rectal examination along with the evaluation of serum levels of prostate-specific antigen. If a tumor is suspected, trans-rectal, ultrasound-guided, prostate needle biopsies are conducted, followed by a histopathology examination. Major limitations of this approach include a relatively high false negative rate and the requirement for annual follow-ups and repeat sampling, an invasive approach with significant burden on patient health care and lifestyle (2). As such, a significant effort has been made to identify more accurate diagnostic and screening biomarkers for PCa, though to date this effort has been met with limited success in clinical application (3).

Among the markers that have gained interest in disease screening is DNA methylation, an epigenetic mechanism that includes modification of the fifth base of cytosine at cytosine-phosphate-guanine (CpG) dinucleotide in the DNA by addition or removal of a methyl group. DNA methylation plays an integral role in the regulation of gene expression and genomic stability, disruption of which can lead to neoplastic transformation, carcinogenesis, and cancer progression (4). The relative stability of the DNA molecule together with the genome-wide distribution of DNA methylation marks has made it an attractive target for disease biomarker discovery (5). DNA methylation has proven to possess a great potential in the prediction of many human traits and conditions, including age, postpartum depression, childhood stress exposure, tumor class, response to therapy, and cancer progression/metastasis (5).

Epigenetic changes have been previously demonstrated in multiple cancers including PCa (6), with epigenomic signatures of PCa being proposed to be used as both diagnostic and prognostic markers (3, 7). PCa-associated DNA methylation changes have also been found in the urine and serum of patients, raising hopes for non-invasive methods of PCa screening (3). These methylation changes have been found in various genes involved in hormonal response, cell-cycle regulation, cell invasion, and DNA damage repair (6). In particular, hypermethylation in the promoters of tumor suppressor genes *APC*, *RAR β* , and *GSTP1*, has been proposed to be used as a diagnostic marker for PCa (8). However, these assays have not generated optimal performances, partly because they target selected candidate genes instead of markers discovered in a genome-wide search. The number of studies attempting to produce a diagnostic methylation algorithm for PCa using a genome-wide approach is limited. Furthermore, few studies have conducted a comprehensive genome-wide DNA methylation analysis to reveal the role of DNA methylation in PCa (3).

In the present study, we have conducted a genome-wide DNA methylation analysis to describe the DNA methylation changes in PCa. We have investigated potential mechanisms by which differentially methylated genes may be involved in the development of PCa, by examining pathway enrichment and protein interactions. In addition, we have addressed mechanisms that regulate the establishment of differentially methylated regions (DMRs) in PCa by investigating the transcription factor (TF)-binding patterns nearby the identified segments. Also, we report the copy number variations (CNVs) recurrently happening in PCa and evaluate their associations with the methylation levels of the overlapping DMRs. Finally, using the most differentiating CpGs, we have generated a classification model to distinguish the tumors from the adjacent benign tissues, and have validated this algorithm using a large independent cohort of benign and tumor prostate samples.

MATERIALS AND METHODS

Specimen Collection

Archival formalin-fixed paraffin-embedded tissues were collected from the Department of Pathology at McMaster University, ON, Canada. The slides were reviewed by a pathologist to identify representative tumor tissue blocks. Enrichment of tumor cells was performed by tissue macro-dissection of predetermined areas, as outlined by Haematoxylin and Eosin staining. Seven-micron thick sections in isolated areas of interest with at least 50% tumor cellularity were dissected to be used as tumor samples. Normal prostate samples were obtained from the adjacent tissues demonstrating no tumor involvement.

Methylation Assay and Quality Assessment

Genomic DNA extraction was conducted using the Illumina FFPE DNA recovery kit. Following bisulfite conversion, DNA methylation analysis of the samples was performed using the Illumina Infinium methylation 450 k bead chip array (San Diego, CA, USA), according to the manufacturer's protocol. Both tumor and normal samples were assayed in one experiment to avoid batch effect. The resulting methylated and unmethylated signal intensity data were imported into R 3.4.2 for analysis. Normalization was performed using Illumina normalization method with background correction using the *minfi* package (9). Probes with detection *p*-value > 0.01, and those known to contain single nucleotide polymorphisms at the CpG interrogation or single nucleotide extension were removed. All of the samples were examined for genome-wide methylation density, and those deviating from a bimodal distribution were excluded. Factor analysis using multiple dimensional scaling did not reveal any

unexplained variation or outliers among the samples. The raw and processed genome-wide methylation microarray data has been deposited to gene expression omnibus (GEO) (accession ID: GSE112047).

Identification of a PCa-Related Methylation Profile

DNA methylation analysis was performed using a modification of our previously published protocol (10–15). Methylation level for each probe was measured as a beta value, calculated from the ratio of the methylated signals vs. the total sum of unmethylated and methylated signals, ranging between 0 (no methylation) and 1 (full methylation). This value was used for biological interpretation and visualization. For statistical analysis, beta values were logit transformed to *M*-values using the following equation: $\log_2(\text{beta}/1 - \text{beta})$. A linear regression modeling using the *limma* package (16) was used to identify the differentially methylated probes by testing the association of every CpG site with the level of tumor involvement (0–90%). The generated *p*-values were moderated using the *eBayes* function in the *limma* package and were corrected for multiple testing using *Bonferroni* method. Probes with a corrected *p*-value < 0.01 and a methylation difference > 20% were considered significant. The effect size cut-off (20%) was determined following the examination of the volcano plot of the analysis, as conducted in our previous study (17). The identified probes were examined using an unsupervised hierarchical clustering to ensure their ability in separating the tumors from normal samples.

Gene Enrichment Analysis and Identification of Differential Methylation Interaction Hotspots

To identify the gene ontology (GO) terms overrepresented in the genes harboring differentially methylated probes, a gene-set enrichment analysis was performed using the *missMethyl* package (18), taking into account the number of CpG sites per gene. GO terms with a *Bonferroni* corrected *p*-value < 0.01 were considered significant. Only the biological processes were reported. The redundant GO terms were reduced, and following a multiple dimensional scaling, visualized using *REViGO* tool (19). The relationships across the significant GO terms were plotted in a hierarchical order using *GORILLA* tool (20).

We used *EpiMod* algorithm (21) to search for the “interactome hotspots” of differential promoter methylation. In this algorithm, protein expression changes are inferred according to a model of inverse association between the promoter methylation and gene expression. Among the differentially expressed genes in an interactive network, a hotspot [or epigenetic module (EpiMods)] is a sub-network with an exceptionally large average edge-weight density (combined methylation statistics of the neighboring genes) as compared to the rest of the network (21). To assign a statistical significance to the identified hotspots, 1,000 Monte Carlo randomization of the molecular profiles were conducted as suggested by the algorithm. Interactive network hotspots composed of at least ten genes and FDR < 0.01 were reported.

Identification of DMRs in PCa

To identify genomic regions harboring methylation changes (DMRs), a bump-hunting approach was used by the *bumphunter* package (22). The analysis considered regions with >20% change in the overall methylation between tumor and normal samples with gaps no more than 500 bp among neighboring CpGs. As suggested by the package, 1,000 bootstrapping procedures were performed to compute family-wise error rate (FWER). We selected regions containing a minimum of three consecutive probes and FWER < 0.01. The identified regions were mapped to CpG islands and coding genes. *Gviz* package was used for visualization of the DMRs.

Identification of Copy Number Changes Using DNA Methylation Array

To estimate the copy number alterations in the prostate tumor from the Infinium methylation array, the raw methylated and unmethylated intensities from every sample were summed, and quantile normalized using the *preprocessCore* package (23). The normalized matrix values were divided by the median values of every probe across the normal samples. The divided ratios were then log₁₀ transformed, smoothed, and segmented using the *DNAcopy* Bioconductor package (24) to identify genomic regions in every sample showing a copy number change. A *p*-value of < 0.01 obtained from 10,000 permutations was used to define a change point during segmentation. The segmented regions with a minimum of five probes and an average log ratio > 0.2 or < -0.3 (corresponding to at least one copy amplification/deletion) were selected. Using the *GenomicRanges* package (25), the segments with a similar pattern of copy number change in both normal and tumor samples were excluded. Segments on chromosome six, overlapping HLA genes, were not considered due to normal variations in their copy numbers. From the remaining segments, those occurring in more than one tumor sample were reported.

TF-Binding Site Enrichment Analysis in DMRs

We investigated the enrichment of TF-binding motifs in both the identified DMRs and their immediate surroundings (± 5 kb) using the pipeline recommended by the *RTFBSDB* package (26). First, the entire direct and inferred Homo sapiens TF-binding motifs were downloaded from the *Cis-BP* database (27). Every DMR sequence and its ± 5 kb surroundings were scanned for motif enrichment relative to a background genome of 100,000 base pairs. As recommended by the package, the difference in the GC content of the DMRs and the background genome was accounted for by re-sampling of the background genome to avoid a potential bias due to GC content. The motifs with an enrichment ratio ≥ 1.5 folds and a *Bonferroni* corrected *p*-value < 0.01 were selected for assessment.

Construction and Validation of a DNA Methylation-Based Diagnostic Model for PCa

The differentially methylated probes were used to build a classification model to differentiate tumor samples from the normal

tissues. Only CpG probes located outside the regions with copy number change were utilized for feature selection. First, a receiver operating characteristic curve analysis (ROC) was performed to identify the most differentiating probes. Those probes with an area under the curve (AUC) of 1.00 were retained. Next, pairwise correlations among the remaining probes were measured to identify and exclude the redundant features with R -squared >0.90 . A least absolute shrinkage and selection operator (LASSO) was used to further narrow down the features and train the model, using the *glmnet* package (28). Following 1,000 permutations, the shrinkage parameter of lambda with lowest misclassification error was selected and incorporated into the final model. The model was set to generate probability scores ranging 0–1, representing the chance of a given sample being a tumor. The default binary classifier's probability threshold of 0.5 was used as classification cut-off. To assess the performance of the model, publically available methylation data from normal and tumor prostate tissues, benign prostate hyperplasia (BPH), and samples of PCa metastasis in other tissues were downloaded from GEO and blindly supplied to the model. Based on the scores and classifications produced by the algorithm for these samples, we measured accuracy, sensitivity, specificity, and AUC as indices of performance for our classifier. The GEO accession IDs and their related publications of the methylation data files used for validation of our algorithm are mentioned here to give credit to the researchers providing these data. Tumor and adjacent benign array files were downloaded from GSE76938 (29), GSE52955 (30), GSE47915 (31), GSE55479 (32), GSE83917 (32), and GSE38240 (33). The normal prostate array data from radical prostatectomy due to bladder cancer, and those with BPH were obtained from GSE55599 (34). Methylation data of tissues with metastases from prostate were included in the files from GSE38240 (33).

Ethics Statement

This study has been approved by the Hamilton Integrated Research Ethics Board (#14-700-T). All of the samples and records were de-identified prior to analysis.

RESULTS

Study Cohort

The study cohort included 31 samples of archival prostate gland tissue with a confirmed diagnosis of prostate adenocarcinoma, assessed by a licensed pathologist. From 16 of these, adjacent normal tissues were collected to be used as the control group. The tumor sections were dissected ensuring a minimum of 50% cancer involvement using macrodissection. There was no overall difference between the age of the subjects in the control group (55.7 ± 7.3) and the tumor group (55.2 ± 9.9).

PCa Generates a Hypermethylation Profile

From a total of 440,532 CpG probes that passed the quality controls, a total of 6,167 were revealed to have a methylation difference of $>20\%$ between the two groups with a multiple testing corrected p -value <0.01 (Table S1 in Supplementary

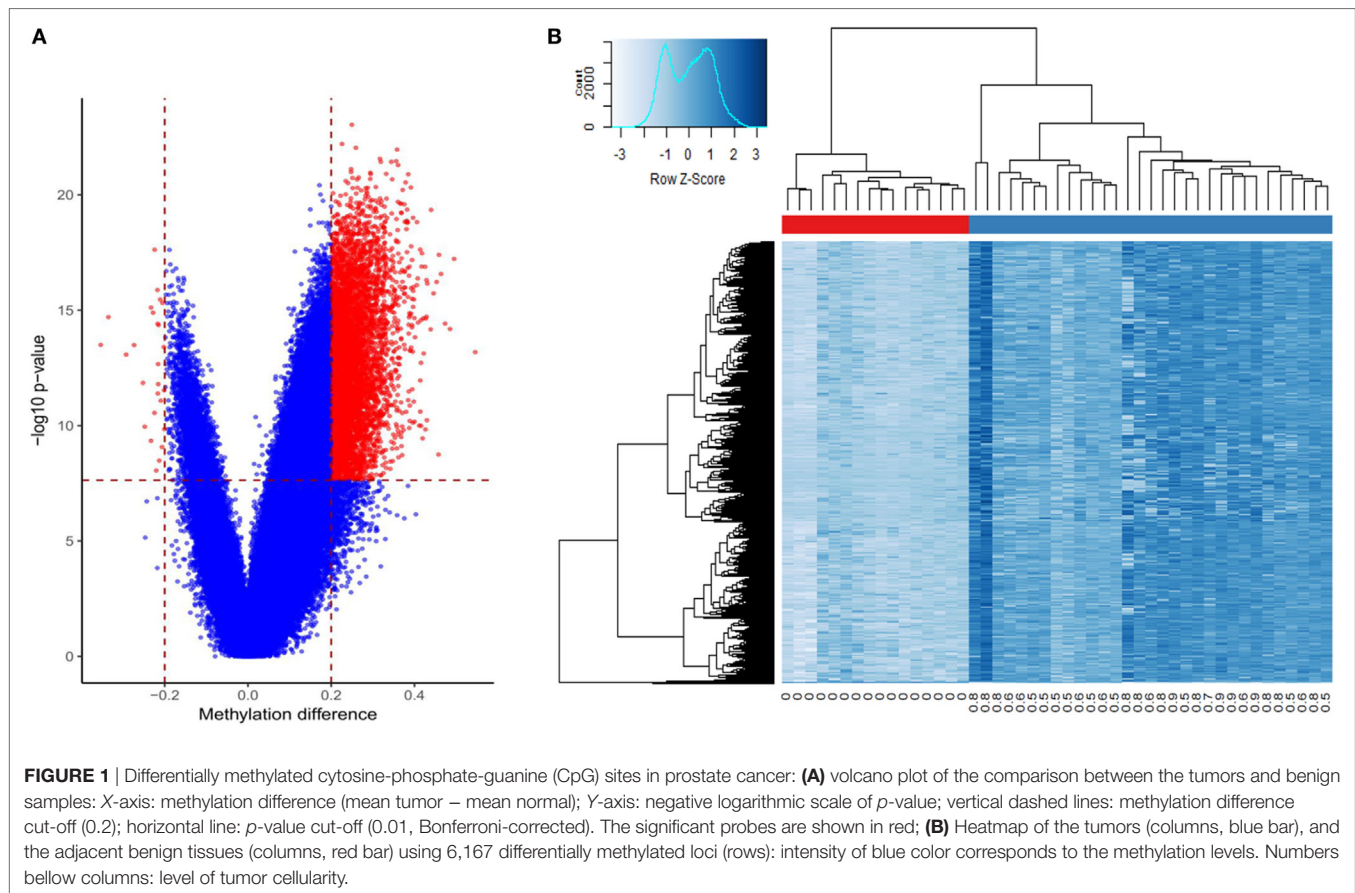
Material; **Figure 1A**). Out of these CpG sites, only 27 (0.4%) showed a hypomethylation change. The vast majority of these probes were annotated to a coding gene ($n = 4,334$, 70.2%). A significant number showed an association with regulatory features, including 5,011 (81.2%) being located in or nearby a CpG island, 1,959 (31.7%) being annotated to DNase Hypersensitivity Sites, and 2,000 (32.3%) located in an enhancer element. Using an unsupervised hierarchical clustering, these probes were shown to be able to completely separate the two groups. The degree of hypermethylation was observed to be correlated with the level of tumor cellularity (**Figure 1B**). A correlation analysis between the mean methylation levels of the entire profile and tumor cellularity revealed a Pearson correlation coefficient of 0.92 (Figure S1 in Supplementary Material).

Cell Communication Is the Central Functional Entity of PCa Methylation Profile

Gene-set enrichment analysis of the CpG sites with differential methylation identified 466 enriched GO terms (biological processes) with a multiple testing corrected p -value < 0.01 , considering the number of probes in every gene (Table S2 in Supplementary Material). The most frequent terms in this list include various forms of cell–cell communications, growth, senescence, bone morphogenetic protein signaling, neurogenesis, and differentiation (**Figure 2**). Protein interaction analysis of the differentially methylated promoters using the *EpiMod* algorithm identified a total of 9 genes as a hotspot for protein–protein interactions containing a minimum of 10 interacting partners and an FDR < 0.01 (Tables S3 and S4 in Supplementary Material; Figure S2 in Supplementary Material). Among these are genes known to be involved in carcinogenesis, including C-X-C motif chemokine ligand 5 (*CXCL5*), estrogen-related receptor gamma (*ESRRG*), and sclerostin domain containing 1. The most active hotspot, however, was found for collagen type-3 alpha-1, located at the center of an interactive network of *COL5A2*, *COL5A1*, *COL11A1*, *COL11A2*, *PCOLCE*, and *HTRA1*, all being members of extracellular matrix regulation pathways (Table S4 in Supplementary Material; Figure S2 in Supplementary Material).

Genomic Regions Differentially Methylated (DMRs) in PCa

Using a bump hunting approach, we found 613 genomic coordinates containing a minimum of three consecutive CpGs, an average regional methylation difference >0.2 , and a FWER < 0.01 (Table S5 in Supplementary Material). All of these regions were hypermethylated, and the vast majority was mapped to coding genes and CpG Islands. Among the largest and most differentially methylated segments are hypermethylation in the promoters of tumor suppressors (*APC*, 10 probes, 30%; *KLK10*, 4 probes, 28%), genes involved in regulation of cellular senescence (*HIF3A*, 8 probes, 31%), cell communication and adhesion [protocadherin (PCDH) gene cluster ($n = 88$), 3–9 probes, 21–36%], and genes regulating the growth including *PDE4D* (6 probes, 23%) and estrogen receptor-related gamma (*ESRRG*, 6 probes, 24%).



In addition, a number of chromatin regulators such as histone deacetylase 9 (*HDAC9*, 6 probes, 26%) and *TET1* (4 probes, 31%), as well as genes coding for histone subunits (*HIST1H-1A*, *3G*, *4D*, and *4F*) were found to show increased methylation in their promoters (3–7 probes, 22–27%).

Copy Number Changes (CNVs) in PCa Correlate With DMR Status

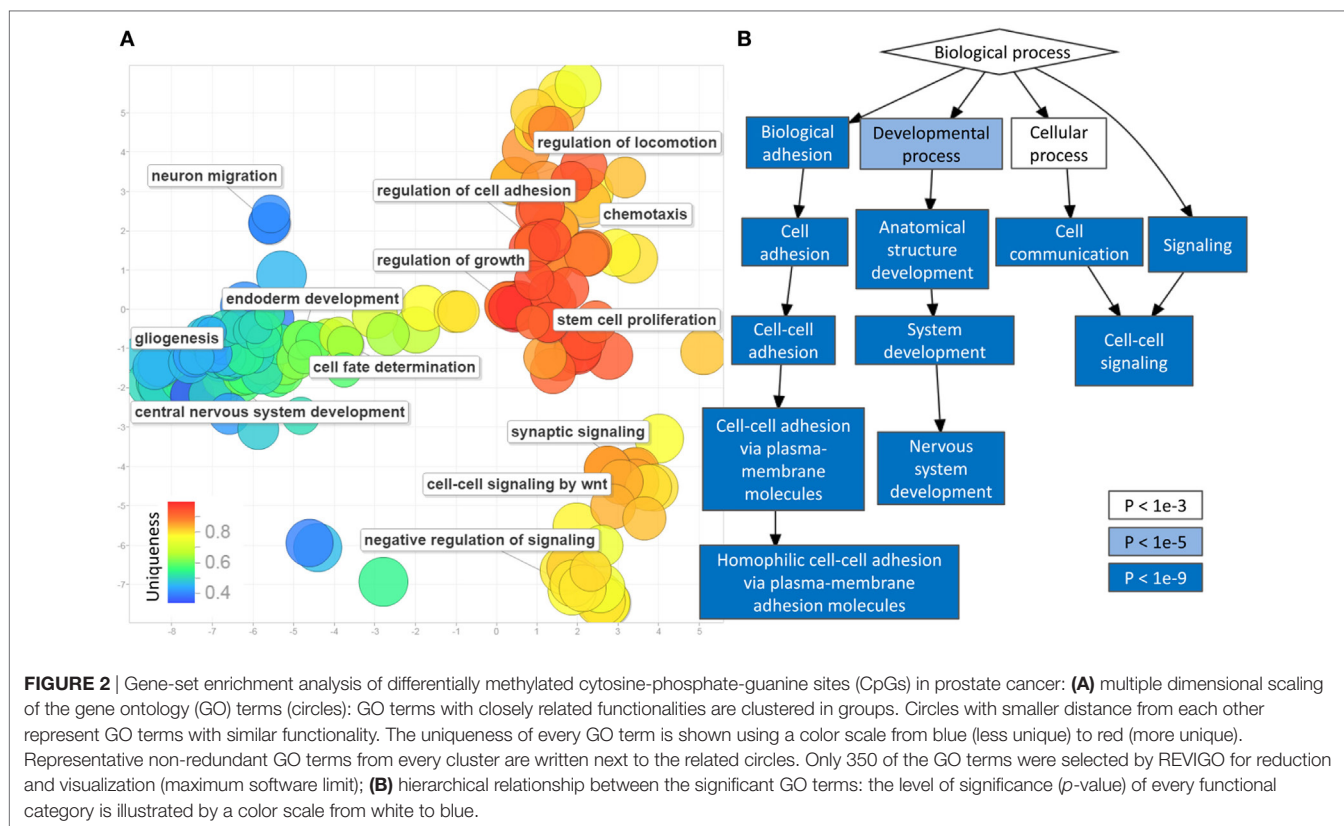
Copy number analysis identified a total of 33 segments in the tumor samples containing a minimum of five markers and one copy of loss/gain, observed in more than one tumor (Table S6 in Supplementary Material). Most of these regions were located nearby coding genes. Out of these, two segments, one in chr2: 29338077–29338258 and the other in chr15: 45421860–45422325, annotating to *CLIP4* and *DUOX1*, were most recurrently amplified among the tumors (in seven and eight samples, respectively). Comparison of the CNV and DMR lists revealed that both of these regions also show an increased level of methylation (Table S5 in Supplementary Material; **Figure 3**). The same incidence was also observed for regions in the promoters of *KRT3* and *PLEC* (each amplified in three tumors), but not among other CNVs. For instance, a segment in the promoter of *PNCK* (amplified in three tumors) did not reveal any hypermethylation change. These results suggested that the hypermethylation pattern seen in some of the DMRs might be associated with their CNV status.

DMRs in PCa Are Surrounded by Tumor Suppressor-Binding Motifs

Analysis of 1,946 Homo-sapiens-specific TF-binding motifs from the Cis-BP database revealed that the DMR sequences rarely bind to any of these TFs. Similarly, a TF-binding site enrichment analysis found an inverse enrichment for any of these motifs among these DMR sequences (enrichment ratios < 1, data not shown). We hypothesized that this might be related to the possibility that DMRs are not a direct target for regulatory elements; rather, they occur as a result of *cis*-regulation by TFs binding to the sites around them. Therefore, we expanded the DMRs by 5 kb upstream and downstream and re-analyzed the data. We found enrichment for 25 motifs specific to 18 TFs (minimum fold change: 1.5, Bonferroni corrected p -value < 0.01, **Figure 4**). Except for few being chromatin regulators (CGBP, MBD2, and DNMT1), all of these TFs were found to have tumor suppressor functionality (*BRCA1*, *E2F1&2&5&7*, *EGR1-4*, *FLI1*, *NRF1*, *TFDP2*, and *STAT1*), or indirectly mediate the suppression of tumorigenesis and tumor suppressor activity (*SP4* and *ZBTB33*).

A Four-CpG Classification Model Enables Accurate Diagnosis of PCa

To design a classification model with the ability of detecting PCa tumor status from methylation profile, we first performed feature selection from the significant PCa-related probes and then trained



a LASSO model with a binary outcome (tumor vs. normal). To avoid a potential bias from the variations in DNA copy numbers, we excluded the probes located in CNV regions. Next, we selected the non-redundant and fully differentiating probes (AUC = 1.00), which resulted in the retention of 46 CpGs. These probes were incorporated into a LASSO model, which following penalization, assigned coefficients equal to 0 to 42 of the probes, dropping them out of the model. Therefore, the final model was trained using four CpGs (**Figure 5A**) on 16 normal samples and 31 tumors (model details in Table S7 in Supplementary Material). These four CpGs are located in the first intron of *OLFM1*, promoter of *RFX7*, 12th intron of *PTPRN2*, and the promoter of *FLOT1*, respectively (Table S7 in Supplementary Material).

To test the model, we first assessed its performance on the training data. As expected, all of the samples were assigned a correct classification of tumor vs. benign (accuracy, sensitivity, specificity, and AUC of 100%; **Figure 5B**). Next, to validate our algorithm, we downloaded DNA methylation array data of prostate samples publically available from GEO, and blindly supplied their methylation levels from the four CpGs to our model for classification. Of a total of 234 tumor and 76 benign prostate arrays (adjacent tissue to tumor) that we found in GEO, the algorithm correctly classified 225 as having a tumor profile and 74 as being benign, respectively. Next, we downloaded six normal prostate arrays from radical prostatectomy due to bladder cancer and ten arrays from BPH samples and supplied them to our model, all of which were classified as normal, suggesting that our algorithm can also differentiate BPH from the tumor

(**Figure 5C**). Altogether, this analysis revealed a sensitivity of 0.96 (9 false negatives in 234 tumors), specificity of 0.98 (2 false positives among 74 adjacent benign, 6 BPH, and 10 normal), an overall accuracy of 0.97, and AUC of 0.98 for the validating dataset (**Figure 5B**). Among other files available from GEO were arrays from bone ($n = 3$), lymph node ($n = 2$), and soft tissue ($n = 1$) with metastases from PCa, all of which were confidently predicted to have a methylation profile similar to PCa. In addition, a total of 61 technical replicates of the prostate tumors were available from datasets GSE83917 and GSE55479. Prediction scores generated for these files remained consistent with their original pairs, including the false negatives, suggesting that our model is not sensitive to technical variation. The prediction scores for all of the above array files are presented in **Figure 5C** and Tables S8 and S9 in Supplementary Material.

DISCUSSION

In this report, we have described DNA methylation changes in PCa as compared to the adjacent benign tissue and reported a hypermethylation profile, which is overrepresented in genes and interactive networks that regulate cell-cell signaling, cell communications, growth, and differentiation. Our study has revealed that the DMRs in PCa do not directly overlap regulatory elements; instead, they are surrounded by TF binding motifs specific to tumor suppressors. And finally, we have demonstrated that by using as few as four CpGs one can accurately classify prostate specimens as malignant or benign.

Insights Into the Involvement of DNA Methylation in PCa

A hypermethylation profile has been reported by previous DNA methylation studies of PCa, which appears to be maintained in

all stages of the disease, likely due to selection pressure (29, 35). Previous reports have also found that, in contrast to the hypermethylated CpGs that will gain further methylation during tumor progression, hypomethylation events are less likely

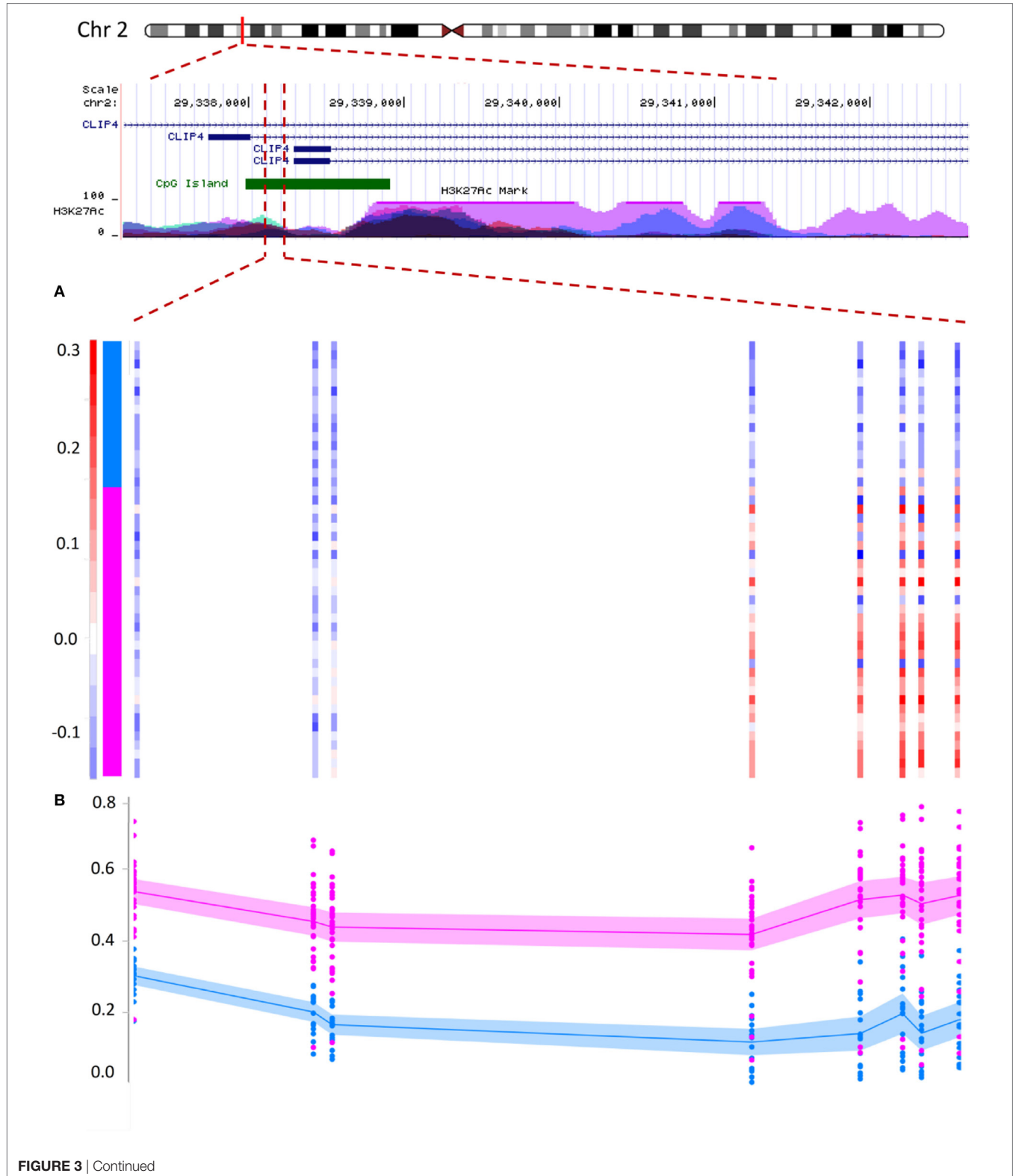


FIGURE 3 | Continued

FIGURE 3 | Correlation between copy number variation (CNV) status and DNA methylation in prostate cancer: hypermethylation in the promoter of the *CLIP4* gene partially correlates with its CNV status. The figure illustrates a segment in the promoter of *CLIP4* gene, located in the short arm of chromosome 2. The region is marked with high levels of acetylation of the 27th lysine residue of Histone 3, (H3K27Ac), a marker associated with active promoters. The segment is also recognized as a cytosine-phosphate-guanine (CpG) island (green pane). Panels (A,B) represent 181 base-pairs of this segment, harboring a total of 8 CpG probes, which were identified to concurrently show both hypermethylation and CNV amplification. (A) The color scale of the vertical bars (each probe) represents the log ratios of the copy numbers. The color scale above 0.2 is shown with red and indicates a minimum of one copy amplification for the region. Color scales of white and light blue represent no CNV change (we defined a CNV loss with a log ratio less than -0.3, which is not observed for this region). Samples are sorted from top to bottom. The top 16 samples represent normal tissues and the lower samples indicate the tumors (vertical pane as an indicator: blue and pink). Within this segment, the right five probes show CNV amplification in tumors, but not in the adjacent benign tissues. (B) The methylation status of the same eight probes shows a hypermethylation in tumors (pink) relative to normal adjacent benign tissues (blue). Methylation level of every probe from every sample is presented with a dot, representing a methylation range between 0 and 1 (bottom to top). Lines represent mean, and shadows around the lines indicate 95% confidence intervals of the mean in every group. The region with CNV amplification is significantly hypermethylated; however, this hypermethylation extends beyond the CNV to three probes in the upstream.

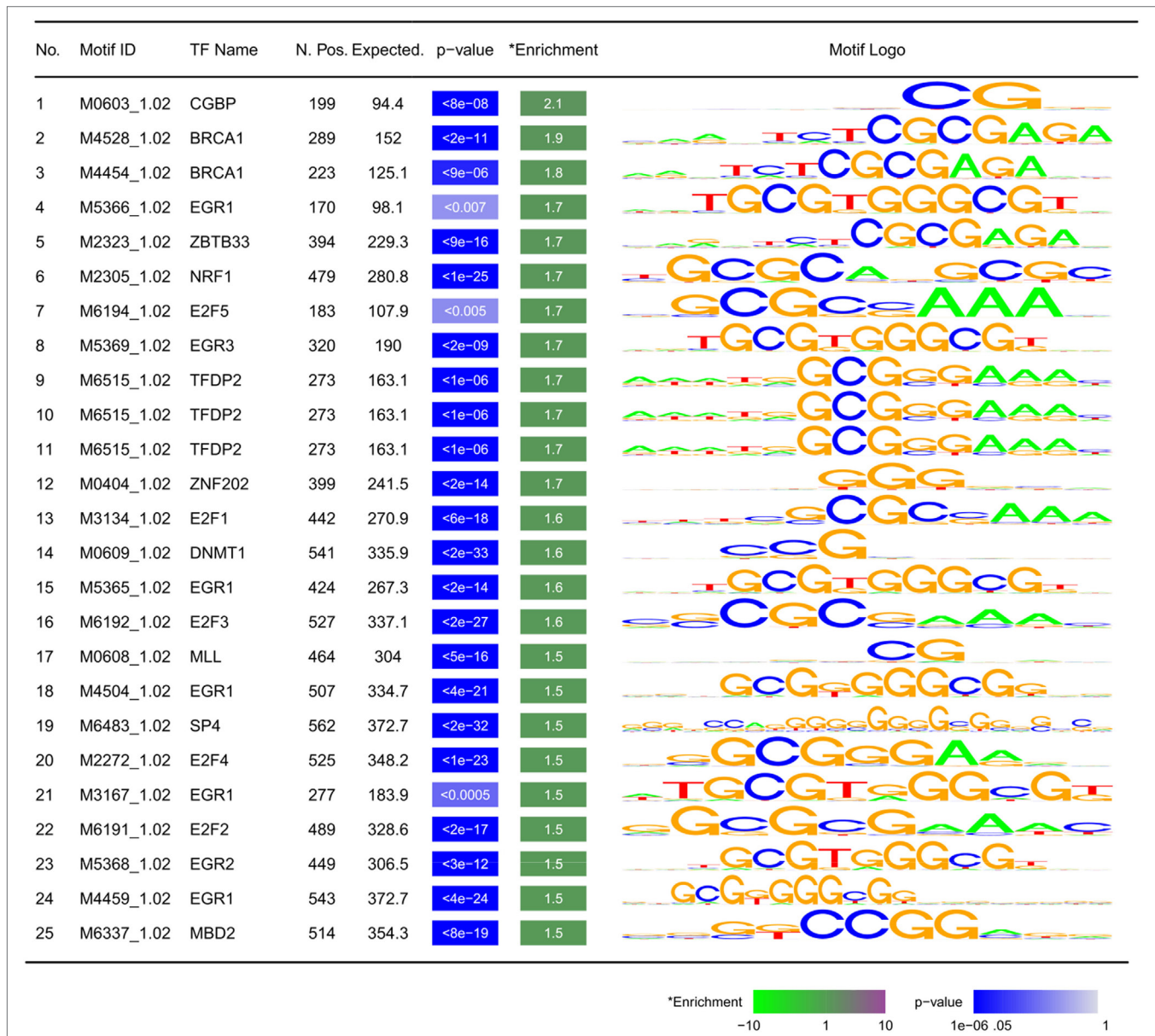
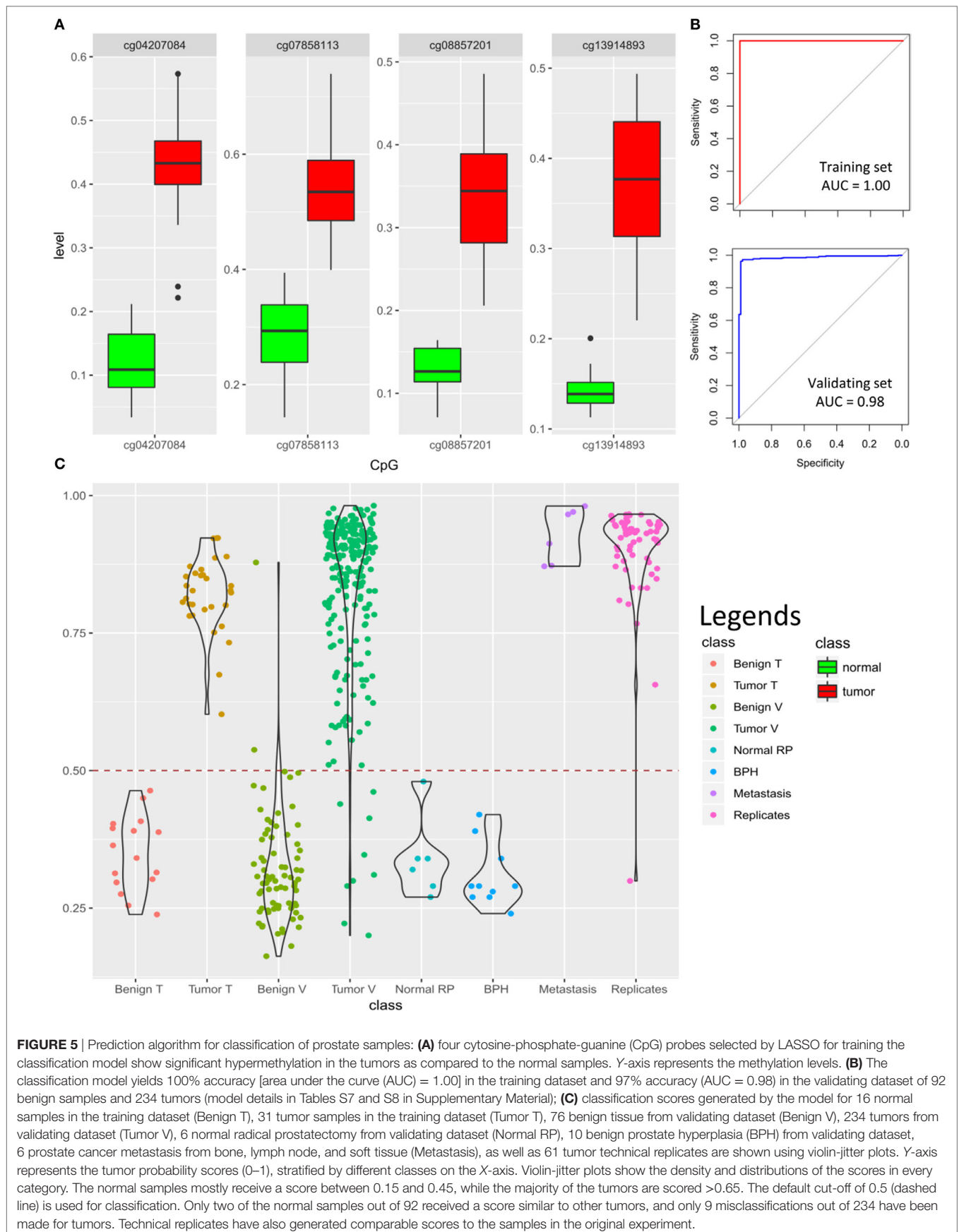


FIGURE 4 | Motifs enriched within ±5 kb of differentially methylated regions: TF, transcription factor; N. Pos.: number of observed motifs; expected: number of expected motifs; enrichment: fold enrichment; p-value is corrected for multiple testing using Bonferroni method. Motifs are sorted by fold enrichment.



to be maintained in the more advanced stages of the disease (33). Consistent with these results, we observed that the level of hypermethylation in PCa directly correlates with the degree of tumor cellularity in the primary prostate specimens (Figure S1 in Supplementary Material). The postulated link between a hypermethylation profile and cancer is a model of translational silencing of tumor suppressors by promoter hypermethylation (6). Indeed, we have observed increased methylation in the promoters of several tumor suppressor genes including *APC* (Figure S3 in Supplementary Material), a well-characterized tumor suppressor gene associated with familial adenomatous polyposis, and *KLK10*, which has been shown to repress proliferation and induce apoptosis in PCa cells (36). Among other DMRs in genes involved in growth and differentiation in our study are *PDE4D*, encoding a signal transduction molecule with cyclic-AMP phosphodiesterase activity that is shown in mice to promote proliferation of PCa (37). As well, we have observed promoter hypermethylation in *ESRRG*, which was also found to act as a hotspot for protein-protein interactions. Downregulation of *ESRRG* has been found in certain types of prostate carcinomas, and it has been shown that its increased expression can repress tumor proliferation regardless of androgen sensitivity (38).

Besides tumor suppressors with a potential role in cell growth and proliferation, analyses of gene-set enrichment and protein-protein interaction hotspots have revealed that methylation changes in PCa encompass a combination of biological processes extending beyond the regulation of growth to neurogenesis and cell-cell communication. The best-known involvement of neurons in cancer is perineural invasions, where tumor invades the neural tissue. Another form of neural involvement in cancer, namely neurogenesis, has been recently reported in prostate tumors as the formation of neural components, axonogenesis, and increased number of neurons (39). Neurogenesis has been found to correlate with both perineural invasions and poor clinical outcomes in PCa (39). Our gene enrichment analysis contains multiple GO terms related to the development of the nervous system, including brain development, axonogenesis, and neurogenesis. In parallel, these analyses provide evidence for the involvement of diverse forms of cell-cell signaling and communications, ranging from regulation of the extracellular matrix to chemotaxis and cellular response to stimuli. Among all of the genes in our results that are known to take part in these processes, the *PCDH* genes are mutually involved in both neurogenesis and cell-cell communications. Our DMRs contain extensive regions that overlap the promoters of close to 90 *PCDH gamma* genes and isoforms. Various members of this cluster have been shown to be present in cell-cell adhesion, neural stroma and synapses (40), and homophilic trans-interactions (41). Overexpression of *PCDHs in vitro* has been demonstrated to suppress Wnt signaling and to inhibit colony formations of cancer cells (42). Limited data suggest an involvement of *PCDH 8, 11, and PC* in PCa; however, no information is available regarding the *PCDH gamma* genes in PCa (43–45). Hypermethylation of various subtypes of *PCDH gamma* promoter is reported in Wilms' tumor, colorectal carcinoma, uterine leiomyosarcoma, esophageal adenocarcinoma, and Barrett's esophagus (46).

Overall, the major biological mechanisms that can be inferred from the methylation changes in this study represent the driver events, required for the maintenance, integrity, and survival of the tumor rather than those that may initiate tumorigenesis resulting from the loss of tumor suppressor functions. Similarly, the most recurrent CNV changes in tumors were found as amplifications in the promoters of *DUOX1*, a NADPH oxidase involved in the maintenance of tissue homeostasis (47), as well as *CLIP4*, an intracellular linker protein whose knock-down has been shown to increase cell migration and viability in clear cell renal cell carcinomas (48). These two segments were also found to be hypermethylated, potentially by the expansion of the CG repeats or methylation quantitative trait loci resulting from CNV amplification (4). Therefore, some of the hypermethylation changes we have observed in PCa might be related to the CNV status of the regions. However, the more common phenomenon that is potentially involved in the establishment of the DMRs appears to be a dysregulation following a change in tumor suppressor activity. The interesting finding that almost the entire enriched motifs nearby DMRs are specific to tumor suppressors indicates that DMRs in PCa, which mainly represent cancer driver events, are potentially generated or maintained secondary to a dysregulation in tumor suppressors' function.

Clinical Implications of the 4-CpG Classification Model

The current method widely in use for the diagnosis of PCa is an 8–12 core needle biopsy, which is well known to encompass a high false-negative rate (49). A significant number of cancer-free reports of the prostate biopsies are shown to be at risk of having an undiagnosed PCa (49). Approximately 25–50% of these men are diagnosed with PCa in a second biopsy performed within 1 year (50). On the other hand, a positive needle biopsy is not confidently replicable. Serefoglu et al. have shown that a repeat PCa diagnosis can be made only in 67.8% of biopsies from post-operative prostate glands of men undergone radical prostatectomy due to PCa, using the same 12-core mapping as performed pre-operatively (51). The main reason for this discrepancy is the spatiotemporal, molecular, and morphological heterogeneity of the prostate adenocarcinoma that leads to a non-uniform presentation of the tumor by the involved tissue (52). This indicates that a significant portion of negative biopsies can contain molecular changes associated with cancer without completely representing histological features of adenocarcinoma (53, 54), and thus a reliable biological marker would assign a diagnosis to a considerable number of suspicious samples undiagnosed following histopathology examination. Consistent with this, Troyer et al. was among the first to publish a report that DNA methylation markers can detect PCa in histopathologically cancer-negative prostate tissues of men who went on to have subsequently positive biopsies (55). Within a few years of this report, methylation markers in the promoters of *GSTP1*, *APC*, and *RASSF1* were introduced as a commercial tissue-based assay to identify patients in need of repeat biopsies (56). However, clinical validation studies of this assay have only revealed a sensitivity of 68% and a specificity of 64% (57, 58), questioning the reliability of a candidate gene approach in PCa

screening. We have also observed hypermethylation to a variable extent in all these three genes (Tables S1 and S5 in Supplementary Material); however, this is not consistently found across all specimens (Figure S3 in Supplementary Material for APC), thereby resulting in reduced sensitivity for cancer tissue specification. As a consequence, efforts to identify DNA methylation biomarkers shifted toward a genome-wide approach for PCa. Goh et al. (59), following a genome-wide DNA methylation analysis, have proposed a 55-CpG classification model, which has yielded 89.8% sensitivity and 66.7% specificity in the validation step. More recently, Kirby et al. have reported a 3-CpG model for classification of tumor samples from the adjacent benign tissues, which reached a sensitivity of 90% and specificity of 82% as measured in a validation cohort of 49 benign and 213 tumor samples (29).

Here, we have further improved the classification of prostate samples by designing a 4-CpG classification model, which has revealed 100% accuracy in separating tumors from benign tissue in our internal dataset. The validation of this model using heterogeneous cohorts of 326 prostate samples from publically available resources has revealed 96% sensitivity and 98% specificity with an overall accuracy of 97% (AUC = 0.98). To our knowledge, this far supersedes performance of any reported DNA methylation-based algorithm for the diagnosis of PCa. Our model is sensitive enough to detect PCa metastasis in other tissues, including bone, soft tissue, and lymph nodes (Figure 5), while it is still specific enough to discriminate BPH from the tumor. This method can be used to develop a low-cost high-throughput targeted assay for diagnosis of PCa or to complement the pathological examination of cancer-negative needle biopsies. It may also be used for confirmation of metastatic lesions outside of the prostate.

Future Directions

The current classification algorithm presented in this study is designed and validated using tumor samples and normal adjacent tissues obtained from radical prostatectomy. Before this method is translated into clinical use, further computational training, validation, and clinical trials are required. First, the performance of the model needs to be assessed on samples with diverse ranges of tumor cellularity. In the present study, we did not assess samples with tumor involvement less than 50%, and thus, the accuracy remains unknown with regards to samples with lower levels of tumor cellularity. In a small subset of the tumor samples in the validation dataset, we observed scores close to the ones found in normal tissues, which might be caused by a potentially low level of tumor cellularity in them. Computationally, we can re-train the algorithm to enable the prediction of tumor cellularity as a continuous measure. In addition, the model will have to be tested and potentially re-trained on DNA obtained from needle core

biopsies. In order to make this technology more broadly applicable, its accuracy should be demonstrated on limited amounts of DNA obtained from needle biopsy specimens, and address whether the performance is similar across the biopsy specimens and the entire prostate gland. Another important aspect to assess is whether this assay can detect neoplastic transformation in specimens that appear to be normal in an initial pathological examination, but are demonstrated to harbor tumor in subsequent assessments of the same core (i.e., false negatives).

The markers and methodology presented here could be expanded to research applications with the aim of developing novel approaches in the non-invasive screening of PCa using circulating tumor DNA. The study we have described herein reveals the power of high-throughput genomic technologies, in combination with machine learning, to increase our understanding of the biology of cancer and, at the same time, to provide us with novel, precise approaches for cancer screening and diagnosis.

ETHICS STATEMENT

This study has been approved by the Hamilton Integrated Research Ethics Board (#14-700-T). All of the samples and records were de-identified prior to analysis.

AUTHOR CONTRIBUTIONS

EA-E: performed statistical and bioinformatics analysis and machine learning, and wrote the manuscript; LS and J-CC: collected the specimens; PA, HL, and DR: commented on the manuscript; BS: oversaw the study protocol.

ACKNOWLEDGMENTS

EA-E. was supported by Children's Health Research Institute Epigenetics Trainee Award, funded by the Children's Health Foundation, London, Ontario, Canada.

FUNDING

This study was funded by the Juravinski Hospital and Cancer Centre Foundation's (JHCCF) research development grant.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <https://www.frontiersin.org/articles/10.3389/fonc.2018.00100/full#supplementary-material>.

REFERENCES

- Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA Cancer J Clin* (2011) 61(2):69–90. doi:10.3322/caac.20107
- Wojno K, Hornberger J, Schellhammer P, Dai M, Morgan T. The clinical and economic implications of specimen provenance complications in diagnostic prostate biopsies. *J Urol* (2015) 193(4):1170–7. doi:10.1016/j.juro.2014.11.019
- Gaudreau PO, Stagg J, Soulières D, Saad F. The present and future of biomarkers in prostate cancer: proteomics, genomics, and immunology advancements. *Biomark Cancer* (2016) 8(Suppl 2):15–33. doi:10.4137/BIC.S31802
- Schenkel LC, Rodenhiser DI, Ainsworth PJ, Paré G, Sadikovic B. DNA methylation analysis in constitutional disorders: clinical implications of the epigenome. *Crit Rev Clin Lab Sci* (2016) 53(3):147–65. doi:10.3109/10408363.2015.1113496
- Mikeska T, Craig JM. DNA methylation biomarkers: cancer and beyond. *Genes (Basel)* (2014) 5(3):821–64. doi:10.3390/genes5030821

6. Li LC, Carroll PR, Dahiya R. Epigenetic changes in prostate cancer: implication for diagnosis and treatment. *J Natl Cancer Inst* (2005) 97(2):103–15. doi:10.1093/jnci/dji010
7. Geybels MS, Wright JL, Bibikova M, Klotzle B, Fan JB, Zhao S, et al. Epigenetic signature of Gleason score and prostate cancer recurrence after radical prostatectomy. *Clin Epigenetics* (2016) 8(1):97. doi:10.1186/s13148-016-0260-z
8. Baden J, Adams S, Astacio T, Jones J, Markiewicz J, Painter J, et al. Predicting prostate biopsy result in men with prostate specific antigen 2.0 to 10.0 ng/ml using an investigational prostate cancer methylation assay. *J Urol* (2011) 186(5):2101–6. doi:10.1016/j.juro.2011.06.052
9. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* (2014) 30(10):1363–9. doi:10.1093/bioinformatics/btu049
10. Schenkel LC, Schwartz C, Skinner C, Rodenhiser DI, Ainsworth PJ, Pare G, et al. Clinical validation of fragile X syndrome screening by DNA methylation array. *J Mol Diagn* (2016) 18(6):834–41. doi:10.1016/j.jmoldx.2016.06.005
11. Kerkhof J, Schenkel LC, Reilly J, McRobbie S, Aref-Eshghi E, Stuart A, et al. Clinical validation of copy number variant detection from targeted next-generation sequencing panels. *J Mol Diagn* (2017) 19(6):905–20. doi:10.1016/j.jmoldx.2017.07.004
12. Kernohan KD, Schenkel LC, Huang L, Smith A, Pare G, Ainsworth P, et al. Identification of a methylation profile for DNMT1-associated autosomal dominant cerebellar ataxia, deafness, and narcolepsy. *Clin Epigenetics* (2016) 8(1):91. doi:10.1186/s13148-016-0254-x
13. Schenkel LC, Aref-Eshghi E, Skinner C, Ainsworth P, Lin H, Paré G, et al. A peripheral blood epi-signature enables sensitive and specific identification of affected males and female carriers with KDM5C related X-linked intellectual disability. *Clin Epigenetics* (2018) 10:21. doi:10.1186/s13148-018-0453-8
14. Aref-Eshghi E, Schenkel LC, Lin H, Skinner C, Ainsworth P, Paré G, et al. The defining DNA methylation signature of Kabuki syndrome enables functional assessment of genetic variants of unknown clinical significance. *Epigenetics* (2017) 12(11):923–33. doi:10.1080/15592294.2017.1381807
15. Schenkel LC, Kernohan KD, McBride A, Reina D, Hodge A, Ainsworth PJ, et al. Identification of epigenetic signature associated with alpha thalassemia/mental retardation X-linked syndrome. *Epigenetics Chromatin* (2017) 10(1):10. doi:10.1186/s13072-017-0118-4
16. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* (2015) 43(7):e47. doi:10.1093/nar/gkv007
17. Aref-Eshghi E, Rodenhiser DI, Schenkel LC, Lin H, Skinner C, Ainsworth P, et al. Genomic DNA methylation signatures enable concurrent diagnosis and clinical genetic variant classification in neurodevelopmental syndromes. *Am J Hum Genet* (2018) 102(1):156–74. doi:10.1016/j.ajhg.2017.12.008
18. Phipson B, Maksimovic J, Oshlack A. missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. *Bioinformatics* (2015) 32(2):286–8.
19. Supek F, Bošnjak M, Škunca N, Šmuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* (2011) 6(7):e21800. doi:10.1371/journal.pone.0021800
20. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *Bioinformatics* (2009) 10(1):48. doi:10.1186/1471-2105-10-48
21. Jiao Y, Widschwendter M, Teschendorff AE. A systems-level integrative framework for genome-wide DNA methylation and gene expression data identifies differential gene expression modules under epigenetic control. *Bioinformatics* (2014) 30(16):2360–6. doi:10.1093/bioinformatics/btu316
22. Jaffe AE, Murakami P, Lee H, Leek JT, Fallin MD, Feinberg AP, et al. Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. *Int J Epidemiol* (2012) 41(1):200–9. doi:10.1093/ije/dyr238
23. Bolstad B. *preprocessCore: A Collection of Pre-Processing Functions*. R Package Version 1.40.0. (2017). Available from: <https://github.com/bmbolstad/preprocessCore> (Accessed: April 2, 2018).
24. Seshan VE, Olshen A. *DNACopy: DNA Copy Number Data Analysis*. R Package Version 1.52.0. (2017). Available from: <https://www.bioconductor.org/packages/release/bioc/html/DNACopy.html>
25. Lawrence M, Huber W, Pages H, Aboyoun P, Carlson M, Gentleman R, et al. Software for computing and annotating genomic ranges. *PLoS Comput Biol* (2013) 9(8):e1003118. doi:10.1371/journal.pcbi.1003118
26. Wang Z, Martins AL, Danko CG. RTFBSDB: an integrated framework for transcription factor binding site analysis. *Bioinformatics* (2016) 32(19):3024–6. doi:10.1093/bioinformatics/btw338
27. Weirauch MT, Yang A, Albu M, Cote AG, Montenegro-Montero A, Drewe P, et al. Determination and inference of eukaryotic transcription factor sequence specificity. *cell* (2014) 158(6):1431–43.
28. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* (2010) 33(1):1–22. doi:10.18637/jss.v033.i01
29. Kirby MK, Ramaker RC, Roberts BS, Lasseigne BN, Gunther DS, Burwell TC, et al. Genome-wide DNA methylation measurements in prostate tissues uncovers novel prostate cancer diagnostic biomarkers and transcription factor binding patterns. *BMC Cancer* (2017) 17(1):273. doi:10.1186/s12885-017-3252-2
30. Ramalho-Carvalho J, Graça I, Gomez A, Oliveira J, Henrique R, Esteller M, et al. Downregulation of miR-130b~301b cluster is mediated by aberrant promoter methylation and impairs cellular senescence in prostate cancer. *J Hematol Oncol* (2017) 10(1):43. doi:10.1186/s13045-017-0415-1
31. Naem H, Wong NC, Chatterton Z, Hong MK, Pedersen JS, Corcoran NM, et al. Reducing the risk of false discovery enabling identification of biologically significant genome-wide methylation status using the HumanMethylation450 array. *BMC Genomics* (2014) 15(1):51. doi:10.1186/1471-2164-15-51
32. Fraser M, Sabelnikova VY, Yamaguchi TN, Heisler LE, Livingstone J, Huang V, et al. Genomic hallmarks of localized, non-indolent prostate cancer. *Nature* (2017) 541(7637):359–64. doi:10.1038/nature20788
33. Aryee MJ, Liu W, Engelmann JC, Nuhn P, Gurel M, Haffner MC, et al. DNA methylation alterations exhibit intraindividual stability and interindividual heterogeneity in prostate cancer metastases. *Sci Transl Med* (2017) 5(169):169ra10. doi:10.1126/scitranslmed.3005211
34. Paziewska A, Dabrowska M, Goryca K, Antoniewicz A, Dobruch J, Mikula M, et al. DNA methylation status is more reliable than gene expression at detecting cancer in prostate biopsy. *Br J Cancer* (2014) 111(4):781–9. doi:10.1038/bjc.2014.337
35. Geybels MS, Zhao S, Wong CJ, Bibikova M, Klotzle B, Wu M, et al. Epigenomic profiling of DNA methylation in paired prostate cancer versus adjacent benign tissue. *Prostate* (2015) 75(16):1941–50. doi:10.1002/pros.23093
36. Hu J, Lei H, Fei X, Liang S, Xu H, Qin D, et al. NES1/KLK10 gene represses proliferation, enhances apoptosis and down-regulates glucose metabolism of PC3 prostate cancer cells. *Sci Rep* (2015) 5:17426. doi:10.1038/srep17426
37. Rahrmann EP, Collier LS, Knutson TP, Doyal ME, Kuslak SL, Green LE, et al. Identification of PDE4D as a proliferation promoting factor in prostate cancer using a Sleeping Beauty transposon-based somatic mutagenesis screen. *Cancer Res* (2009) 69(10):4388–97. doi:10.1158/0008-5472.CAN-08-3901
38. Misawa A, Inoue S. Estrogen-related receptors in breast cancer and prostate cancer. *Front Endocrinol* (2015) 6:83. doi:10.3389/fendo.2015.00083
39. Ayala GE, Dai H, Powell M, Li R, Ding Y, Wheeler TM, et al. Cancer-related axonogenesis and neurogenesis in prostate cancer. *Clin Cancer Res* (2008) 14(23):7593–603. doi:10.1158/1078-0432.CCR-08-1164
40. Kohmura N, Senzaki K, Hamada S, Kai N, Yasuda R, Watanabe M, et al. Diversity revealed by a novel family of cadherins expressed in neurons at a synaptic complex. *Neuron* (1998) 20(6):1137–51. doi:10.1016/S0896-6273(00)80495-X
41. Schreiner D, Weiner JA. Combinatorial homophilic interaction between γ -protocadherin multimers greatly expands the molecular diversity of cell adhesion. *Proc Natl Acad Sci U S A* (2010) 107(33):14893–8. doi:10.1073/pnas.1004526107
42. Dallosso AR, Hancock AL, Szemes M, Moorwood K, Chilukamarri L, Tsai HH, et al. Frequent long-range epigenetic silencing of protocadherin gene clusters on chromosome 5q31 in Wilms' tumor. *PLoS Genet* (2009) 5(11):e1000745. doi:10.1371/journal.pgen.1000745
43. Bex G, Van Roy F. Involvement of members of the cadherin superfamily in cancer. *Cold Spring Harb Perspect Biol* (2009) 1(6):a003129. doi:10.1101/cshperspect.a003129
44. Terry S, Queires L, Gil-Diez-de-Medina S, Chen MW, De La Taille A, Allory Y, et al. Protocadherin-PC promotes androgen-independent prostate cancer cell growth. *Prostate* (2006) 66(10):1100–13. doi:10.1002/pros.20446
45. Zhang P, Wang H, Wang J, Liu Q, Wang Y, Feng F, et al. Association between protocadherin 8 promoter hypermethylation and the pathological

- status of prostate cancer. *Oncol Lett* (2017) 14(2):1657–64. doi:10.3892/ol.2017.6282
46. El Hajj N, Dittrich M, Haaf T. Epigenetic dysregulation of protocadherins in human disease. *Semin Cell Dev Biol* (2017) 69:172–82. doi:10.1016/j.semcdb.2017.07.007
 47. Luxen S, Belinsky SA, Knaus UG. Silencing of DUOX NADPH oxidases by promoter hypermethylation in lung cancer. *Cancer Res* (2008) 68(4):1037–45. doi:10.1158/0008-5472.CAN-07-5782
 48. Ahn J, Han KS, Heo JH, Bang D, Kang YH, Jin HA, et al. FOXC2 and CLIP4: a potential biomarker for synchronous metastasis of ≤ 7 -cm clear cell renal cell carcinomas. *Oncotarget* (2016) 7(32):51423. doi:10.18632/oncotarget.9842
 49. Kronz JD, Allan CH, Shaikh AA, Epstein JI. Predicting cancer following a diagnosis of high-grade prostatic intraepithelial neoplasia on needle biopsy: data on men with more than one follow-up biopsy. *Am J Surg Pathol* (2001) 25(8):1079–85. doi:10.1097/00000478-200108000-00014
 50. Odowd GJ, Miller MC, Orozco R, Veltri RW. Analysis of repeated biopsy results within 1 year after a noncancer diagnosis. *Urology* (2000) 55(4):553–8. doi:10.1016/S0090-4295(00)00447-7
 51. Serefoglu EC, Altinova S, Ugras NS, Akincioglu E, Asil E, Balbay MD. How reliable is 12-core prostate biopsy procedure in the detection of prostate cancer? *Can Urol Assoc J* (2013) 7(5–6):E293. doi:10.5489/cuaj.11224
 52. Cyll K, Ersvaer E, Vlatkovic L, Pradhan M, Kildal W, Kjaer MA, et al. Tumour heterogeneity poses a significant challenge to cancer biomarker research. *Br J Cancer* (2017) 117(3):367–75. doi:10.1038/bjc.2017.171
 53. Yang B, Bhusari S, Kueck J, Weeratunga P, Wagner J, Levenson G, et al. Methylation profiling defines an extensive field defect in histologically normal prostate tissues associated with prostate cancer. *Neoplasia* (2013) 15(4):399–408. doi:10.1593/neo.13280
 54. Slater MD, Delprado WJ, Murphy CR, Barden JA. Detection of preneoplasia in histologically normal prostate biopsies. *Prostate Cancer Prostatic Dis* (2001) 4(2):92. doi:10.1038/sj.pcan.4500500
 55. Troyer DA, Lucia MS, de Bruïne AP, Mendez-Meza R, Baldewijns MM, Dunscomb N, et al. Prostate cancer detected by methylated gene markers in histopathologically cancer-negative tissues from men with subsequent positive biopsies. *Cancer Epidemiol Biomarkers Prev* (2009) 18(10):2717–22. doi:10.1158/1055-9965.EPI-09-0068
 56. Van Neste L, Bigley J, Toll A, Otto G, Clark J, Delrée P, et al. A tissue biopsy-based epigenetic multiplex PCR assay for prostate cancer detection. *BMC Urol* (2012) 12(1):16. doi:10.1186/1471-2490-12-16
 57. Stewart GD, Van Neste L, Delvenne P, Delrée P, Delga A, McNeill SA, et al. Clinical utility of an epigenetic assay to detect occult prostate cancer in histopathologically negative biopsies: results of the MATLOC study. *J Urol* (2015) 189:1110–6. doi:10.1016/j.juro.2012.08.219
 58. Partin AW, Van Neste L, Klein EA, Marks LS, Gee JR, Troyer DA, et al. Clinical validation of an epigenetic assay to predict negative histopathological results in repeat prostate biopsies. *J Urol* (2014) 192:1081–7. doi:10.1016/j.juro.2014.04.013
 59. Goh LK, Liem N, Vijayaraghavan A, Chen G, Lim PL, Tay KJ, et al. Diagnostic and prognostic utility of a DNA hypermethylated gene signature in prostate cancer. *PLoS One* (2014) 9(3):e91666. doi:10.1371/journal.pone.0091666

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Aref-Eshghi, Schenkel, Ainsworth, Lin, Rodenhiser, Cutz and Sadikovic. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.