

RESEARCH ARTICLE

Open Access

# DNA methylation signatures of Prostate Cancer in peripheral T-cells



Ali Mehdi<sup>1,2†</sup>, David Cheishvili<sup>3,4†</sup>, Ani Arakelian<sup>1</sup>, Tarek A. Bismar<sup>5</sup>, Moshe Szyf<sup>6</sup> and Shafaat A. Rabbani<sup>1,2,7\*</sup> 

## Abstract

**Background:** Prostate Cancer (PCa) is the second most common cancer in men where advancements have been made for early detection using imaging techniques, however these are limited by lesion size. Immune surveillance has emerged as an effective approach for early detection and to monitor disease progression. In recent studies, we have shown that host peripheral blood immune cells undergo changes in DNA methylation in liver and breast cancer.

**Methods:** In the current study, we examined the DNA methylation status of peripheral blood T cells of men with positive biopsy for PCa versus men with negative biopsy having benign prostate tissue, defined as controls. T cells DNA was isolated and subjected to Illumina Infinium methylation EPIC array and validated using Illumina amplicon sequencing and pyrosequencing platforms.

**Results:** Differential methylation of 449 CG sites between control and PCa T cell DNA showed a correlation with Gleason score ( $p < 0.05$ ). Two hundred twenty-three differentially methylated CGs between control and PCa ( $\Delta\beta$   $\pm 10\%$ ,  $p < 0.05$ ), were enriched in pathways involved in immune surveillance system. Three CGs which were found differentially methylated following DMP (Differentially methylated probes) analysis of ChAMP remained significant after BH (Benjamini-Hochberg) correction, of which, 2 CGs were validated. Predictive ability of combination of these 3 CGs (polygenic methylation score, PMS) to detect PCa had high sensitivity, specificity and overall accuracy. PMS also showed strong positive correlation with Gleason score and tumor volume of PCa patients.

**Conclusions:** Results from the current study provide for the first-time a potential role of DNA methylation changes in peripheral T cells in PCa. This non-invasive methodology may allow for early intervention and stratification of patients into different prognostic groups to reduce PCa associated morbidity from repeat invasive prostate biopsies and design therapeutic strategy to reduce PCa associated mortality.

**Keywords:** Prostate cancer, DNA methylation, T cells, Blood, Diagnosis, Immune

\* Correspondence: [shafaat.rabbani@mcgill.ca](mailto:shafaat.rabbani@mcgill.ca)

<sup>†</sup>Ali Mehdi and David Cheishvili contributed equally to this work.

<sup>1</sup>Department of Medicine, McGill University, Montreal, Quebec, Canada

<sup>2</sup>Department of Human Genetics, McGill University, Montreal, Quebec, Canada

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

Prostate Cancer (PCa) is the second most common cancer in men and has been estimated to affect more than a million men worldwide in 2018 [1–4]. PCa is graded using the Gleason score (ranging from 2 to 10) that examines the differentiation of cancer cells according to their histological pattern [1–5]. In its early stage, tumor cells are androgen sensitive however, in advanced stages they are androgen insensitive, metastasize to distant organs and have limited therapeutic strategies available [3, 5]. Therefore, there is an urgent need to identify sensitive biomarkers which can detect initial and recurrent PCa early to allow for effective intervention. Diagnosis for PCa includes digital rectal examination and determination of the levels of prostate specific antigen however, each of these tests have limitations [6, 7]. On the other hand, prostate biopsies are invasive, can provide false-negative results and do not examine the whole prostate [5, 6, 8, 9].

DNA methylation has emerged as a major area of investigation for the identification of cancer specific biomarkers at an early time point which can be utilized as a diagnostic or prognostic biomarker in numerous cancers including PCa [10–18]. DNA methylation alterations are highly common and are particularly able to provide a high degree of sensitivity using plasma and urine of PCa patients [12, 19]. However, these studies were limited in scope due to their focus on a single or few gene loci. Collectively a multi-faceted approach is required to accurately predict and detect PCa for early intervention and develop effective therapeutic strategies.

Host immuno-surveillance plays an important role in the recognition and elimination of transformed and tumor cells [20–22]. T-cells are the most prominent component of this system that control tumor growth [20–22] and hence have the potential of being effective cancer biomarkers [23–25]. In previous studies, we have identified a DNA methylation signature in peripheral blood mononuclear cells and T cells in hepatocellular carcinoma (HCC) which distinguished HCC from chronic hepatitis B and C versus healthy controls [25]. Similarly, DNA methylation signature in T cells of breast cancer patients showed strong correlation with breast cancer progression compared to healthy controls [24].

In the current study, we have carried out genomic DNA methylation assay using peripheral blood T cells from negative-biopsy men having benign prostate tissue, defined as controls, and positive biopsy PCa patients. T cells DNA isolated from control and PCa subjects was analyzed by Illumina Infinium methylation EPIC array platform and validated using amplicon- and pyro- sequencing. We have identified a DNA methylation signature that not only distinguishes healthy controls from PCa patients but also different stages of PCa according to Gleason score.

## Methods

### Study populations

The study design was approved by the ethics committee of Alberta PCa Research Initiative (APCaRI). Buffy coats for positive-biopsy PCa patients and healthy negative-biopsy, having benign prostate tissue defined as controls, men were obtained through APCaRI. All patients were enrolled at the time of diagnosis before initiation of any treatment and had strict exclusion criterion of any autoimmune disease, asthma or infection which could alter T cells characteristics. The study had 12 controls and 20 PCa samples; 9 of which were low Gleason PCa (LGPCa) having a Gleason score  $\leq 7$  and; 11 of which were high Gleason PCa (HGPCa) having a Gleason score  $> 7$  (Cohort 1) (Supplementary Figure 1A and Supplementary Table 1). We used an older cohort (Cohort 2) which had 11 controls and 13 LGPCa samples for some analysis (described below) (Supplementary Table 1). The controls (mean age of 58.0 years) and the PCa patients (mean age of 62.4 years) had no significant difference between mean age ( $p > 0.05$ ). The percentage tumor volume (%) was calculated as the total volume of positive cores divided by all cores tissue and is reflective of the potential cancer volume in the prostate gland.

### T-cell isolation, Methylome assay and pyrosequencing

Firstly, CD3<sup>+</sup> T cells were isolated using anti-CD3 immuno-magnetic beads (Dynabeads®, Invitrogen) from the buffy coats of PCa and healthy controls according to manufacturer's protocol. Genomic DNA was extracted from the T cells using AllPrep DNA/RNA mini kit (Qiagen, Canada) according to manufacture's protocol as previously described [24, 25].

Methylome assay was performed on the genomic DNA (gDNA) extracted from T cells using Infinium® MethylationEPIC BeadChip assay from Illumina (Illumina Inc., CA, USA). Following sodium bisulfite conversion of gDNA Infinium® MethylationEPIC BeadChip assay was carried out as previously described [24, 25].

Pyrosequencing was performed according to PyroMark Q24 (QIAGEN) following manufacturer's protocol except a few modifications according to Tost et al [24–26] All data were expressed as mean  $\pm$  standard error of the mean (SEM). Primers used for the analysis are listed in Supplementary Table 2.

### Amplicon sequencing (using Illumina MiSeq)

The bisulfite converted DNA was quantified by a Nanodrop Spectrophotometer (Thermo Fisher Scientific, MA, USA) followed by two rounds of polymerase chain reaction (PCR) targeting the cg14713996 and cg05133736 regions using Bio-Rad C1000 Touch Thermal Cycler (Bio-Rad Laboratories, CA, USA) for multiplex sequencing. The pooled library was then purified

twice using AMPure XP Beads (Beckman Coulter Life Sciences, CA, USA) and quantified by RealTime PCR using NEBNext® Library Quant Kit for Illumina (New England Biolabs, MA, USA) and sequencing was performed on the Illumina platform using MiSeq Reagent Nano Kit V2 (Illumina, CA, USA).

#### Data and statistical analysis

Data analysis was performed in R software using the ChAMP pipeline from Bioconductor at default setting and quality controlled using BMIQ and SVD while batch corrected with ComBat [27, 28]. Indeed, significant ( $P < 0.05$ ) variation arising from the Slide variable were completely removed after running the ComBat function (Supplementary Figure 2). We carried out further QC as previously described by us and others [24, 25]. Subsequently, Differentially Methylated Probes (DMPs) between PCa and negative-biopsy controls was run in ChAMP with Benjamini-Hochberg (BH) and False Discovery Rate (FDR) as  $< 0.05$ . We performed Pearson correlation between the normalized DNA methylation beta values and Gleason score using the Pearson correlation function in R and corrected for multiple testing using BH method (FDR  $< 0.05$ ). Next, we used an older small cohort (cohort 2) for which we had 11 control and 13 LGPCa T cell DNA samples (Supplementary Table 1). For this cohort, we had run Illumina Infinium Human Methylation 450 K BeadChip in the past but there were no significantly differentially methylated CGs between the groups after BH correction (FDR  $< 0.05$ ). However, several CG methylation levels were significantly correlated with the Gleason score. Hence, we combined the CGs in 450 K and 850 K cohorts that correlated with the Gleason score and plotted a Heatmap and Boxplot of these CGs.

We shortlisted differentially methylated CGs which had delta-beta value of  $\pm 10\%$  in between the groups and plotted a Heat-map. Then the genes associated with these CGs were subjected to a pathway analysis using ConsensusPath DB (with  $p < 0.01$  and  $q < 0.05$ ) [29]. The flow chart of experimental approaches are outlined in Supplementary Figure 1B. Heatmaps were generated using GENE-E of broad institute [30]. The statistical analyses were performed using the computing environment R version 3.4.4 and GraphPad Prism 6.0 (GraphPad Software Inc., California, US).

## Results

### Differentially methylated genes in T-cells of PCa patients are enriched in immune-related pathways

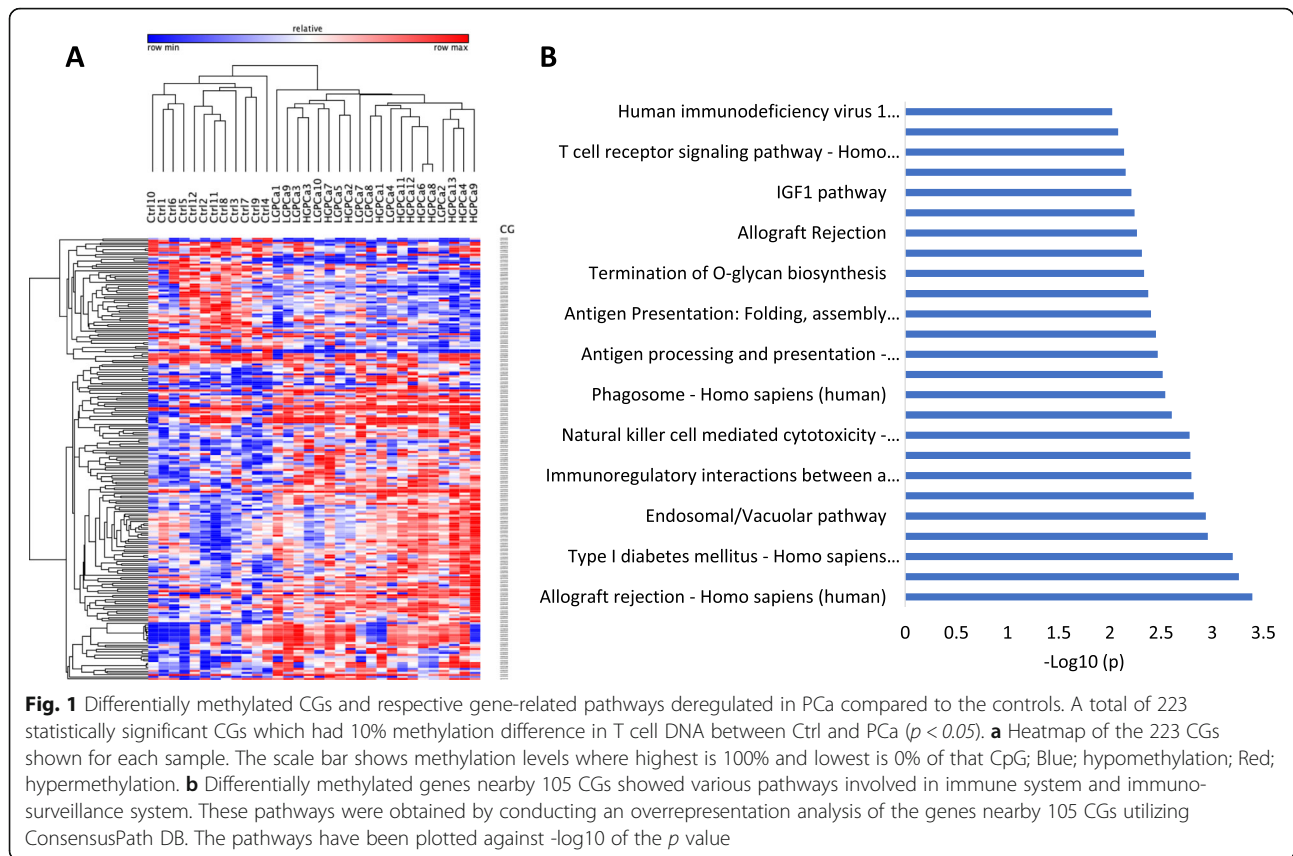
We delineated 223 differentially methylated CGs between the two groups of healthy individuals and PCa patients, using the Bioconductor package Limma, implemented in ChAMP [31, 32] with a delta beta threshold

of  $\pm 10\%$  (Supplementary File S1) and plotted a heatmap (Fig. 1a). The differentially methylated genes were subjected to pathway analysis. Notably, genes associated with differentially methylated CG sites were enriched in the pathways involved in immuno-editing and immune-surveillance systems (Fig. 1b).

### Correlation between DNA methylation levels and Gleason score in PCa

To delineate a set of differentially methylated CGs that are correlated with Gleason score progression, we performed a Pearson correlation analysis (Hmisc R) (0 for healthy individuals and 6, 8 and 9 for Gleason score 6, 8, 9). We corrected for multiple testing using the FDR (Q of  $< 0.05$ ) Benjamini-Hochberg method. This analysis revealed a signature of DNA methylation comprised of 449 CGs which correlates with Gleason scores in T cells from PCa patients (Fig. 2a-d). Four hundred sixteen CGs, out of 449 CGs, were hypermethylated (Fig. 2a and b) and 33 were hypomethylated (Fig. 2c and d). Boxplot of the methylation values of the differentially methylated sites included in the signature demonstrates the magnitude and progressive change in average DNA methylation as the Gleason score progresses from normal individuals to PCa patients with the highest differences observed at the highest Gleason score of 9 (Fig. 2a and c). Heatmap and hierarchical clustering analysis of 416 hypermethylated and 33 hypomethylated sites accurately grouped normal and PCa patients with Gleason score 6, 8 and 9 with exception of one PCa patient in Fig. 2b and two normal in Fig. 2d.

We next performed a Pearson analysis between methylation levels and Gleason scores of 354,692 CG sites that were also included in a 450 K Illumina bead array analysis of a different cohort (cohort 2) which we performed earlier (hereafter Combined cohort) (Supplementary Table 1). Eleven thousand seven hundred twenty-two CG sites showed a statistically significant correlation after FDR correction (Supplementary File S2); 8478 were hypermethylated and 3244 were hypomethylated. We shortlisted 1181 (Supplementary File S3) progressively methylated CGs that had a 10% difference in methylation level between average DNA methylation of PCa patients with Gleason score 9 and in normal individuals. Seven hundred forty-nine out of 1181 were hypermethylated and 432 hypomethylated (Fig. 2e-h). As demonstrated by the heatmap in Fig. 2f and h, the methylation profile correlates with increasing Gleason score from 6, 7, 8 to 9. We next overlapped 184 out of 449 differentially methylated CGs from EPIC (850 K) cohort with 11,722 differentially methylated in combined cohort (265 CGs were excluded as being specific to EPIC array CG list). This analysis resulted in 89 CGs that were



significantly correlated with Gleason score in both (hypergeometric,  $p = 1.56e-90$ ).

A genome-wide distribution on IGV browser [33] of 1181 differentially methylated CGs whose methylation changed progressively with Gleason score ( $p < 0.05$ ) is shown in Supplementary Figure 3. Most sites were hypermethylated, as Gleason score progress from 6 to 9, while a small fraction was hypomethylated. These data are consistent with our previous data in breast and liver cancer and support the hypothesis that epigenome-wide methylation changes occur in T cells genome as the cancer progress [24, 25]. Interestingly, there was significant overlap ( $p = 8.93e-42$ ) between the list of progressively changed methylation of CG sites as cancer progresses in prostate and liver, but not between prostate and breast or breast and liver (Supplementary Figure 4).

#### Differentially methylated regions (DMRs)

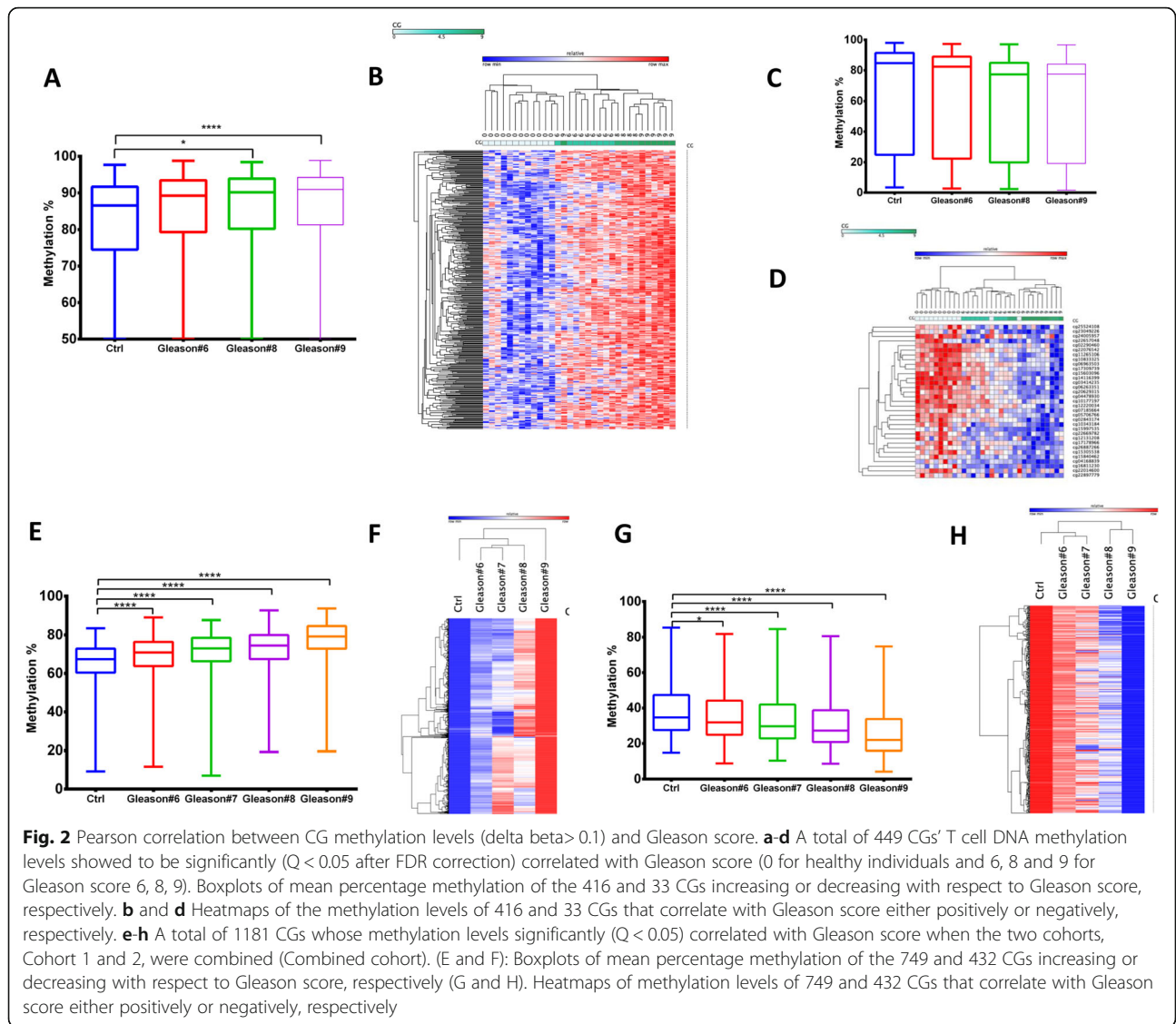
We analyzed differentially methylated regions (DMRs) utilizing the DMR function in ChAMP pipeline. We found 10 DMRs between the PCa patients and controls (Supplementary Table 3 and Supplementary Figure 5). We visualized these DMRs in IGV using the reference genome Human Hg19 to identify genes overlapping with or nearby these regions (Supplementary Table 3). The DMRs were present near genes at different sites

including transcriptions start sites (TSS), intergenic and intragenic regions. Pathway analysis on these genes showed only apoptosis pathway significantly different from control samples (data not shown). Although DMR function provides a good overview of DMRs around the genome, however, we were interested in CGs that could distinguish between healthy and PCa patients specifically. Thus, we moved to differentially methylated probes (DMPs) analysis in ChAMP pipeline.

#### Differentially methylated probes (DMPs) between PCa and normal progressively change with Gleason score

We next performed differentially methylation analysis between healthy individuals and PCa patients with both, high and low Gleason score (HGPCa and LGPCa) using stringent criteria of keeping BH correction at  $FDR < 0.05$ . This analysis resulted in three significant CGs (cg14713996, cg02701909, cg05133736) whose methylation levels discriminated between healthy controls and PCa patients (Table 1 and Fig. 3).

We then determined whether a polygenic methylation score of these 3 sites could serve as an accurate predictor of PCa. We calculated the regression coefficient of each CG and intercept using a multivariate linear regression model that included three methylation positions: polygenic methylation score (PMS) = ( $[\beta_1 * 27.184 +$

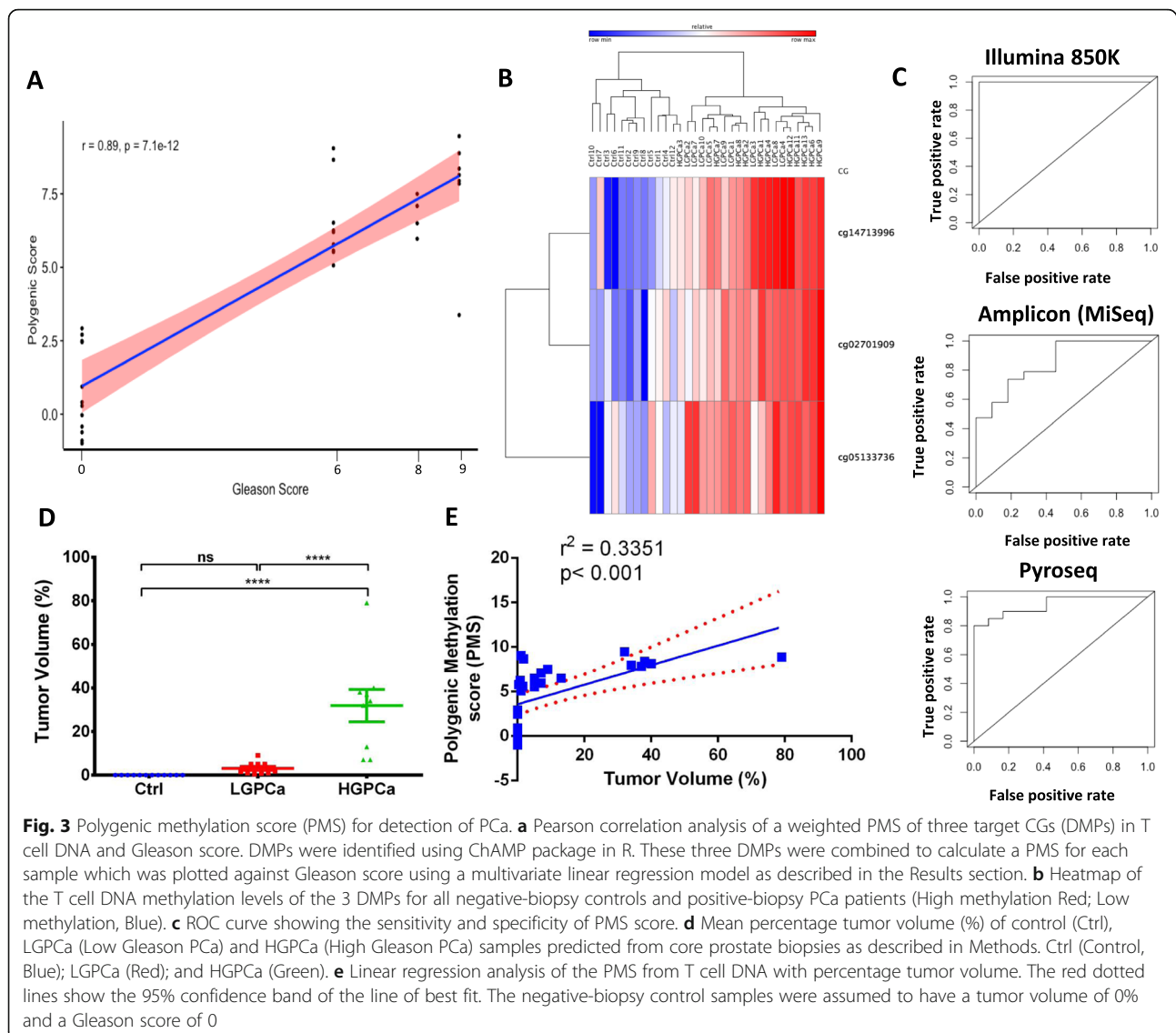


$\beta_2 * 28.842 + \beta_3 * 24.57 - 58.774$ ]; where  $\beta_1$  = methylation level of cg14713996,  $\beta_2$  = methylation level of cg02701909 and  $\beta_3$  = methylation level of cg05133736). This polyvariable linear regression equation, composed of weighted methylation level of three CG sites, was highly significant for the prediction of Gleason score in PCa patients ( $r = 0.89$ ,  $p = 7.1e^{-12}$ ) as demonstrated in Fig. 3a and b.

We generated receiver operating characteristic (ROC) curves and measured sensitivity, specificity and the area under the ROC curve (AUC) for each CG (cg14713996, cg02701909, cg05133736) and for PMS. Figure 3c shows ROC curves plotting the clinical performance of polygenic methylation markers for the detection of PCa. The combination of the 3 CGs (PMS) had higher sensitivity,

**Table 1** Differentially methylated probes (DMPs) between negative-biopsy controls and positive-biopsy PCa patients identified by Illumina EPIC methylation assay

CpG	Chromosome Location	Gene nearby	Delta Beta /LogFC	FDR	Site of differential methylation	CGI
cg14713996	X		0.12	0.01	IGR	opensea
cg02701909	1	CTRC	0.03	0.03	1stExon	opensea
cg05133736	10		0.08	0.04	IGR	opensea
cg06915331	11	PDE2A	0.03	0.12	Body	shore
cg01235591	13	ATP11A	0.03	0.12	3'UTR	shore



specificity and overall accuracy particularly using Illumina EPIC (Table 2 and Fig. 3c).

Percentage tumor volume, which is reflective of the potential cancer volume in the prostate gland of PCa patients, also showed a strong significant positive correlation with Gleason score (Supplementary Figure 6A;  $r^2 = 0.4032$ ;  $p < 0.0001$ ). However, the data points best fit was a non-linear exponential curve (Supplementary Figure 6B;  $r^2 = 0.8344$ ). Indeed, when stratified into healthy controls (Ctrl), LGPCa and HGPCa, the HGPCa showed significantly higher mean tumor volumes (%) than both the other groups (Fig. 3d). Moreover, PMS also showed strong significant positive correlation with tumor volume (Fig. 3e;  $r^2 = 0.3351$ ;  $p < 0.001$ ) indicating that PMS was able to predict aggressiveness of the tumor. Hence, a higher PMS in peripheral T cells' DNA

would potentially indicate a higher tumor volume in the prostate of that patient and vice versa.

#### Validation of differentially methylated cg14713996 and cg05133736

We next validated cg14713996 and cg05133736 using pyrosequencing and amplicon sequencing and due to insufficient amount of T cells DNA we couldn't validate cg02701909 for amplicon sequencing and some samples for pyrosequencing (Fig. 4, Fig. 5 and Supplementary Figure 7). The validation set included 20 PCa patients and 12 healthy controls.

cg14713996 showed a 12% ( $\pm 1.45\%$ ) difference in methylation between the controls and PCa in the Illumina EPIC (850 K) assay (Fig. 4a; Table 2). The methylation level of this site progressively changes

**Table 2** Sensitivity, specificity, accuracy and the area under the ROC curve (AUC) for each CG and for PMS, tested for Illumina EPIC (850 K), amplicon sequencing and pyrosequencing platforms

		Sensitivity	Specificity	Accuracy	AUC
<b>Illumina 850 K</b>	cg14713996	0.95	0.92	0.94	0.98
	cg02701909	0.95	0.92	0.94	0.98
	cg05133736	0.9	0.92	0.9	0.97
	PMS	1	1	1	1
<b>Pyrosequencing</b>	cg14713996	0.95	0.83	0.9	0.95
	cg02701909	na	na	na	na
	cg05133736	0.8	0.83	0.81	0.8
	PMS	0.85	0.92	0.87	0.95
<b>Amplicon sequencing</b>	cg14713996	0.79	0.82	0.8	0.83
	cg02701909	na	na	na	na
	cg05133736	0.85	0.66	0.78	0.73
	PMS	0.74	0.82	0.83	0.85

between controls, LGPCa and HGPCa (Fig. 4b). We examined correlation between Illumina EPIC array, amplicon sequencing and pyrosequencing assays ( $r^2 = 0.55$  and  $r^2 = 0.618$ , respectively,  $P < 0.0001$ ), as well as amplicon sequencing and pyrosequencing ( $r^2 = 0.5032$ ,  $P < 0.0001$ ) (Fig. 4c).

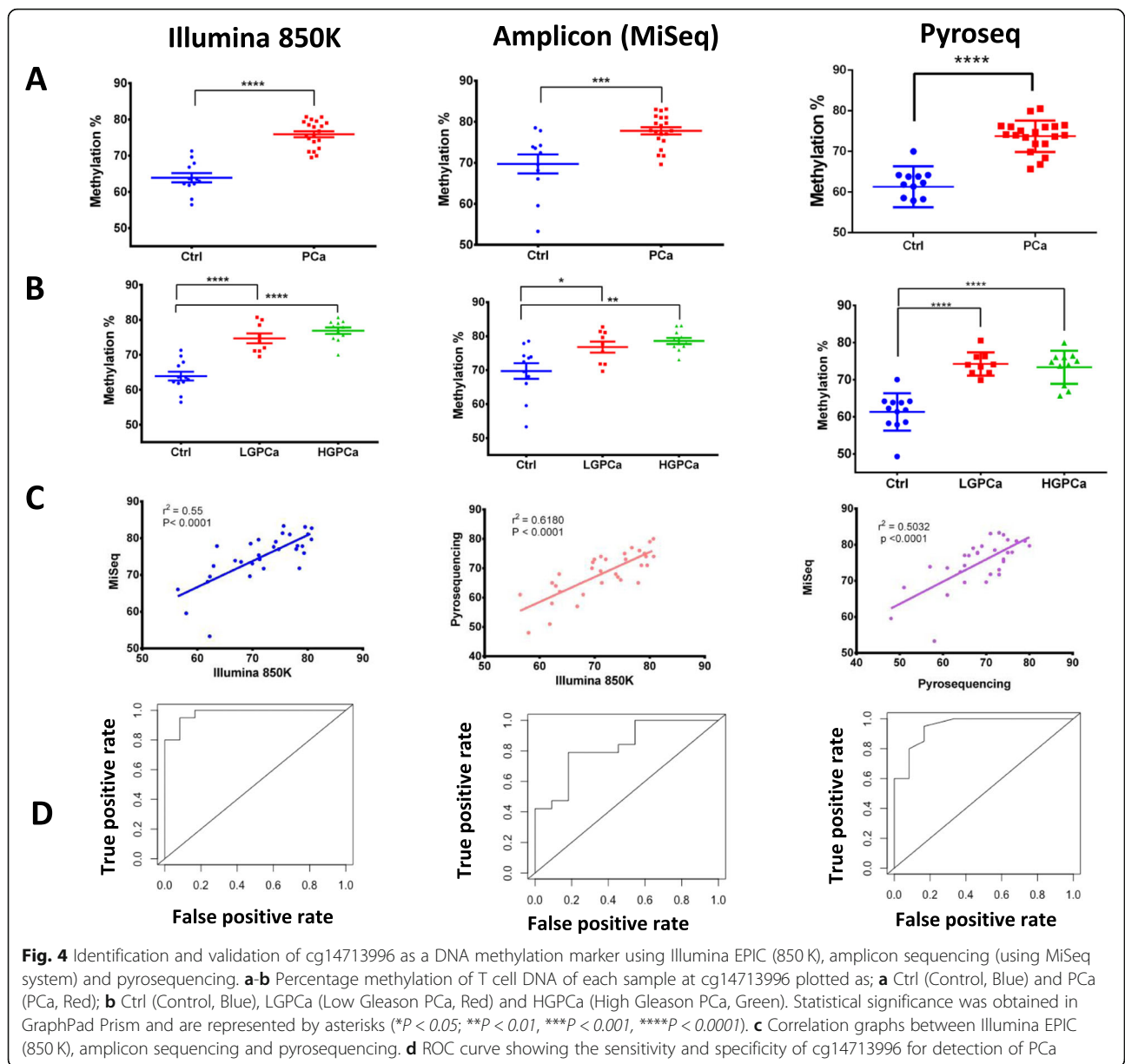
cg05133736 showed an 8% ( $\pm 1.15\%$ ) difference in DNA methylation between the controls and PCa in the Illumina EPIC (850 K) assay (Fig. 5a; Table 2). Validations and correlations were performed for cg05133736 as well (Fig. 5a-c). Lastly, both cg14713996 and cg05133736 had high sensitivity, specificity and overall accuracy particularly using Illumina EPIC (Table 2, and Figs. 4d and 5d, respectively).

## Discussion

DNA methylation alterations in leukocytes have been reported by us and others in many cancers including head and neck squamous cell carcinoma, ovarian [34, 35], colorectal [36], breast [24], and HCC [25, 37]. We focused on T cells which are essential for controlling tumorigenesis and epigenetic changes like DNA methylation which are likely to occur first in T cell as they are activated even before malignant transformation. Detection of these approaches which are highly sensitive, specific and non-invasive will allow stratification of men for the risk of developing or having PCa [24, 25]. In the current study, identified genes and CGs were enriched in signaling pathways implicated in immune-editing and -surveillance which allowed differentiation of control and PCa subjects, and also PCa subjects with various Gleason scores. Consistent with immune-editing theory, these results provide additional support that the DNA methylation signature can be utilized in assessing risk of early stages of PCa.

Previous studies have shown differential methylation of individual gene promoters in PCa. A study investigating clinical significance of gene promoter DNA methylation in blood DNA found six genes promoter methylation of PCa patients to be significantly different from controls [38]. However, we did not see such differences which could be due to several reasons; firstly, T cells DNA was used in our study while blood DNA was used in previous study [38]; secondly, previous study was carried out in African American men while our study had mostly Caucasian men [38]. Also, using whole blood DNA, no association between the DNA methylation on all CG sites and risk of PCa or aggressive disease was observed [39]. Similarly, in a small cohort, we also did not find any CG site with differential methylation between controls and PCa patients T cell DNA (data not shown) which could be due to small sample size and use of Illumina 450 K which does not include FANTOM5 and ENCODE enhancer regions and many other sites that were added in EPIC (850 K) array [40]. Our previous studies in breast and liver cancer validated the use of T cells to distinguish between control and cancer and also differentiated between different disease stages [24, 25]. Similarly, our study is not only consistent in distinguishing between negative and positive biopsy for PCa but also according to Gleason score.

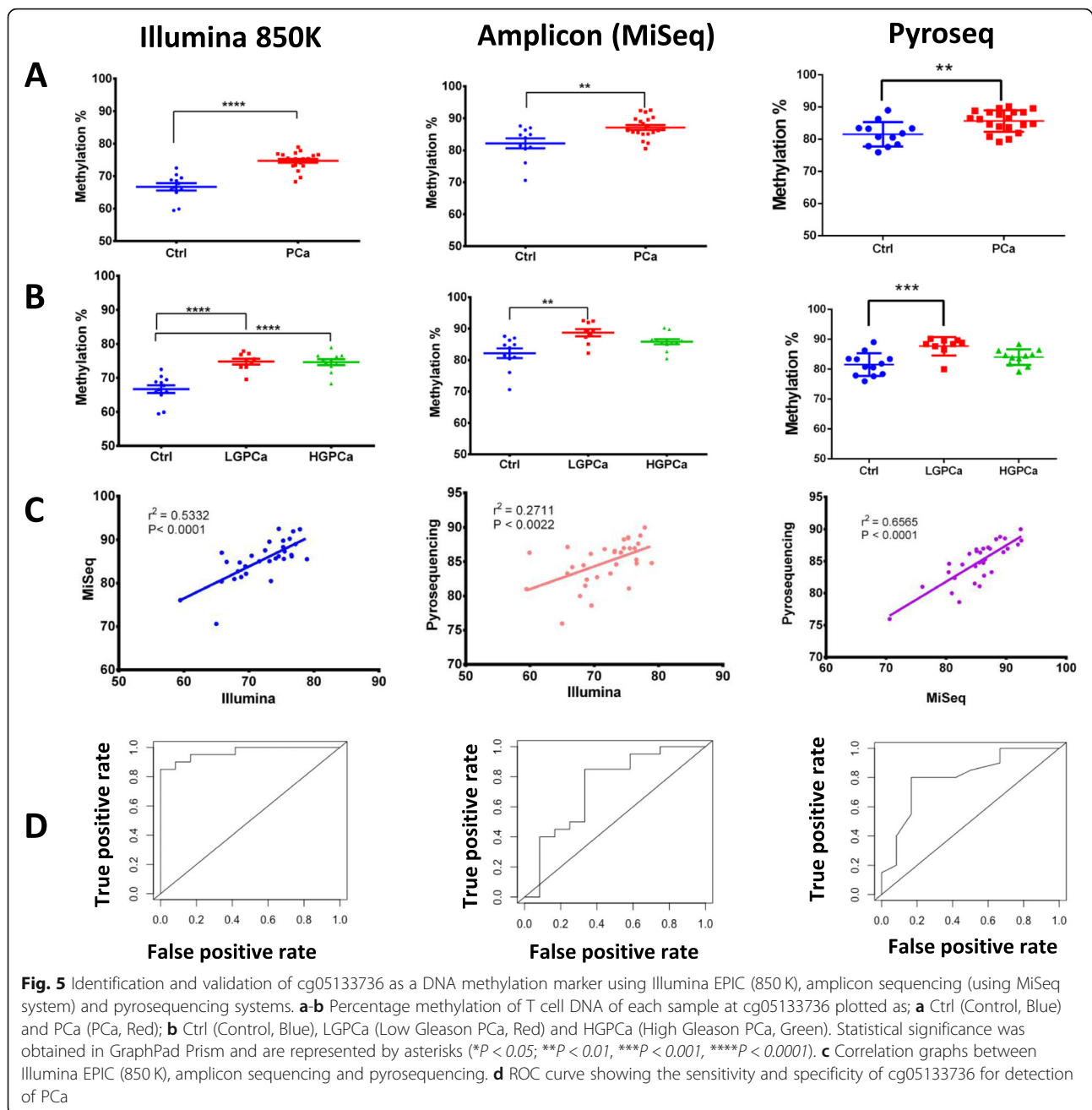
Despite these convincing results we are aware of the limitation of our study due to the small sample size. However, these results are robust in differentiating normal and PCa subjects, identifying several CGs which were validated using amplicon sequencing and pyrosequencing [26, 27, 41–43]. In addition, our previous study involving T cells from breast cancer had smaller sample size of 19 samples, but was able to distinguish between healthy controls and cancer patients [24]. Our DNA methylation signature gets stronger between controls and PCa patients across various Gleason scores and the highest difference is seen between



Gleason score of 9 and control subjects. Importantly, the PMS was able to not only predict PCa patients from healthy controls, but also Gleason score indicating progressive stages of PCa with strong predictive power of PMS with high sensitivity, specificity and accuracy (Fig. 3c and Table 2). Indeed, the PMS in the T cells' DNA showed strong correlation with tumor volume of PCa patients that is strongly associated with Gleason score as showed herein and previously (Fig. 3 and Supplementary Figure 6). While all subjects had no history of immune disorders or infection, due to the samples size we were unable to take into account additional confounding states including smoking and alcohol intake which can affect the state of DNA methylation [44–48].

Results of this study are the first to identify the DNA methylation signature in immune cells of PCa subjects using non-invasive approach for early detection and stratification of PCa. It is imperative to carry out follow up studies using a larger cohort of subjects which will also consider history of smoking, alcohol intake and stratify the subjects based on their racial background. Future studies will also monitor differences among different T cells, PCa sub-types, stages and Gleason score. It is anticipated that the T cell DNA methylation signature will be identifiable in whole blood for large scale screens, active surveillance, early intervention with targeted therapies including the use of epigenetic based therapeutic agents.





## Conclusions

The current study provides for the first-time a potential role of DNA methylation changes in peripheral T cells in PCa. Using this non-invasive approach, changes in the DNA methylation status of T cells can predict men at risk of developing PCa and disease progression. Furthermore, the non-invasive methodology may allow for early intervention and stratification of patients into different prognostic groups to reduce PCa associated morbidity from repeat invasive prostate biopsies and design therapeutic strategy to reduce PCa associated mortality.

## Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12885-020-07078-8>.

**Additional file 1: Supplementary Figure 1.** (A) The number of T cell DNA samples used in the cohort 1. Ctrl, Negative-biopsy healthy control; LGPCa, Positive-biopsy low Gleason PCa; HGPCa, Positive biopsy high Gleason PCa. (B) Outline of DNA methylation signature determination in DNA of peripheral T cells in buffy coat of PCa patients and healthy controls using Illumina Infinium EPIC methylation array followed by validation with Illumina MiSeq and Pyrosequencing platforms. **Supplementary Figure 2.** Quality control (QC) of the data from Illumina Infinium EPIC methylation array using QC in ChAMP pipeline. (A) Density plot of raw

data showing the beta distributions for each healthy control and PCa sample. None of the samples deviate from each other. (B) Singular value decomposition analysis (SVD) plot shows the significant components of variation in T cell DNA methylation. SVD analysis before (Pre) and after (Post) running the ComBat function in ChAMP pipeline removes the batch effects such as "Slide" variations as shown here. **Supplementary Figure 3.** Global distribution of the differentially methylated CGs that showed correlation with Gleason score in combined cohort. 1181 significant ( $p < 0.05$ ) differentially methylated CGs that had 10% methylation difference (beta value) in T cell DNA between PCa patients (with Gleason score 9) and in healthy controls were taken and input into IGV browser. Each row represents a Gleason score in a sequence of 6, 7, 8 and 9 while the first row is Chromosome numbers (1-22, X and Y). Hypomethylation (Blue) and hypermethylation (Red). **Supplementary Figure 4.** Venn diagram showing the overlap of CGs between breast, liver and prostate cancer studies. Methylated CGs in T cell DNA that were correlating with Gleason score in this study were overlapped with CGs that were correlating with breast cancer and liver cancer progression in breast and liver cancer studies, respectively. ns, not significant; \*\*\*\*,  $p < 0.0001$ . **Supplementary Figure 5.** Global distribution of the differentially methylated regions (DMRs). DMRs of the T cell DNA between the PCa and healthy control obtained using DMR function in the ChAMP pipeline. The data for the IGV input was obtained from Supplementary Table 3. The first row represents Chromosome numbers (1-22, X and Y), second row shows DMR positions and the last row shows the reference genome H19. **Supplementary Figure 6.** Correlation analysis of percentage tumor volume and Gleason score. The percentage tumor volume was predicted from core prostate biopsies as described in Methods. Negative-biopsy healthy controls had no tumor and therefore assigned a 0% tumor volume and a Gleason score of 0. (A) Linear regression analysis of the percentage tumor volume of the PCa patients with the Gleason score. The red dotted lines show the 95% confidence band of the line of best fit. (B) Non-linear regression analysis of the percentage tumor volume of the PCa patients with Gleason score. The equation for the graph was an exponential growth curve with least square method as the fitting method. **Supplementary Figure 7.** Identification and validation of cg02701909 as a DNA methylation marker using Illumina EPIC (850K) and pyrosequencing. (A-B) Percentage methylation of T cell DNA at cg02701909 plotted as; (A) Ctrl (Control, Blue) and PCa (PCa, Red); (B) Ctrl (Control, Blue), LGPCa (Low Gleason PCa, Red) and HGPCa (High Gleason PCa, Green). The number of samples for pyrosequencing were; Ctrl ( $n = 9$ ), LGPCa ( $n = 9$ ) and HGPCa ( $n = 9$ ). Statistical significance was obtained in GraphPad Prism and are represented by asterisks (\* $P < 0.05$ ; \*\* $P < 0.01$ , \*\*\* $P < 0.001$ , \*\*\*\* $P < 0.0001$ ). (C) Correlation graph between Illumina EPIC (850K) and pyrosequencing.

**Additional file 2: Supplementary Table 1.** List of negative-biopsy healthy controls and positive-biopsy PCa patient sample ids with their respective Gleason score.

**Additional file 3: Supplementary Table 2.** List of the primers used for pyrosequencing and amplicon sequencing (Illumina MiSeq system).

**Additional file 4: Supplementary Table 3.** Differentially methylated regions (DMRs) between the PCa T cell DNA and healthy control T cell DNA. The data was obtained using DMR function in the ChAMP pipeline and input into the IGV browser to identify nearby genes and location of each DMR manually. Reference genome H19 was used for alignment. TSS, Transcription start site.

**Additional file 5.**

**Additional file 6.**

**Additional file 7.**

## Abbreviations

PCa: Prostate Cancer; LGPCa: Low Gleason Prostate Cancer; HGPCa: High Gleason Prostate Cancer; ChAMP: Chip Analysis Methylation Pipeline; DMP: Differentially methylated probes; BH: Benjamini-Hochberg; HCC: Hepatocellular carcinoma; PCR: Polymerase Chain Reaction; BMIQ: Beta-Mixture Quantile; SVD: Singular Value Decomposition; FDR: False Discovery Rate; IGV: Integrative Genomics Viewer; ROC: Receiver Operating Characteristic; AUC: Area Under Curve; FANTOM5: Functional Annotation of

the Mammalian Genome 5; ENCODE: The Encyclopedia of DNA Elements; PMS: Polygenic Methylation Score

## Acknowledgements

Ali Mehdi is the recipient of Queen Elizabeth II scholarship.

## Authors' contributions

SAR and MS conceived the study and experimental design. AM and AA carried out experimental procedures. All data analysis was carried out by DC and AM. AM wrote the manuscript with SAR and MS. TAB provided access to samples and assisted in manuscript revisions. All authors read and approved the final manuscript.

## Funding

This work was supported by grants from the Canadian Institutes for Health Research (PJT-156225) to SAR. There was no role of funding agency in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

## Availability of data and materials

All data generated or analysed during this study are included in this published article [and its supplementary information files].

## Ethics approval and consent to participate

The study design was approved by the ethics committee of Alberta PCa Research Initiative (APCaRI) and informed consents were obtained in writing from all subjects for the APCaRI.

## Consent for publication

Not applicable.

## Competing interests

MS and DC are co-founders of Montreal Epiterapia Inc. TAB is an Editorial Board member for the journal. The remaining authors declare that they have no conflict of interest.

## Author details

<sup>1</sup>Department of Medicine, McGill University, Montreal, Quebec, Canada.

<sup>2</sup>Department of Human Genetics, McGill University, Montreal, Quebec, Canada.

<sup>3</sup>HKG Epitherapeutics, Hong Kong, China. <sup>4</sup>Department of Oncology, McGill University, Montreal, Quebec, Canada.

<sup>5</sup>Departments of Pathology & Laboratory Medicine, Oncology, Biochemistry & Molecular Biology, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada.

<sup>6</sup>Department of Pharmacology, McGill University, Montreal, Quebec, Canada.

<sup>7</sup>McGill University Health Centre, 1001 Décarie Blvd. (Glen site), Room

EM1.3232, Montréal, QC H4A3J1, Canada.

Received: 12 July 2019 Accepted: 15 June 2020

Published online: 23 June 2020

## References

1. Society CC. Prostate cancer statistics - Canadian Cancer Society. Published 2020. <http://www.cancer.ca/en/cancer-information/cancer-type/prostate/statistics/?region=qc>. Accessed 1 Apr 2020.
2. Research AlFC. Prostate cancer statistics. Published 2018. <https://www.wcrf.org/dietandcancer/cancer-trends/prostate-cancer-statistics>. Updated 2018-08-22. Accessed 1 Apr 2020.
3. Gandhi J, Afridi A, Vatsia S, et al. The molecular biology of prostate cancer: current understanding and clinical implications. *Prostate Cancer Prostatic Dis.* 2018;21(1):22–36.
4. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2017. *CA Cancer J Clin.* 2018; 67(1):7–30.
5. Litwin MS, Tan HJ. The diagnosis and treatment of prostate Cancer: a review. *JAMA.* 2017;317(24):2532–42.
6. Carter HB, Albertsen PC, Barry MJ, et al. Early detection of prostate cancer: AUA guideline. *J Urol.* 2013;190(2):419–26.
7. Lilja H, Ulmert D, Vickers AJ. Prostate-specific antigen and prostate cancer: prediction, detection and monitoring. *Nat Rev Cancer.* 2008;8(4):268–78.
8. 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.

9. Stabile A, Giganti F, Emberton M, Moore CM. MRI in prostate cancer diagnosis: do we need to add standard sampling? A review of the last 5 years. *Prostate Cancer Prostatic Dis.* 2018;21(4):473–87.
10. Bird A. The essentials of DNA methylation. *Cell.* 1992;70(1):5–8.
11. Baylin SB, Herman JG, Graff JR, Vertino PM, Issa JP. Alterations in DNA methylation: a fundamental aspect of neoplasia. *Adv Cancer Res.* 1998;72:141–96.
12. Hoque MO. DNA methylation changes in prostate cancer: current developments and future clinical implementation. *Expert Rev Mol Diagn.* 2009;9(3):243–57.
13. Merlo A, Herman JG, Mao L, et al. 5' CpG island methylation is associated with transcriptional silencing of the tumour suppressor p16/CDKN2/MTS1 in human cancers. *Nat Med.* 1995;1(7):686–92.
14. Hoque MO, Topaloglu O, Begum S, et al. Quantitative methylation-specific polymerase chain reaction gene patterns in urine sediment distinguish prostate cancer patients from control subjects. *J Clin Oncol.* 2005;23(27):6569–75.
15. Cooper CS, Foster CS. Concepts of epigenetics in prostate cancer development. *Br J Cancer.* 2009;100(2):240–5.
16. Nelson WG, Yegnasubramanian S, Agoston AT, et al. Abnormal DNA methylation, epigenetics, and prostate cancer. *Front Biosci.* 2007;12:4254–66.
17. Dobosy JR, Roberts JL, Fu VX, Jarrard DF. The expanding role of epigenetics in the development, diagnosis and treatment of prostate cancer and benign prostatic hyperplasia. *J Urol.* 2007;177(3):822–31.
18. Li L, Choi JY, Lee KM, et al. DNA methylation in peripheral blood: a potential biomarker for Cancer molecular epidemiology. *J Epidemiol.* 2012;22(5):384–94.
19. Kirby MK, Ramaker RC, Roberts BS, 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.
20. Dunn GP, Old LJ, Schreiber RD. The immunobiology of cancer immunosurveillance and immunoediting. *Immunity.* 2004;21(2):137–48.
21. Smyth MJ, Dunn GP, Schreiber RD. Cancer immunosurveillance and immunoediting: the roles of immunity in suppressing tumor development and shaping tumor immunogenicity. *Adv Immunol.* 2006;90:1–50.
22. Finn OJ. A believer's overview of cancer immunosurveillance and immunotherapy. *J Immunol.* 2018;200(2):385–91.
23. Gubin MM, Artyomov MN, Mardis ER, Schreiber RD. Tumor neoantigens: building a framework for personalized cancer immunotherapy. *J Clin Invest.* 2015;125(9):3413–21.
24. Parashar S, Cheishvili D, Mahmood N, et al. DNA methylation signatures of breast cancer in peripheral T-cells. *BMC Cancer.* 2018;18(1):574.
25. Zhang Y, Petropoulos S, Liu J, et al. The signature of liver cancer in immune cells DNA methylation. *Clin Epigenetics.* 2018;10:8.
26. Tost J, Gut IG. DNA methylation analysis by pyrosequencing. *Nat Protoc.* 2007;2(9):2265–75.
27. Morris TJ, Butcher LM, Feber A, et al. ChAMP: 450k Chip Analysis Methylation Pipeline. *Bioinformatics.* 2014;30(3):428–30.
28. Tian Y, Morris TJ, Webster AP, Yang Z, Feber A, Teschendorff AE. ChAMP: updated methylation analysis pipeline for Illumina BeadChips. *Bioinformatics.* 2017;33(24):3982–4.
29. Kamburov A, Pentchev K, Galicka H, Wierling C, Lehrach H, Herwig R. ConsensusPathDB: toward a more complete picture of cell biology. *Nucleic Acids Res.* 2011;39(Database issue):D712–7.
30. Institute B. GENE-E. <https://software.broadinstitute.org/GENE-E/index.html>. Published 2019. Accessed 1 May 2019.
31. Smyth GK. Limma: linear models for microarray data. In: Gentleman R, Dudoit S, Irizarry R, Huber W, editors. *Bioinformatics and computational biology solutions using R and bioconductor*, vol. 1. New York: Springer; 2005. p. 397–420.
32. Smyth GK, Michaud J, Scott HS. Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics.* 2005;21(9):2067–75.
33. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform.* 2013;14(2):178–92.
34. Fridley BL, Armasu SM, Cicek MS, et al. Methylation of leukocyte DNA and ovarian cancer: relationships with disease status and outcome. *BMC Med Genet.* 2014;7:21.
35. Koestler DC, Marsit CJ, Christensen BC, et al. Peripheral blood immune cell methylation profiles are associated with nonhematopoietic cancers. *Cancer Epidemiol Biomarkers Prev.* 2012;21(8):1293–302.
36. Noshok K, Baba Y, Tanaka N, et al. Tumour-infiltrating T-cell subsets, molecular changes in colorectal cancer, and prognosis: cohort study and literature review. *J Pathol.* 2010;222(4):350–66.
37. Kao WY, Yang SH, Liu WJ, et al. Genome-wide identification of blood DNA methylation patterns associated with early-onset hepatocellular carcinoma development in hepatitis B carriers. *Mol Carcinog.* 2017;56(2):425–35.
38. Moses-Fynn E, Tang W, Beyene D, et al. Correlating blood-based DNA methylation markers and prostate cancer risk in African-American men. *PLoS One.* 2018;13(9):e0203322.
39. FitzGerald LM, Naeem H, Makalic E, et al. Genome-wide measures of peripheral blood DNA methylation and prostate cancer risk in a prospective nested case-control study. *Prostate.* 2017;77(5):471–8.
40. Moran S, Arribas C, Esteller M. Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. *Epigenomics.* 2016;8(3):389–99.
41. Royo JL, Hidalgo M, Ruiz A. Pyrosequencing protocol using a universal biotinylated primer for mutation detection and SNP genotyping. *Nat Protoc.* 2007;2(7):1734–9.
42. Ronaghi M, Uhlen M, Nyren P. A sequencing method based on real-time pyrophosphate. *Science.* 1998;281(5375):363–5.
43. Teschendorff AE, Marabita F, Lechner M, et al. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics.* 2013;29(2):189–96.
44. Wilson LE, et al. Alcohol and DNA methylation: an epigenome-wide association study in blood and normal breast tissue. *Am J Epidemiol.* 2019;188:1055–65 *Epidemiology Branch NIEHS, Research Triangle Park, North Carolina, Department of Population Health Sciences DUSoM, Durham, North Carolina.*
45. Varela-Rey M, Woodhoo A, Martinez-Chantar ML, Mato JM, Lu SC. Alcohol, DNA methylation, and Cancer. *Alcohol Res.* 2013;35(1):25–35.
46. Liu C, Marioni RE, Hedman ÅK, et al. A DNA methylation biomarker of alcohol consumption. *Mol Psychiatry.* 2016;23(2):422.
47. Lee KW, Pausova Z. Cigarette smoking and DNA methylation. *Front Genet.* 2013;4:132.
48. Li S, Wong EM, Bui M, et al. Causal effect of smoking on DNA methylation in peripheral blood: a twin and family study. *Clin Epigenetics.* 2018;10:18.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

