Title: A three gene DNA methylation biomarker accurately classifies early stage prostate cancer

Palak G. Patel,1,2 Thomas Wessel,3 Atsunari Kawashima,1,2,4 John B. A. Okello,1,2,5 Tamara Jamaspishvili,1,2 Karl-Philippe Guérard,6 Laura Lee,7 Anna Ying-Wah Lee,7 Nathan E. How,1,2 Dan Dion,7 Eleonora Scarlata,6 Chelsea L. Jackson,1,2 Suzanne Boursalie,1,2 Tanya Sack,1,2 Rachel Dunn,1,2 Madeleine Moussa,8 Karen Mackie,8 Audrey Ellis,8 Elizabeth Marra,8 Joseph Chin,8 Khurram Siddiqui,8 Khalil Hetou,8 Lee-Anne Pickard,9 Vinolia ArthurHayward,9 Glenn Bauman,10 Simone Chevalier,6 Fadi Brimo,11 Paul Boutros,7 Jacques Lapointe,6 John M.S. Bartlett,12 Robert J. Gooding,2,13 and David M. Berman,1,2

1Department of Pathology & Molecular Medicine, Queen's University, Kingston, ON, Canada
2Division of Cancer Biology & Genetics, Queen's Cancer Research Institute, Queen's University, Kingston, ON, Canada
3Life Sciences Group, ThermoFisher Scientific, Waltham, MA, US
4Department of Urology, Graduate School of Medicine, Osaka University, Suita, Osaka, Japan
5Cardiac Genome Clinic, Ted Rogers Centre for Heart Research, Peter Gilgan Centre for Research and Learning, The Hospital for Sick Children, Toronto, ON, Canada
6Division of Urology, Department of Surgery, McGill University and the Research Institute of the McGill University Health Centre, Montreal, QC, Canada
7Ontario Institute for Cancer Research (OICR), Toronto, ON, Canada
8Department of Pathology and Laboratory Medicine, Division of Surgical Pathology, London Health Sciences Centre, London, ON, Canada
9Ontario Tumor Bank, Toronto, ON, Canada
10Division of Radiation Oncology, London Regional Cancer Program, London Health Sciences Centre, London, Ontario, Canada; Department of Physics and Astronomy, University of Western Ontario, London, Ontario, Canada
11Department of Pathology, McGill University Health Center and McGill University, Montreal, QC, Canada
12Diagnostic Development, Ontario Institute for Cancer Research (OICR), Toronto, ON, Canada
13Department of Physics, Engineering Physics & Astronomy, Queen’s University, Kingston, ON, Canada

Running title: Three gene DNA methylation biomarker for prostate cancer.

Keywords: DNA methylation, prostate cancer diagnosis, biomarker discovery and validation, cancer epigenetics.

Corresponding Author: David M. Berman

10 Stuart St., QCRI 302A
Kingston, ON K7L 3N6
1- 613-533-3022
Abstract

**Background:** To identify and validate accurate diagnostic biomarkers for prostate cancer through systematic evaluation of DNA methylation alterations.

**Materials and Methods:** We constructed three early prostate cancer cohorts (total patients = 699) from which we collected and processed over 1300 prostatectomy tissue samples for DNA extraction. Using real-time methylation-specific PCR (qMSP), we measured normalized methylation levels at 15 frequently methylated loci. After partitioning sample sets into independent training and validation cohorts, classifiers were developed using logistic regression, analyzed, and validated.

**Results:** In the training dataset, DNA methylation levels at seven of fifteen genomic loci (GSTP1, CCDC181, HAPLN3, GSTM2, GAS6, RASSF1, and APC) showed large differences between cancer and benign samples. The best binary classifier was the GAS6/GSTP1/HAPLN3 logistic regression model, with an AUC of 97%, which showed a sensitivity of 94%, and a specificity of 93% after external validation.

**Conclusion:** We created and validated a multi-gene model for the classification of benign and malignant prostate tissue. With false positive and negative rates below 7%, this 3 gene biomarker represents a promising basis for more accurate prostate cancer diagnosis.
Introduction

*Diagnostic challenges in Prostate Cancer*

Its high prevalence and low risk for progression have complicated efforts to screen and manage prostate cancer. Serum prostate-specific antigen (PSA) testing is the most commonly used tool to identify men suspected of harboring prostate cancer. Patients with elevated PSA levels are typically referred for biopsy testing for definitive diagnosis. With a false positive rate of >75% and positive predictive value of ~25%, PSA results are most often inconclusive. Accordingly, each year in United States alone, approximately 4.7 million men are identified with elevated PSA levels, and 1.3 million of these men are recommended to undergo biopsy for prostate cancer. However, 975,000 prostate biopsy results are negative and can be considered unnecessary, which exposes patients to complications such as infections, bleeding and thousands of hospitalizations.

In an effort to identify a more accurate non-invasive biomarker than PSA, numerous studies have investigated the use of mRNA, microRNA, prostate-specific proteins, and genetic mutations as biomarkers. However, prostate cancers possess few informative biomarkers in these categories. Only a handful of biomarkers (i.e. PSA, PCA3, TMPRSS2-ERG gene fusions) are currently available for diagnostic use in prostate cancer. Unfortunately, these tests exhibit low balanced accuracy, with high false positive or false negative rates. Pathologically evaluated prostatectomy tissue can serve as the gold-standard for investigating the maximum possible accuracy of a diagnostic biomarker. But to our knowledge, no study has optimized and validated a biomarker for prostate cancer detection in tissue.

*Methylation in Prostate Cancer*

Aberrant DNA methylation may be a superior substrate for biomarker discovery in early prostate cancer. In contrast to other genomic abnormalities, cancer-specific DNA methylation alterations are highly prevalent in prostate cancer, making them a sustained focus of research, with growing evidence supporting their role in progression and risk stratification. Some investigators study therapeutic strategies based on methylation inhibitors, others are working to develop DNA methylation alterations as useful diagnostic biomarkers. To date, most efforts in this area have focused broadly on describing the epigenetic landscape of prostate cancer. Notably, Yegnasubramanian found that GSTP1, APC, RASSF1A, PTGS2, and ABCBJ loci were hypermethylated in >85% of cancers. Haldrup et al. added AOX1, CCDC181, GAS6, HAPLN3, KLF8, and MOB3B loci as exhibiting cancer-specific hypermethylation and association with biochemical recurrence, and Vanaja, Mahapatra and others identified additional DNA methylation alterations associated with recurrence or risk of progression using genome-wide approaches.

Most prior studies have relied on vast sets of genes tested on comparatively few samples. Few studies have validated their results or generated methylation-based classifiers for clinically important outcomes. Exceptions include ConfirmMDx, a DNA methylation-based test performed on negative biopsy tissues to address the risk of finding cancer on a subsequent biopsy.
multicenter validation with 848 patients, the test yielded a high negative predictive value (89%), but limited specificity (64%) and sensitivity (65%).\textsuperscript{32,33} Similarly, Haldrup et al.\textsuperscript{21} developed a three gene DNA methylation classifier for risk of treatment failure (biochemical recurrence after prostatectomy) using a training cohort of archival samples from 293 patients which was validated on an independent cohort of fresh frozen samples from 114 prostatectomy patients, achieving a hazard ratio of 2.33.\textsuperscript{21} We are not aware of any validated DNA methylation classifiers that are designed to directly detect or diagnose prostate cancer or to help select patients for biopsies.

The goal of this project was to determine the maximum possible accuracy of a methylation-based biomarker in classifying prostate cancer. With superior performance to single gene tests, compact multi-gene biomarkers are ideal for clinical implementation.\textsuperscript{32,34–37} Compact tests overcome hurdles associated with genome-wide testing, including high cost and challenges with validation and analysis across thousands of data elements.\textsuperscript{38–40} To build a compact and reproducible diagnostic classifier, we chose real-time methylation-specific PCR (qMSP), a highly sensitive and cost-effective assay platform.\textsuperscript{41–44} We prioritized and selected 15 DNA methylation alterations for having been individually validated in multiple previous reports (Supplementary Tables S1 and S3). In this study, we tested each locus alone and in optimal combinations to determine its accuracy in classifying pathologically reviewed prostatectomy tissue as benign or cancer. As “ground truth,” we profiled benign and cancer tissue samples from 699 prostatectomy cases using real-time methylation-specific PCR (qMSP), a highly sensitive and cost-effective assay for quantitative DNA methylation analysis.\textsuperscript{41,43,44} For statistical power, DNA methylation profiling data from over 1250 cancer and 96 benign samples were divided into independent training and validation cohorts. Using this data, we constructed and validated a highly sensitive and specific classifier for detecting prostate cancer in tissue. Future studies will test its performance in non-invasive settings (i.e., urine, blood).

### Materials and Methods

#### Patient material

As part of a larger genomic profiling study, three patient cohorts were analyzed. They comprised of consecutive radical prostatectomies (RP) performed with curative intent for histologically verified, clinically localized prostate cancer (Table 1). Cohorts were obtained from Queen’s University/Kingston General Hospital (2000 - 2012), McGill University/Montreal General Hospital (1994 - 2013) and London Health Science Centre (LHSC) (2003 - 2009). In total, 699 patients were included in this study.

Using a previously published protocol \textsuperscript{45} we macrodissected and extracted DNA from index tumour foci from 699 RP cases and contralateral benign tissue (at least 5mm away from tumor foci) from 96 of those cases. Multiple samples were collected from each case, yielding over 1300 tissue samples. DNA was quantified on a Qubit 3.0 Fluorometer (ThermoFisher Scientific) using
the dsDNA HS (High Sensitivity) kit. A summary of final sample numbers for each DNA methylation assay is shown in Supplementary Table S2.

**Real-time Methylation-specific PCR (qMSP) analysis**

By searching PubMeth and PubMed databases, we identified 77 loci that were frequently hypermethylated in prostate cancer. Of these, we prioritized 24 that were documented in multiple studies containing >250 cases/samples, further signifying their potential reliability. ALDH1A2 and GSTM2 were added on the basis of preliminary data (N. How, not shown). We were able to design robust qMSP assays targeting 15 of 26 loci (Supplementary Table S3), and the rest were omitted from the study. An assay targeting ALU repeat elements was used as the reference control and distilled water was used as a negative control.

We modified previously described protocols for real time MSP (qMSP) assays according to MIQE guidelines and quantified changes in 15 DNA methylation alterations (Supplementary Table S3) in DNA samples collected from three RP cohorts. Briefly, individual DNA samples (50 ng) were bisulfite converted according to the manufacturer's protocol (EpiTect Bisulfite Kit, Qiagen). A mastermix was prepared that contained one of 15 primer pairs (400 nM; ThermoFisher Scientific) and probe sets (150 nM; ThermoFisher Scientific) (Supplementary Table S3), nucleotides (250 μM; Invitrogen), MgCl₂ (1.2 mM; NEB), BSA (0.5 mg/mL; NEB), ROX reference dye (24.5 nM; Invitrogen), EpiMark Taq polymerase (0.25 U; NEB) and 1X EpiMark reaction buffer (NEB). Next, Bisulfite-converted DNA (1 μL) was added to the mastermix and 10 uL reactions were amplified using a VIIA7 thermocycler (Applied Biosystems). Cycling conditions included denaturation at 95°C for 30 s, 7 cycles of touch-down PCR with annealing temperatures decreasing by 2°C per cycle and extension at 68°C for 30 s, followed by 48 cycles of 30 s at 95°C, 30 s at 58°C, 30 s at 68°C, and a final extension step of 5 min at 68°C.

CpG methylated Jurkat DNA (New England Biolabs) was used as a positive control sample, and assay efficiency of each qMSP assay was determined by generating standard curves as described previously (Supplementary Table S3).

**Data analysis and Statistics**

The relative threshold method, Cₙ (Applied Biosystems Relative Quantification “RQ” application on ThermoFisher Cloud) was used to determine cycle quantification (Cₙ) values for each amplification curve. Cₙ parameter optimization (Early access version, ThermoFisher Scientific Cloud) was conducted to enhance reliable detection of amplification. Sample reactions with inconclusive amplification curves, contamination, or poor reaction efficiency were excluded from further analysis. Reactions with negative amplification were assigned a Cq two higher than the maximum observed Cq value in their respective cohort. Amplification data at each locus and for each sample type are listed in Supplementary Table S2. Normalized methylation levels were calculated using delta-delta Ct method as described below:

\[
\text{Normalized methylation levels} = \frac{2^{(P_t-S_t)}}{2^{(P_r-S_r)}}
\]
Where,

\[ P_t = C_q \] of positive control DNA control for target gene;
\[ S_t = C_q \] of sample for target gene;
\[ P_r = C_q \] of positive control DNA for reference gene \((ALU)\);
\[ S_r = C_q \] of sample for reference gene \((ALU)\)

Exploratory analyses were performed using the training dataset, and differential methylation levels of 15 selected DNA methylation alterations were assessed as fold changes using a Mann-Whitney test. \( p \) values were adjusted for false discovery using the family-wise Bonferroni method.\(^52,53\) All DNA methylation alterations with significant enrichment in cancer samples compared to benign were considered for downstream analysis. Univariate and multivariate logistic regression analysis assessing all possible combinations of significant DNA methylation alterations were performed and the resulting models were ranked according to their balanced accuracy. Receiver operating characteristic (ROC) curve analysis, areas under these curves (AUC), and confusion matrices were generated for best-performing models using model thresholds determined from the "closest topleft" method.\(^53,54\) The best model was selected using the training cohort dataset and was then applied to the validation cohort dataset. Statistical analysis was performed in R (v3.4.1) using "pROC", "caret", “ggrepel” and "ggplot2" packages.\(^53–56\)

Results

Common methylation alterations in prostate cancer

We assembled three radical prostatectomy cohorts and extracted over 1300 DNA samples – (see Table 1 and Supplementary Table S2). These cohorts were selected originally to study prognostic biomarkers. However, as a by-product of that work, we identified diagnostic biomarkers using the following approach. Cases from two cohorts with 41 benign samples and 890 cancer samples from 480 patients were merged into a training set. An independent cohort from a 3rd hospital contained 55 benign samples and 377 cancer samples from 219 patients and was used for validation.

In prostate cancer, CpG island hypermethylation takes place in large blocks \((> 1 \text{ kb})\).\(^57–61\) For example, methylation levels of seven Illumina 450K probes covering a \(GSTP1\) CpG island were found to be consistently higher in cancer compared to benign tissue (Supplementary Figure S1).\(^57\) Thus, small regions of CpG islands can be viewed as representative when assessing DNA methylation status in cancer. Using real-time MSP assays, we profiled methylation changes in ~100 bp regions representing CpG islands at 15 different genomic loci which are among the most frequently reported as hypermethylated in prostate cancer (see Supplementary Table S1).\(^15,19–22,30,62–67\) In the training cohort, 14 out of 15 of these loci were significantly hypermethylated (adjusted \( P \) value < 0.01) with normalized methylation levels > 2 in 890 cancer samples compared to 41 benign samples (Figure 1). In contrast, methylation levels of the HIC1
CpG island were hypermethylated at similar levels in cancer and benign tissues, possibly representing a cancerization field effect. 

DNA methylation at seven loci, GSTP1, CCDC181, HAPLN3, GSTM2, GAS6, RASSF1, and APC, showed the largest differences between cancer and benign tissues (Figure 2). For each of these seven regions, DNA methylation levels in benign tissue was minimal with low variation (Figure 2). In a univariate logistic modelling of the training dataset, the area under the curve (AUC) from ROC analysis for each of these regions ranged from 83% to 95%, individually. The specificity of these univariate logistic models ranged from 77% to 90%, and the sensitivity ranged from 72% to 91% (Figure 2). The GSTP1 locus was highly methylated in cancer, but not in benign samples. As a cancer classifier, DNA methylation at GSTP1 locus demonstrated an AUC of 95% and balanced accuracy of 88%. TCGA prostate cancer methylation data show similar results (Supplementary Figure S1). Two other loci, GAS6 and APC, demonstrated strong diagnostic capabilities with comparable balanced accuracies to GSTP1, but with individual AUCs of < 90%. We found that regardless of the model threshold chosen, each single gene had false positive and/or false negative rates of 10% or higher. Therefore, to improve accuracy we performed multigene logistic modelling.

**Multigene diagnostic model in prostate cancer**

The multivariate approach chosen relied on the simplest binary classifier model, logistic regression. Using the training dataset, we tested all possible combinations of all 15 methylation regions to identify a multigene model with higher sensitivity and specificity. We identified the GAS6/GSTP1/HAPLN3 model as the best binary (cancer/benign) classifier with an AUC of 97% for the ROC curve (Table 2, Figure 3, panel A). Using the closest top-left method, we identified the threshold of 0.917 for the three-gene model, which produced specificity and sensitivity of 92% (Figure 3, panel A). A summary of the performance of one, two, or three gene models using GAS6/GSTP1/HAPLN3 DNA methylation is shown in Supplementary Table S4. Judging by its large coefficient, methylation levels at GSTP1 locus contributed most significantly to the classifier. This is consistent with its frequent inclusion in other prostate classifiers (Supplementary Table S4). However, methylation levels at GAS6 and HAPLN3 loci also made significant contributions to the model (p-value < 1.0e-06; Table 2).

Having optimized an accurate classifier, we then used the same threshold to validate the GAS6/GSTP1/HAPLN3 model in an independent cohort. As illustrated in in Table 3 and Figure 3 (panel B), the best three gene model (GAS6/GSTP1/HAPLN3), misclassified only 2 of 30 benign samples (6.7%) from the validation dataset as cancer. As for the cancer samples, only 12 out of 212 samples (5.6%) were misclassified as benign. The three-gene model showed sensitivity of 94% and specificity of 93% in the validation dataset. Overall, the GAS6/GSTP1/HAPLN3 model showed a significant improvement over univariate approaches, with a balanced accuracy of 94%, positive predictive value (PPV) of 99% and a negative predictive value (NPV) of 70% in the validation dataset (Table 3).
Discussion

PSA screening leads to over 900,000 negative biopsies per year in the United States alone, many of which would be unnecessary if an accurate non-invasive test was available. Despite significant efforts, very few molecular features of prostate cancer have been validated for this purpose. PROGENSA is the only FDA approved test currently utilized for prostate cancer diagnosis. This test detects changes in PCA3 and PSA RNA levels in urine samples of patients suspected of harboring prostate cancer. With balanced accuracy of 67% and PPV of 34%, this test however, produces a large number of false positives.

In considering richer sources of potential biomarkers for early prostate cancer, DNA hypermethylation is by far the most diverse and prevalent genomic aberration. It is found at higher levels in cancer tissues compared to benign histological samples such as normal/adjacent normal, benign prostate hyperplasia (BPH) and prostatic intraepithelial neoplasia (PIN). Its stability and ease of detection in archival tissues further enhance the feasibility and appeal of its use in clinical applications.

Few methylation-based classifiers have been rigorously tested. Studies to date have small sample sizes and few have validated a DNA methylation-based diagnostic classifier for prostate cancer. To date, only one epigenetic assay, ConfirmMDx, is commercially available, which measures methylation at APC, GSTP1, and RASSF1 loci, and is intended for use on benign prostate biopsies to identify men who are likely to have cancer in a subsequent biopsy. This test, designed to detect molecular features of a field cancerization effect, represents a challenging application since field effects are reportedly variable in time and space. It is therefore not surprising that this test shows limited specificity (64%) and sensitivity (66%) with an overall misclassification rate exceeding 35%. The current work addresses a simpler question: Whether or not a biospecimen contains cancer, and represents the largest independently validated study of its kind. Methylation levels at GSTP1, HAPLN3, and GAS6 loci formed the basis of the classifier, which demonstrated high accuracy. In univariate analysis, none of the DNA methylation alterations investigated here possessed false negative and false positive rates below 10%. Through multivariate analysis, the best model containing GAS6, GSTP1, and HAPLN3 DNA methylation alterations demonstrated high balanced accuracy with false positive and false negative rates of approximately 6%.

Since cancer-specific DNA hypermethylation is heavily documented in the literature, the biological relevance of genes in the three-gene model merits attention. Methylation at the GSTP1 locus contributed most significantly to the final three-gene classifier. GSTP1 belongs to the glutathione S-transferases (GSTs) family, which help maintain cell integrity and protect against DNA damage by detoxifying electrophilic substances. In recent studies, additional non-enzymatic functions of GSTP1 (i.e. protein-protein interactions) were elucidated as playing a major role in cell proliferation. In prostate cancer, GSTP1 DNA methylation effectively silences the gene. GSTP1 silencing can activate Stat3, which has been implicated as an oncoprotein in prostate cancer progression.
The functional relevance of GAS6 and HAPLN3 methylation, both of which contributed roughly equally to the final three-gene classifier (Table 2), are not well understood. GAS6 encodes an extracellular signalling peptide called growth arrest-specific 6. However the effect of DNA methylation on its expression has not been investigated. In vitro cell line studies have shown that GAS6 signalling promotes invasion but inhibits cell proliferation.\(^\text{86}\) In addition, GAS6 signals may also protect against apoptosis induced by chemotherapy.\(^\text{86,87,88}\) HAPLN3 (Hyaluronan And Proteoglycan Link Protein 3) belongs to the hyaluronan and proteoglycan binding link protein gene family which function to maintain extracellular matrix to support tissue architecture.\(^\text{89}\) Although this family of proteins has been previously found to be involved in drug resistance in multiple myeloma,\(^\text{90}\) the functional role of HAPLN3 in prostate epithelia is poorly understood, as are the consequences of DNA methylation at this locus.\(^\text{15}\) Their consistent hypermethylation levels in prostate cancer point to the importance of further investigation to elucidate the roles for GAS6 and HAPLN3 in prostate biology and disease.

Although there are no DNA methylation-based tests currently marketed for prostate cancer diagnosis, several other types of molecular tests have been marketed for this purpose. These tests show high negative predictive values, but limited specificity and sensitivity.\(^\text{11–13}\) For successful validation and implementation of non-invasive tests, biomarker assays can first be optimized and validated on tissue samples to demonstrate their specificity and sensitivity prior to testing them on liquid biopsy samples such as urine or blood.\(^\text{91}\) In this study, we used prostatectomy tissue samples instead of liquid biopsy samples (i.e. urine or blood) to build a prostate cancer classifier. We see this as a necessary step since many of the previous studies investigating DNA methylation in prostate cancer were based on small sample sizes.

Few DNA methylation tests investigated for prostate cancer diagnosis. GSTP1 methylation was tested in multiple studies showing sensitivity and specificity as a single gene urine biomarker of 52 % and 89 %, respectively.\(^\text{67,92}\) With highly sensitive new techniques such as next generation sequencing, increasing sensitivity should be feasible. The data presented here demonstrate theoretical maximum accuracy of a urine test for the best single gene classifier (GSTP1) vs. the three-gene classifier (GAS6/GSTP1/HAPLN3). In a hypothetical cohort of 1000 men with elevated PSA levels, 293 of whom have prostate cancer,\(^\text{4}\) optimal testing for GSTP1 would incorrectly classify 250 men (122 false positive (FP), 128 false negative (FN)) (Supplementary Table S4), whereas the three-gene classifier would cut the misclassification rate by over two-fold, misclassifying only 124 men (67 FP and 57 FN) (Table 3).

Current prostate cancer guidelines recommend that men with elevated PSA levels undergo prostate biopsy.\(^\text{93,94}\) A highly sensitive and specific urine test could potentially avoid hundreds of thousands of unnecessary biopsies annually.\(^\text{6}\) Upon validating the three-gene classifier (in urine samples) it could be used to identify a significantly smaller population of men who should undergo confirmatory biopsy.

**Limitations of this study:** As part of a larger forthcoming study of prognostic biomarkers in prostate cancer, the cohorts used are skewed towards low and intermediate risk patients. Therefore, the potential performance of this classifier on high grade/high stage cancers has not...
been evaluated. This limitation could be mitigated by limiting the use of such a classifier to men who lack high risk clinical and laboratory features, such as PSA levels above 10 ng/ml and/or suspicious findings on digital rectal examination.\textsuperscript{95} Likewise, although DNA methylation alterations are thought to be one of the first events in carcinogenesis,\textsuperscript{72,79,83} patients with familial prostate cancer show different patterns of DNA methylation than sporadic cases, and these rare patients may not be captured in the present study.\textsuperscript{96} In addition, unlike sensitivity and specificity, the large differences in number of benign and cancer samples used in this study preclude us from accurately determining NPV and PPV.\textsuperscript{97} These limitations will need to be addressed in subsequent investigations. Nevertheless, this work is unique in developing and validating a list of differentially and consistently hypermethylated genomic loci in prostate cancer, along with inexpensive assays that should be compatible with routine workflow in clinical laboratories.

**Disclosure of Potential Conflicts of Interest**

No potential conflicts of interest to report.

**Acknowledgments**

This work is proudly funded by the Movember Foundation through Prostate Cancer Canada (Grant #T2014.01); research grants from Ride for Dad and Southeastern Ontario Academic Medical Association (SEAMO) to D. M. B., and Terry Fox Foundation Training Program in Transdisciplinary Cancer Research in Partnership with CIHR, Ontario Graduate Scholarship and Queen’s Graduate awards to P.G.P.
References


Figure Legends:

Figure 1: Genomic loci with cancer-specific hypermethylation. Fourteen cancer-specific DNA methylation alterations were identified with fold change of > 2 and Bonferroni-corrected p-value < 0.01 (highlighted in pink) in the training dataset. Changes in DNA methylation levels between benign and cancer samples are represented as a fold change, and their corresponding adjusted p-values (Mann-Whitney U test after Bonferroni correction) are shown in this volcano plot.

Figure 2: Methylation levels at GSTP1, CCDC181, HAPLN3, GSTM2, GAS6, RASSF1, and APC loci are far higher in cancer than in benign prostate tissue. Boxplots and ROC curves show the distribution of the normalized methylation levels in cancer (red) and benign (blue) samples for each of the top DNA methylation alterations in the training set. The area under the curve for each of the ROC curves is annotated with sensitivity and specificity corresponding to the best threshold (according to the “closest topleft” method).

Figure 3: Independent validation of a three-gene binary cancer classifier (GAS6/GSTP1/HAPLN3). Panel A) An ROC curve of the three-gene classifier is shown, along with its AUC, sensitivity and specificity. Using the closest top-left method, we identified the model threshold of 0.917 for this three-gene model, which produced the specificity and sensitivity of 92% in the training dataset. Panel B) This binary classifier was tested on the validation dataset, and the classification of the benign and cancer samples is shown in blue and red, respectively. A red horizontal line is also plotted showing the model threshold from Figure 3, panel A, and only 14 out of 242 samples from the validation dataset were misclassified, showing error rate of < 6%.