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

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]

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

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.[7–10] 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.[11–13] 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,[14–25] with growing evidence supporting their role in progression and risk stratification.[26–28] Some investigators study therapeutic strategies based on methylation inhibitors,[14,29] others are working to develop DNA methylation alterations as useful diagnostic biomarkers.[30] 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 *ABCB1* loci were hypermethylated in >85% of cancers.[19] Haldrup et al. added *AOX1, CCDC181, GAS6, HAPLN3, KLF8*, and *MOB3B* loci as exhibiting cancer-specific hypermethylation and association with biochemical recurrence,[21] and Vanaja, Mahapatra and others identified additional DNA methylation alterations associated with recurrence or risk of progression using genome-wide approaches.[20,25,28,31]

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.[32,33] In a

multicenter validation with 848 patients, the test yielded a high negative predictive value (89%), but limited specificity (64 %) and sensitivity (65 %).[32,33] Similarly, Haldrup et al.[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.[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.[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.[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.[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.[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 [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[46] 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.[44,47]

We modified previously described protocols for real time MSP (qMSP) assays according to MIQE guidelines[42,44,48,49] and quantified changes in 15 DNA methylation alterations (Supplementary Table S3) in DNA samples collected from three RP cohorts.[45,50] 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), $MgCl_2$ (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 95ºC for 30 s, 7 cycles of touch-down PCR with annealing temperatures decreasing by 2ºC per cycle and extension at 68ºC 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).[49]

Data analysis and Statistics

The relative threshold method, C_{rt} (Applied Biosystems Relative Quantification "RQ" application on ThermoFisher Cloud) was used to determine cycle quantification (C_q) values for each amplification curve. C_{rt} 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[51] 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 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.[19]

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).[57] 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).[32,33,62] 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.[6] 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.[68–70]

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).[19,71–73] 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[62,74,75] and few have validated a DNA methylation-based diagnostic classifier for prostate cancer. To date, only one epigenetic assay, ConfirmMDx,[32,33] 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.[66,71,76] It is therefore not surprising that this test shows limited specificity (64 %) and sensitivity (66 %) with an overall misclassification rate exceeding 35 %.[32,33] 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,[77–79] 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.[80] In prostate cancer, *GSTP1* DNA methylation effectively silences the gene.[67,72,81–83] *GSTP1* silencing can activate Stat3, which has been implicated as an oncoprotein in prostate cancer progression.[84,85]

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.[86] In addition, GAS6 signals may also protect against apoptosis induced by chemotherapy.[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.[89] Although this family of proteins has been previously found to be involved in drug resistance in multiple myeloma,[90] the functional role of HAPLN3 in prostate epithelia is poorly understood, as are the consequences of DNA methylation at this locus.[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.[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.[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.[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,[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.[93,94] A highly sensitive and specific urine test could potentially avoid hundreds of thousands of unnecessary biopsies annually.[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.[95] Likewise, although DNA methylation alterations are thought to be one of the first events in carcinogenesis,[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.[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.[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

1. Brawley OW. Trends in prostate cancer in the United States. *J Natl Cancer Inst Monogr*. 2012;2012(45):152-156. doi:10.1093/jncimonographs/lgs035

2. Delongchamps NB, de la Roza G, Jones R, Jumbelic M, Haas GP. Saturation biopsies on autopsied prostates for detecting and characterizing prostate cancer. *BJU Int*. 2009;103(1):49-54. doi:10.1111/j.1464-410X.2008.07900.x

3. Haas GP, Delongchamps N, Brawley OW, Wang CY, de la Roza G. The worldwide epidemiology of prostate cancer: perspectives from autopsy studies. *Can J Urol*. 2008;15(1):3866-3871.

4. Schröder FH, Hugosson J, Roobol MJ, et al. Screening and Prostate-Cancer Mortality in a Randomized European Study. *N Engl J Med*. 2009;360(13):1320-1328. doi:10.1056/NEJMoa0810084

5. Wolf AMD, Wender RC, Etzioni RB, et al. American Cancer Society guideline for the early detection of prostate cancer: update 2010. *CA Cancer J Clin*. 2010;60(2):70-98. doi:10.3322/caac.20066

6. Aubry W, Lieberthal R, Willis A, Bagley G, Willis SM, Layton A. Budget Impact Model: Epigenetic Assay Can Help Avoid Unnecessary Repeated Prostate Biopsies and Reduce Healthcare Spending. *Am Health Drug Benefits*. 2013;6(1):15-24.

7. Tokheim CJ, Papadopoulos N, Kinzler KW, Vogelstein B, Karchin R. Evaluating the evaluation of cancer driver genes. *Proc Natl Acad Sci*. 2016;113(50):14330-14335. doi:10.1073/pnas.1616440113

8. Rizzardi AE, Rosener NK, Koopmeiners JS, et al. Evaluation of protein biomarkers of prostate cancer aggressiveness. *BMC Cancer*. 2014;14:244. doi:10.1186/1471-2407-14-244

9. Wang L-Y, Cui J-J, Zhu T, et al. Biomarkers identified for prostate cancer patients through genome-scale screening. *Oncotarget*. 2017;8(54):92055-92063. doi:10.18632/oncotarget.20739

10. Souza MF de, Kuasne H, Barros-Filho M de C, et al. Circulating mRNAs and miRNAs as candidate markers for the diagnosis and prognosis of prostate cancer. *PLOS ONE*. 2017;12(9):e0184094. doi:10.1371/journal.pone.0184094

11. Alford AV, Brito JM, Yadav KK, Yadav SS, Tewari AK, Renzulli J. The Use of Biomarkers in Prostate Cancer Screening and Treatment. *Rev Urol*. 2017;19(4):221-234. doi:10.3909/riu0772

12. Gaudreau P-O, 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

13. Kohaar I, Petrovics G, Srivastava S. A Rich Array of Prostate Cancer Molecular Biomarkers: Opportunities and Challenges. *Int J Mol Sci*. 2019;20(8). doi:10.3390/ijms20081813

14. Ngollo M, Dagdemir A, Karsli-Ceppioglu S, et al. Epigenetic modifications in prostate cancer. *Epigenomics*. 2014;6(4):415-426. doi:10.2217/epi.14.34

15. Strand SH, Orntoft TF, Sorensen KD. Prognostic DNA methylation markers for prostate cancer. *Int J Mol Sci*. 2014;15(9):16544-16576. doi:10.3390/ijms150916544

16. Lee WH, Morton RA, Epstein JI, et al. Cytidine methylation of regulatory sequences near the pi-class glutathione S-transferase gene accompanies human prostatic carcinogenesis. *Proc Natl Acad Sci*. 1994;91(24):11733-11737. doi:10.1073/pnas.91.24.11733

17. Lee WH, Isaacs WB, Bova GS, Nelson WG. CG island methylation changes near the GSTP1 gene in prostatic carcinoma cells detected using the polymerase chain reaction: a new prostate cancer biomarker. *Cancer Epidemiol Prev Biomark*. 1997;6(6):443-450.

18. Cairns P, Esteller M, Herman JG, et al. Molecular Detection of Prostate Cancer in Urine by GSTP1 Hypermethylation. *Clin Cancer Res*. 2001;7(9):2727-2730.

19. Yegnasubramanian S, Kowalski J, Gonzalgo ML, et al. Hypermethylation of CpG islands in primary and metastatic human prostate cancer. *Cancer Res*. 2004;64(6):1975-1986.

20. Aryee MJ, Liu W, Engelmann JC, et al. DNA methylation alterations exhibit intraindividual stability and interindividual heterogeneity in prostate cancer metastases. *Sci Transl Med*. 2013;5(169):169ra10. doi:10.1126/scitranslmed.3005211

21. Haldrup C, Mundbjerg K, Vestergaard EM, et al. DNA methylation signatures for prediction of biochemical recurrence after radical prostatectomy of clinically localized prostate cancer. *J Clin Oncol Off J Am Soc Clin Oncol*. 2013;31(26):3250-3258. doi:10.1200/JCO.2012.47.1847

22. Rouprêt M, Hupertan V, Yates DR, et al. Molecular detection of localized prostate cancer using quantitative methylation-specific PCR on urinary cells obtained following prostate massage. *Clin Cancer Res Off J Am Assoc Cancer Res*. 2007;13(6):1720-1725. doi:10.1158/1078-0432.CCR-06-2467

23. Ge Y-Z, Xu L-W, Jia R-P, et al. The association between RASSF1A promoter methylation and prostate cancer: evidence from 19 published studies. *Tumour Biol J Int Soc Oncodevelopmental Biol Med*. 2014;35(4):3881-3890. doi:10.1007/s13277-013-1515-3

24. Kobayashi Y, Absher DM, Gulzar ZG, et al. DNA methylation profiling reveals novel biomarkers and important roles for DNA methyltransferases in prostate cancer. *Genome Res*. 2011;21(7):1017-1027. doi:10.1101/gr.119487.110

25. Mahapatra S, Klee EW, Young CYF, et al. Global methylation profiling for risk prediction of prostate cancer. *Clin Cancer Res Off J Am Assoc Cancer Res*. 2012;18(10):2882-2895. doi:10.1158/1078-0432.CCR-11-2090

26. Fraser M, Sabelnykova VY, Yamaguchi TN, et al. Genomic hallmarks of localized, non-indolent prostate cancer. *Nature*. 2017;541(7637):359-364. doi:10.1038/nature20788

27. Ruggero K, Farran-Matas S, Martinez-Tebar A, Aytes A. Epigenetic Regulation in Prostate Cancer Progression. *Curr Mol Biol Rep*. 2018;4(2):101-115. doi:10.1007/s40610-018-0095-9

28. Vanaja DK, Ehrich M, Van den Boom D, et al. Hypermethylation of genes for diagnosis and risk stratification of prostate cancer. *Cancer Invest*. 2009;27(5):549-560. doi:10.1080/07357900802620794

29. Perry AS, Watson RWG, Lawler M, Hollywood D. The epigenome as a therapeutic target in prostate cancer. *Nat Rev Urol*. 2010;7(12):668-680. doi:10.1038/nrurol.2010.185

30. Valdés-Mora F, Clark SJ. Prostate cancer epigenetic biomarkers: next-generation technologies. *Oncogene*. 2015;34(13):1609-1618. doi:10.1038/onc.2014.111

31. Lin P-C, Giannopoulou EG, Park K, et al. Epigenomic alterations in localized and advanced prostate cancer. *Neoplasia N Y N*. 2013;15(4):373-383.

32. Partin AW, Van Neste L, Klein EA, et al. Clinical validation of an epigenetic assay to predict negative histopathological results in repeat prostate biopsies. *J Urol*. 2014;192(4):1081-1087. doi:10.1016/j.juro.2014.04.013

33. Stewart GD, Van Neste L, Delvenne P, et al. Clinical utility of an epigenetic assay to detect occult prostate cancer in histopathologically negative biopsies: results of the MATLOC study. *J Urol*. 2013;189(3):1110-1116. doi:10.1016/j.juro.2012.08.219

34. Van Neste L, Hendriks RJ, Dijkstra S, et al. Detection of High-grade Prostate Cancer Using a Urinary Molecular Biomarker-Based Risk Score. *Eur Urol*. 2016;70(5):740-748. doi:10.1016/j.eururo.2016.04.012

35. Haese A, Trooskens G, Steyaert S, et al. Multicenter Optimization and Validation of a 2-Gene mRNA Urine Test for Detection of Clinically Significant Prostate Cancer Prior to Initial Prostate Biopsy. *J Urol*. April 2019:101097JU0000000000000293. doi:10.1097/JU.0000000000000293

36. Health Quality Ontario. Prolaris Cell Cycle Progression Test for Localized Prostate Cancer: A Health Technology Assessment. *Ont Health Technol Assess Ser*. 2017;17(6):1-75.

37. Klein EA, Cooperberg MR, Magi-Galluzzi C, et al. A 17-gene assay to predict prostate cancer aggressiveness in the context of Gleason grade heterogeneity, tumor multifocality, and biopsy undersampling. *Eur Urol*. 2014;66(3):550-560. doi:10.1016/j.eururo.2014.05.004

38. Ioannidis JPA, Bossuyt PMM. Waste, Leaks, and Failures in the Biomarker Pipeline. *Clin Chem*. 2017;63(5):963-972. doi:10.1373/clinchem.2016.254649

39. Ransohoff DF. Challenges and opportunities in evaluating diagnostic tests. *J Clin Epidemiol*. 2002;55(12):1178-1182.

40. Ioannidis JPA. Biomarker Failures. *Clin Chem*. 2013;59(1):202-204. doi:10.1373/clinchem.2012.185801

41. Herman JG, Graff JR, Myöhänen S, Nelkin BD, Baylin SB. Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proc Natl Acad Sci U S A*. 1996;93(18):9821-9826.

42. Huang Z, Bassil CF, Murphy SK. Methylation-specific PCR. *Methods Mol Biol Clifton NJ*. 2013;1049:75-82. doi:10.1007/978-1-62703-547-7_7

43. Ku J-L, Jeon Y-K, Park J-G. Methylation-specific PCR. *Methods Mol Biol Clifton NJ*. 2011;791:23-32. doi:10.1007/978-1-61779-316-5_3

44. Olkhov-Mitsel E, Zdravic D, Kron K, Kwast T van der, Fleshner N, Bapat B. Novel Multiplex MethyLight Protocol for Detection of DNA Methylation in Patient Tissues and Bodily Fluids. *Sci Rep*. 2014;4:4432. doi:10.1038/srep04432

45. Patel P, Selvarajah S, Boursalie S, et al. Extraction of Both RNA and DNA from Formalin-fixed Paraffin Embedded Tissue Cores | Protocol. *J Vis Exp JoVE*. 2016;e54299. doi:10.3791/54299

46. Ongenaert M, Van Neste L, De Meyer T, Menschaert G, Bekaert S, Van Criekinge W. PubMeth: a cancer methylation database combining text-mining and expert annotation. *Nucleic Acids Res*. 2008;36(suppl_1):D842-D846. doi:10.1093/nar/gkm788

47. Weisenberger DJ, Campan M, Long TI, et al. Analysis of repetitive element DNA methylation by MethyLight. *Nucleic Acids Res*. 2005;33(21):6823-6836. doi:10.1093/nar/gki987

48. Ku J-L, Jeon Y-K, Park J-G. Methylation-specific PCR. *Methods Mol Biol Clifton NJ*. 2011;791:23-32. doi:10.1007/978-1-61779-316-5_3

49. Bustin SA, Benes V, Garson JA, et al. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem*. 2009;55(4):611-622. doi:10.1373/clinchem.2008.112797

50. Patel PG, Selvarajah S, Guérard K-P, et al. Reliability and performance of commercial RNA and DNA extraction kits for FFPE tissue cores. *PloS One*. 2017;12(6):e0179732. doi:10.1371/journal.pone.0179732

51. Pfaffl MW. A new mathematical model for relative quantification in real-time RT–PCR. *Nucleic Acids Res*. 2001;29(9):e45.

52. Bonferroni C. Teoria statistica delle classi e calcolo delle probabilita. *Pubblicazioni R Ist Super Sci Econ E Commericiali Firenze*. 1936;8:3-62.

53. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; 2017.

54. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011;12:77. doi:10.1186/1471-2105-12-77

55. Kamil Slowikowski. *Ggrepel: Repulsive Text and Label Geoms for "Ggplot2."*; 2017.

56. Wickham H. *Ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag; 2009.

57. The Cancer Genome Atlas Research Network. The molecular taxonomy of primary prostate cancer. *Cell*. 2015;163(4):1011-1025. doi:10.1016/j.cell.2015.10.025

58. Bhasin JM, Lee BH, Matkin L, et al. Methylome-wide Sequencing Detects DNA Hypermethylation Distinguishing Indolent from Aggressive Prostate Cancer. *Cell Rep*. 2015;13(10):2135-2146. doi:10.1016/j.celrep.2015.10.078

59. Bhasin JM, Hu B, Ting AH. MethylAction: detecting differentially methylated regions that distinguish biological subtypes. *Nucleic Acids Res*. 2016;44(1):106-116. doi:10.1093/nar/gkv1461

60. Gaspar JM, Hart RP. DMRfinder: efficiently identifying differentially methylated regions from MethylC-seq data. *BMC Bioinformatics*. 2017;18. doi:10.1186/s12859-017-1909-0

61. Wu H, Xu T, Feng H, et al. Detection of differentially methylated regions from whole-genome bisulfite sequencing data without replicates. *Nucleic Acids Res*. 2015;43(21):e141. doi:10.1093/nar/gkv715

62. Ahmed H. Promoter Methylation in Prostate Cancer and its Application for the Early Detection of Prostate Cancer Using Serum and Urine Samples. *Biomark Cancer*. 2010;2:17-33. doi:10.4137/BIC.S3187

63. Costa VL, Henrique R, Jerónimo C. Epigenetic Markers for Molecular Detection of Prostate Cancer. *Dis Markers*. 2007;23(1-2):31-41. doi:10.1155/2007/356742

64. Hopkins TG, Burns PA, Routledge MN. DNA Methylation of GSTP1 as Biomarker in Diagnosis of Prostate Cancer. *Urology*. 2007;69(1):11-16. doi:10.1016/j.urology.2006.10.008

65. Kang GH, Lee S, Lee HJ, Hwang KS. Aberrant CpG island hypermethylation of multiple genes in prostate cancer and prostatic intraepithelial neoplasia. *J Pathol*. 2004;202(2):233-240. doi:10.1002/path.1503

66. Mehrotra J, Varde S, Wang H, et al. Quantitative, spatial resolution of the epigenetic field effect in prostate cancer. *The Prostate*. 2008;68(2):152-160. doi:10.1002/pros.20675

67. Wu T, Giovannucci E, Welge J, Mallick P, Tang W-Y, Ho S-M. Measurement of GSTP1 promoter methylation in body fluids may complement PSA screening: a meta-analysis. *Br J Cancer*. 2011;105(1):65-73. doi:10.1038/bjc.2011.143

68. Chunhua L, Zhao H, Zhao H, et al. Clinical Significance of Peripheral Blood PCA3 Gene Expression in Early Diagnosis of Prostate Cancer. *Transl Oncol*. 2018;11(3):628-632. doi:10.1016/j.tranon.2018.02.019

69. Deras IL, Aubin SMJ, Blase A, et al. PCA3: a molecular urine assay for predicting prostate biopsy outcome. *J Urol*. 2008;179(4):1587-1592. doi:10.1016/j.juro.2007.11.038

70. Gittelman MC, Hertzman B, Bailen J, et al. PCA3 molecular urine test as a predictor of repeat prostate biopsy outcome in men with previous negative biopsies: a prospective multicenter clinical study. *J Urol*. 2013;190(1):64-69. doi:10.1016/j.juro.2013.02.018

71. Møller M, Strand SH, Mundbjerg K, et al. Heterogeneous patterns of DNA methylation-based field effects in histologically normal prostate tissue from cancer patients. *Sci Rep*. 2017;7:40636. doi:10.1038/srep40636

72. Martignano F, Gurioli G, Salvi S, et al. GSTP1 Methylation and Protein Expression in Prostate Cancer: Diagnostic Implications. *Dis Markers*. 2016;2016. doi:10.1155/2016/4358292

73. Henrique R, Jerónimo C. Molecular detection of prostate cancer: a role for GSTP1 hypermethylation. *Eur Urol*. 2004;46(5):660-669; discussion 669. doi:10.1016/j.eururo.2004.06.014

74. Srivastava S, Kramer BS. Validation: a critical step in bringing biomarkers to clinical fruition. *Ann Epidemiol*. 2018;28(2):135-138. doi:10.1016/j.annepidem.2017.10.013

75. Zapała P, Dybowski B, Poletajew S, Radziszewski P. What Can Be Expected from Prostate Cancer Biomarkers A Clinical Perspective. *Urol Int*. 2018;100(1):1-12. doi:10.1159/000479982

76. Massie CE, Mills IG, Lynch AG. The importance of DNA methylation in prostate cancer development. *J Steroid Biochem Mol Biol*. 2017;166:1-15. doi:10.1016/j.jsbmb.2016.04.009

77. Esteller M. Epigenetics in cancer. *N Engl J Med*. 2008;358(11):1148-1159. doi:10.1056/NEJMra072067

78. Taby R, Issa J-PJ. Cancer epigenetics. *CA Cancer J Clin*. 2010;60(6):376-392. doi:10.3322/caac.20085

79. Wu Y, Sarkissyan M, Vadgama JV. Epigenetics in breast and prostate cancer. *Methods Mol Biol Clifton NJ*. 2015;1238:425-466. doi:10.1007/978-1-4939-1804-1_23

80. Singh S. Cytoprotective and regulatory functions of glutathione S-transferases in cancer cell proliferation and cell death. *Cancer Chemother Pharmacol*. 2015;75(1):1-15. doi:10.1007/s00280-014-2566-x

81. Bastian PJ, Yegnasubramanian S, Palapattu GS, et al. Molecular Biomarker in Prostate Cancer: The Role of CpG Island Hypermethylation. *Eur Urol*. 2004;46(6):698-708. doi:10.1016/j.eururo.2004.07.022

82. Greene KL, Li L-C, Okino ST, Carroll PR. Chapter 33 - Molecular Basis of Prostate Cancer. In: Mendelsohn J, Howley PM, Israel MA, Gray JW, Thompson CB, eds. *The Molecular Basis of Cancer (Third Edition)*. Philadelphia: W.B. Saunders; 2008:431-440. doi:10.1016/B978-141603703-3.10033-0

83. Knudsen BS, Vasioukhin V. Chapter 1 - Mechanisms of Prostate Cancer Initiation and Progression. In: Vande Woude GF, Klein G, eds. *Advances in Cancer Research*. Vol 109. Academic Press; 2010:1-50. doi:10.1016/B978-0-12-380890-5.00001-6

84. KOU X, CHEN N, FENG Z, LUO L, YIN Z. GSTP1 negatively regulates Stat3 activation in epidermal growth factor signaling. *Oncol Lett*. 2013;5(3):1053-1057. doi:10.3892/ol.2012.1098

85. Abdulghani J, Gu L, Dagvadorj A, et al. Stat3 Promotes Metastatic Progression of Prostate Cancer. *Am J Pathol*. 2008;172(6):1717-1728. doi:10.2353/ajpath.2008.071054

86. Shiozawa Y, Pedersen EA, Patel LR, et al. GAS6/AXL axis regulates prostate cancer invasion, proliferation, and survival in the bone marrow niche. *Neoplasia N Y N*. 2010;12(2):116-127.

87. Sainaghi PP, Castello L, Bergamasco L, Galletti M, Bellosta P, Avanzi GC. Gas6 induces proliferation in prostate carcinoma cell lines expressing the Axl receptor. *J Cell Physiol*. 2005;204(1):36-44. doi:10.1002/jcp.20265

88. Wu G, Ma Z, Hu W, et al. Molecular insights of Gas6/TAM in cancer development and therapy. *Cell Death Dis*. 2017;8(3):e2700. doi:10.1038/cddis.2017.113

89. Spicer AP, Joo A, Bowling RA. A Hyaluronan Binding Link Protein Gene Family Whose Members Are Physically Linked Adjacent to Chrondroitin Sulfate Proteoglycan Core Protein Genes THE MISSING LINKS. *J Biol Chem*. 2003;278(23):21083-21091. doi:10.1074/jbc.M213100200

90. Huynh M, Pak C, Markovina S, et al. Hyaluronan and proteoglycan link protein 1 (HAPLN1) activates bortezomib-resistant NF-κB activity and increases drug resistance in multiple myeloma. *J Biol Chem*. 2018;293(7):2452-2465. doi:10.1074/jbc.RA117.000667

91. Altintas Z, Tothill IE. Molecular biosensors: promising new tools for early detection of cancer. Nanobiosensors in Disease Diagnosis. doi:10.2147/NDD.S56772

92. Leygo C, Williams M, Jin HC, et al. DNA Methylation as a Noninvasive Epigenetic Biomarker for the Detection of Cancer. *Dis Markers*. 2017;2017. doi:10.1155/2017/3726595

93. Carroll PH, Mohler JL. NCCN Guidelines Updates: Prostate Cancer and Prostate Cancer Early Detection. *J Natl Compr Cancer Netw JNCCN*. 2018;16(5S):620-623. doi:10.6004/jnccn.2018.0036

94. Mohler JL, Antonarakis ES, Armstrong AJ, et al. Prostate Cancer, Version 2.2019, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Cancer Netw JNCCN*. 2019;17(5):479-505. doi:10.6004/jnccn.2019.0023

95. D'Amico AV, Whittington R, Malkowicz SB, et al. Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer. *JAMA*. 1998;280(11):969-974.

96. Taylor RA, Fraser M, Livingstone J, et al. Germline *BRCA2* mutations drive prostate cancers with distinct evolutionary trajectories. *Nat Commun*. 2017;8:13671. doi:10.1038/ncomms13671

97. Shaikh SA. Measures Derived from a 2 x 2 Table for an Accuracy of a Diagnostic Test. *J Biom Biostat*. 2011;2(5). doi:10.4172/2155-6180.1000128

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 horizonal 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%.