Development and Clinical Validation of an \textit{in situ} Biopsy Based Multi-Marker Assay for Risk Stratification in Prostate Cancer

Authors:

Peter Blume-Jensen, M.D., Ph.D.\(^1\), David M. Berman, M.D., Ph.D.\(^2\), David L. Rimm, M.D., Ph.D.\(^3\), Michail Shipitsin, Ph.D.\(^1\), Mathew Putzi, M.D.\(^4\), Thomas P. Nifong, M.D.\(^1\), Clayton Small\(^1\), Sibgat Choudhury, Ph.D.\(^1\), Teresa Capela\(^1\), Louis Coupal\(^5\), Christina Ernst\(^1\), Aeron Hurley\(^1\), Alex Kaprelyants, M.D.\(^1\), Hua Chang, Ph.D.\(^1\), Eldar Giladi, Ph.D.\(^1\)*, Julie Nardone, Ph.D.\(^1\), James Dunyak, Ph.D.\(^1\), Massimo Loda, M.D.\(^6\), Eric A. Klein, M.D.\(^7\), Cristina Magi-Galluzzi, M.D., Ph.D.\(^8\), Mathieu Latour, M.D.\(^9\), Jonathan I. Epstein, M.D.\(^10\), Philip Kantoff, M.D.\(^6\), and Fred Saad, M.D.\(^9\)

\(^1\)Metamark Genetics Inc., Cambridge, MA, \(^2\)Department of Pathology and Molecular Medicine, Queen’s University, Kingston, ON, Canada, \(^3\)Department of Pathology, Yale University Medical School, New Haven, CT, \(^4\)Urology Austin, Austin, TX, \(^5\)Impacts Inc., Montreal, QC, Canada, \(^6\)Dana-Farber Cancer Institute, Boston, MA, \(^7\)Glickman Urological and Kidney Institute, Cleveland Clinic, Cleveland, OH, \(^8\)Department of Anatomic Pathology, Cleveland Clinic, Cleveland, OH, \(^9\)Centre Hospitalier de l’Université de Montréal, Montreal, QC, Canada, \(^10\)Department of Pathology, Urology and Oncology, Johns Hopkins Hospital, Baltimore, MD

Corresponding author: Fred Saad, M.D., FRCS, Centre Hospitalier de l’Université de Montréal, Montreal, QC, Canada. Email: fred.saad@umontreal.ca. Tel. 514-890-8000 ext 27466

* Present address: Atreca Inc., San Carlos, CA.

Research support: This project was funded by Metamark Genetics, Inc.

Running head: Biopsy Test for Discriminating Prostate Cancer Outcomes
Acknowledgments

The authors would like to thank Drs. Giovanni Parmigiani, Lynda Chin, Raju Kucherlapati, Greg Critchfield and collaborators at Metamark for insightful comments and suggestions. We thank Rowena Hughes and Winnie McFadzean from Oxford PharmaGenesis™ Ltd, and Denise Spring for editorial support in collating comments from authors and finalization of the manuscript for submission.
Statement of translational relevance

Early stage prostate cancers exhibit significant heterogeneity on the genomic, pathological and clinical levels, motivating efforts to identify prognostic determinants capable of guiding patient management at the time of biopsy. Current pathological and clinical parameters are insufficient for accurate prediction of the disease state from prostate cancer biopsies, particularly those with Gleason grades 3+3 or 3+4. Here we sought to establish a novel assay for automated quantitative measurements of protein biomarkers in defined areas of intact formalin fixed paraffin-embedded (FFPE) biopsy-derived tissue. The assay provides independent prognostic information about the disease state at the time of biopsy and could aid in stratification of patients for active surveillance versus therapeutic intervention.

Abstract

Purpose

Prostate cancer aggressiveness and appropriate therapy are routinely determined following biopsy sampling. Current clinical and pathologic parameters are insufficient for accurate risk prediction leading primarily to overtreatment but also missed opportunities for curative therapy.

Experimental design

An 8-biomarker proteomic assay for intact tissue biopsies predictive of prostate pathology was defined in a study of 381 patient biopsies with matched prostatectomy specimens. A second blinded study of 276 cases validated this assay’s ability to distinguish ‘favorable’ versus ‘non-favorable’ pathology independently and relative to current risk classification systems (NCCN and D’Amico).
Results

A favorable biomarker risk score of ≤0.33, and a non-favorable risk score of >0.80 (possible range between 0 and 1) were defined on ‘false negative’ and ‘false positive’ rates of 10% and 5%, respectively. At a risk score ≤0.33, predictive values for favorable pathology in very low- and low-risk NCCN and low-risk D’Amico groups were 95%, 81.5%, and 87.2%, respectively, higher than for these current risk classification groups themselves (80.3%, 63.8%, and 70.6%, respectively). The predictive value for non-favorable pathology was 76.9% at biomarker risk scores >0.8 across all risk groups. Increased biomarker risk scores correlated with decreased frequency of favorable cases across all risk groups. The validation study met its two co-primary endpoints, separating favorable from non-favorable pathology (AUC, 0.68, P<0.0001, odds ratio=20.9) and GS-6 versus non-GS-6 pathology (AUC, 0.65, P<0.0001, OR=12.95).

Conclusion

The 8-biomarker assay provided individualized, independent prognostic information relative to current risk stratification systems, and may improve the precision of clinical decision-making following prostate biopsy.
INTRODUCTION

In 2014, there will be an estimated 233,000 new diagnoses of prostate cancer in the USA.(1) The majority of these patients will have early-stage, clinically localized disease.(1-3) Given the marked genomic and pathological heterogeneity of prostate cancer and concerns regarding its overtreatment,(4, 5) significant effort has been devoted to the discovery and validation of biomarkers capable of distinguishing indolent cases with good prognosis from more aggressive cases with poor survival.(6) Pathologic evaluation of needle biopsy tissue establishes a prostate cancer diagnosis and provides data to define a patient’s risk category that will guide treatment decisions.(6) Although a number of useful classification systems have been developed that combine available clinical and pathological parameters, (6, 7) it is widely recognized that improved classification methods are needed to ascribe an individualized risk score and help clinical management.(8-10)

A significant proportion of patients undergoing biopsy ultimately will have their tumor pathology (as defined by Gleason score) upgraded or downgraded following assignment of a more accurate ‘surgical’ or pathologic Gleason score after prostatectomy.(11) For example, 25-30% of patients with low-risk disease at biopsy are subsequently upgraded upon pathological review of their radical prostatectomy.(10-12) These revisions may reflect initial biopsy sampling error,(13) or pathologist discordance in tumor grading,(14) both of which can contribute either to overtreatment or delayed treatment of disease.(4) In particular, there are concerns around over-calling or under-calling Gleason pattern 4 in needle biopsy samples, since these cases usually fall in a grey zone between clear active surveillance and surgical intervention.(11, 15-17) In addition,
there is an ongoing need to determine whether a cancer is organ-confined, or non-organ-confined with ultimate metastatic potential in patients with low- to intermediate-grade disease.

Advances have been made in identifying genetic markers that may inform clinical risk prognostication in prostate cancer.(18, 19) There has been focus on identifying in situ protein biomarkers on the cellular level that, under circumstances of intra-tumor cellular heterogeneity, would enable more accurate classification of the most aggressive tumor areas which may consist of only few cancer cells.(20, 21) We recently reported the development and use of a robust quantitative multiplex proteomics in situ imaging (QMPI) approach to identify and measure protein biomarker candidates resistant to such sampling error(22) (see Supplementary Appendix). Starting with a large pool of potential candidates, this biopsy simulation study identified 12 biomarker candidates that predict both prostate pathology aggressiveness and lethal outcome despite sampling error (22) based on prostatectomy tissue samples (Supplementary Appendix). The 12 biomarkers are ACTN1, CUL2, DERL1, FUS, HSPA9, PDSS2, PLAG1, pS6, SMAD2, SMAD4, VDAC1, and YBX1.

In this study, we report further development and subsequent blinded validation of an 8-biomarker assay derived from 11 of these markers in two independent clinical biopsy studies. Specifically, we first designed a biomarker development study to define an optimal biomarker subset and develop a predictive mathematical model. This effort resulted in an 8-biomarker prognostic model that provided “risk scores” predictive of final disease pathology. Next, benchmarked against Gleason scores and TNM staging data of matched post-surgical prostatectomy tissue, the performance of this 8-biomarker biopsy assay to predict the clinically relevant endpoint of
favorable versus non-favorable pathology was evaluated in an independent clinical validation study. The risk score generated by the 8-biomarker assay for each patient was compared with two current risk stratification systems, the NCCN guideline categories and D’Amico system (6, 7) and its potential to provide supplementary data of high predictive value that could aid in treatment decision for individual patients was considered. The intended use of this biomarker assay is to supplement current biopsy based prostate cancer risk assessment methods in cases where a clinical decision regarding active surveillance versus active treatment is not straightforward.

METHODS

Clinical Samples

Biomarker Assay Development: In order to develop a robust assay, multiple institutions were recruited representing typical North American patient cohorts (Table 1): Urology Austin, Chesapeake Urology Associates, Cleveland Clinic, Michigan Urology, and Folio Biosciences. Biopsy sample inclusion/exclusion criteria matched those that would be in place during routine clinical use of the assay. Since they are usually not candidates for active surveillance, patients with biopsy Gleason ≥4+3 were excluded, except for a limited number of biopsies that had been discordantly graded as both 3+4 and 4+3 by two expert pathologists. Annotations including information on matched biopsy and prostatectomy pathology reports were obtained. Researchers were blinded to sample identity during laboratory processing. The final number of cases included in the biomarker assay development cohort was 381.
Clinical Validation: The validation study cohort (N = 276) was separate and independent from the assay development cohort, and comprised biopsy samples with matched prostatectomy annotation from patients managed at the University of Montreal Hospital Center, Canada (Table 1). Consent criteria and IRB approval steps were as for the assay development phase of the study, above. Inclusion criteria were biopsies with a centralized Gleason score 3+3 or 3+4 (biopsies with discordant grading by two expert pathologists of 3+4 and 4+3 were included as well), and matched prostatectomy with pathologic TNM staging, PSA level, and resulting surgical Gleason score. Central review was performed on biopsies, while for prostatectomies we had to rely on the original pathology annotation, due to practical challenges in accessing all the different locations where these had been conducted; however, all prostatectomy pathology annotation was done according to new ISUP classification. Sample sizes were statistically designed, with power greater than 95% for both cohorts.

Quantitative Multiplex Immunofluorescence Format

The QMPI approach for protein in situ measurements was as described (23) (Supplementary Appendix). Briefly, formalin-fixed, paraffin-embedded (FFPE) prostate cancer biopsy tissue slides were analyzed using a quantitative multiplex proteomics imaging (QMPI) platform for intact tissue that integrates morphological object recognition and molecular biomarker measurements from tumor epithelium at the individual slide level. Four combinations of three (triplex) biomarkers each were used: A) PLAG1, SMAD2, ACTN1; B) VDAC1, FUS, SMAD4; C) pS6, YBX1, DERL1; D) PDSS2, CUL2, DCC. When each of the primary antibodies used was validated for specificity, it was found that PLAG1 was insufficiently specific; it was thus excluded from the potential signature. Prostate tissue slides that passed initial quality control
were subjected to the multiplex staining procedure with subsequent image acquisition, Definiens analysis, and bioinformatics analysis. (See Supplementary Appendix for further details of antibody validation, staining protocols, image acquisition, image analysis, and inter-experimental controls.)

**Biomarker Assay Development**

A non-interventional, retrospective clinical assay development study using biopsy case tissue samples was devised to define the best marker subset from those candidates previously shown to correlate with both prostate pathology aggressiveness and lethal outcome. The study goal was to define a model able to distinguish between prostate pathology usually recommended for active surveillance (“GS 6”): surgical Gleason 3+3 and ≤T3a versus those more likely to require prostatectomy (“non-GS 6”): surgical Gleason ≥3+4 or non-localized >T3a or N or M. This GS 6 versus non-GS 6 definition was based on studies showing that tumors with surgical Gleason 3+3 at prostatectomy do not metastasize. Institutional Review Boards (IRBs) approved the biomarker development protocol, and patient consent was obtained or waived accordingly.

**Train-Test to Define Biomarker Set**

To rank our previously identified biomarkers in terms of their predictive importance, we evaluated each potential combination of biomarkers using 500 rounds of computation. In each round, a training set (65% of cases) was obtained by bootstrap and remaining samples (35% of cases) were used as a testing set. A logistic regression model was trained on the training set, and risk scores were obtained for samples in both the training and testing sets. The final marker coefficients were used in a logistic regression model for calculation of the risk score, provided as
a continuous scale from 0 to 1. After computation of these statistics, we sorted the biomarker sets in three different ways. Biomarker sets were characterized by the area under the receiver operating characteristic (ROC) curve (AUC), and sorted by increasing value of Akaike information criterion (AIC), decreasing value of the AUC on the training set, and decreasing value of the AUC on the testing set. To finalize the set of biomarkers for our assay, we determined the frequency of usage of each biomarker in the 10% of most highly ranked sets. Optimal biomarker set size was likewise determined through examination of test-set AUC performance during the exhaustive model search. We defined a resulting risk score, a continuous number between 0 and 1, which estimated the likelihood of “non-GS 6” pathology. Sensitivity analyses confirm the defined, locked-down biomarker assay set. We also explored a number of sensitivity analyses to confirm performance robustness. These analyses identified an optimal set of 8 biomarkers with the best performance in distinguishing GS6 from non-GS6 cases (Figure 2).

Clinical Validation Study

A non-interventional, blinded, prospectively designed, retrospectively collected clinical study was conducted to validate the performance of the 8-biomarker assay set to predict prostate pathology on its own and relative to current systems for patient risk categorization. This study used biopsy samples from a patient cohort independent of the one used for assay development (Table 1), and generated a “risk score” for each sample. Matched prostatectomy samples with annotated surgical Gleason scores ultimately classified the tumor as “favorable” or “non-favorable” for the purposes of evaluating assay results. ROCs and corresponding AUCs determined for the 8-biomarker risk score quantitatively evaluated performance of the assay.

Two Co-Primary Endpoints for Validation Study
Favorable versus non-favorable pathology was chosen for final patient categorization throughout the validation study. It reflects the increasing awareness that organ-confined disease with minimal Gleason 4 pattern is likely to remain indolent with a significantly better long-term prognosis than higher-grade (dominant Gleason 4 pattern) or non-organ-confined disease (16, 24, 25). Thus, one of our two co-primary endpoints was defined as ‘favorable’ pathology: surgical Gleason ≤3+4 and organ-confined disease (≤T2) versus ‘non-favorable’ pathology: surgical Gleason ≥4+3 or non-organ-confined disease (T3a, T3b, N, or M). Additionally, in recognition of clinical variability, a second co-primary endpoint was defined as “GS 6” pathology: surgical Gleason of 3+3 and localized disease (≤T3a) versus “non-GS 6” pathology: surgical Gleason ≥3+4 or non-localized disease (T3b, N, or M).

Statistical Analysis

Using biopsy samples, the 8-biomarker assay-derived risk score validated these two co-primary endpoints for prostate pathology, as assessed by AUC. Secondary analyses included odds ratios (ORs) for the highest quartile versus lowest quartile of risk score, and OR (point estimate) for the continuous scale. We compared the risk outcomes from our diagnostic test with current risk classification categories as defined by the NCCN and D’Amico,(6, 7) using positive predictive values (PPVs). Definition and statistical analysis of the Net Reclassification Improvement (NRI) was done as described by Pencina.(26)

See the Supplementary Appendix for the statistical plan for both clinical studies.

RESULTS
Biomarker Assay Development

We first sought to identify a protein biomarker set and predictive model capable of correlating biopsy-based prognosis with final post-prostatectomy pathology. Biomarker assay development utilized prostate biopsies and post-prostatectomy annotation from 381 patients recruited from several US institutions (Table 1). Eleven previously identified biomarker candidates were evaluated using a 5-step optimization process to determine the top-performing biomarker set from among all possible combinations (Supplementary Appendix, Figure S1). Individual quantitative biomarker measurements were correlated with prostate cancer pathology as an endpoint, and univariate odds ratios (OR) were calculated (Figure 2A). The performance of all possible biomarker combinations were assessed and several high-performing sets were identified; the biomarker set with the highest performance had a test AUC of 0.79 (95% confidence interval [CI], 0.72 to 0.84). Given that many biomarker sets had similar performance in the bootstrapped test AUC, both univariate performance and frequency of occurrence in top-performing models (Figure 2B) were used to choose the final biomarker set for validation. Model coefficients for each marker were derived from logistic regression on the full data set. The resulting locked-down biomarker assay set consisted of eight biomarkers (CUL2, DERL1, FUS, HSPA9, PDSS2, pS6, SMAD4, YBX1) as shown in Figure 2.

Clinical Validation Study

Clinical validation utilized biopsy specimens, and post-prostatectomy annotation from a second independent cohort of 276 patients. As shown in Table 2, the study met its two co-primary endpoints and validated the 8-biomarker assay for both endpoints: “favorable” versus “non-favorable” and “GS 6” versus “non-GS 6” (see Methods for endpoint details). The analysis for
“favorable” pathology yielded an AUC of 0.68 with 95% CI of 0.61 to 0.74. The associated P-value was P<0.0001, with an OR for risk score of 20.9 per unit change. “GS 6” pathology yielded an AUC of 0.65 with 95% CI of 0.58 to 0.72. The associated P-value was P<0.0001, with an OR for risk score of 12.6 per unit change. Further details are shown in Figures S2 and S3 in the Supplementary Appendix.

We had sufficient prostatectomy annotation to classify 256 cases within our validation cohort according to NCCN and D’Amico criteria, two current systems for risk classification. The performance for “favorable” pathology of the 8-biomarker assay on associated biopsy samples from this cohort is shown in Figure S4 and was similar to the full cohort of 276 cases (AUC, 0.69 [95% CI, 0.63 to 0.76]; P<0.0001; OR for risk score, 26.2 per unit change). The AUC for identification of the “favorable” group based on NCCN risk stratification alone was 0.69 [95% CI: 0.62-0.75] and 0.65 [95% CI: 0.59-0.71] based on D’Amico alone. The AUCs for the combined models of the 8-biomarker assay with NCCN and D’Amico were 0.75 [95% CI: 0.69-0.81] and 0.75 [95% CI: 0.69-0.81], respectively.

Multivariate analysis was also performed during development of the combined models. The Odds Ratio for aggressive disease per 0.25 change in the 8-biomarker assay risk score, with NCCN held as a constant, was 3.5 [pv<10^{-3}, 95% CI: 2.6 – 4.7]. The Odds Ratio for aggressive disease per 0.25 change in 8-biomarker assay risk score, when the D’Amico status was held constant, was also 3.5 [pv<10^{-3}, 95% CI: 2.6 – 4.7]. These results are consistent with the notion that the 8-biomarker assay, the results of which are entirely based on integrated molecular and
The 8-marker proteomic assay was also considered in the context of standard clinical parameters, including PSA, % cores positive as a surrogate for tumor volume, and biopsy Gleason pattern. These parameters were available to the investigators for the clinical development (‘train-test’) cohort and held as blinded data by the independent validating statistician for the validation cohort. A logistic regression model was built for the “favorable” endpoint on the train-test cohort using the clinical nomogram parameters alone as well as for the integrated model combining the clinical and 8-marker assays. The clinical parameters alone resulted in an AUC of 0.67 [95% CI: 0.61-0.74] and the combined model resulted in an AUC of 0.71 [95% CI: 0.64-0.77].

Multivariate analysis of the combined model showed that the odds ratio for aggressive disease per 0.25 change in 8-biomarker assay risk score, when clinical parameters were held constant, was 3.1 \[p<10^{-3}, 95\% \text{ CI: } 2.2 - 4.4\]. Again, these results are in accordance with the molecular information captured by the 8-biomarker assay is partly independent of the prognostic features based on the clinical and pathological parameters in the clinical nomogram.

Next, to define the risk score ranges (cut-points) for favorable vs non-favorable pathology, we generated sensitivity and specificity curves for the 8-biomarker assay (Figures 3A and 3B, respectively). Figure 3A shows an example of a favorable risk score category identified in this study population on the basis of the 8-biomarker assay. Here, a Sensitivity of 90\% (95\% CI, 82\% to 94\%) corresponds to a biomarker assay risk score of 0.33 (Sensitivity: \(P[\text{risk score}>0.33 \mid ...\))
non-favorable pathology]). Below this threshold of risk score, the false-negative rate among patients with non-favorable pathology is limited to 10% (95% CI, 6% to 18%). Stated differently, among patients with risk score ≤0.33, the chance of having non-favorable pathology is less than or equal to 1 in 10. Conversely, in Figure 3B, a non-favorable category was defined by a biomarker assay risk score of 0.80, which in this study population corresponded to a Specificity of 95% (95% CI, 90% to 98%). Above this threshold of risk score, the false-positive rate among patients with favorable pathology is limited to 5% (95% CI, 2% to 10%), where Specificity = (P[risk score≤0.80| favorable pathology]). Stated differently, among patients with risk score greater than 0.80, the likelihood of having a favorable pathology is less than 5 in 100.

**Clinical utility of the 8-biomarker assay**

Similar to prognostic assay such as OncotypeDx which stratifies an individual patient to a particular risk level through cut-points that define risk categories(27), the intended use of this 8-biomarker assay is to categorize an individual patient to favorable versus non-favorable disease pathology based on where his risk score is on the Specificity curves of Figure 3. To this end, we first calculated the predictive value of the risk scores. The positive predictive value (PPV) for identifying favorable disease with a risk score of ≤0.33 was 83.6% (specificity, 90%) (Table 3A). Conversely, at a risk score of >0.80, 76.9% had non-favorable disease (i.e. only 23.1% of patients in the high risk score category had favorable disease). This translates to 39% of patients in this study population having risk scores ≤0.33 or >0.8, of which 81% were correctly identified by our 8-biomarker assay (Table 3A). Of patients with intermediate risk scores (0.33 < risk score ≤ 0.8), 58.3% had favorable disease.
Next, we compared the PPV of this 8-biomarker assay with those of the NCCN and D’Amico risk categories. Here, within each NCCN category, we analyzed the distribution of patients categorized as having favorable disease by our biomarker risk score. The positive predictive value for favorable pathology, at a risk score of ≤0.33, was 75% for patients in the NCCN-intermediate-risk category; 81.5% for NCCN-low-risk patients; and 95% for NCCN-very-low-risk, with an overall PPV for favorable pathology of 85% (Table 3B, in green, and Figure 3C). This contrasts favorably with the PPVs obtained for the NCCN risk categories alone, which were 40.9% for intermediate-risk, 63.8% for low-risk, and 80.3% for very-low-risk (Table 3B, blue). In other words, when combined with the current prognostic standard, this 8-biomarker assay can increase the confidence that a patient categorized to NCCN very-low risk group indeed has a favorable pathology based on his biopsy from 80.3% to 95%. Conversely, the predictive value for non-favorable pathology was 76.9% at a risk score >0.8, but reached 100% at a risk score cut-point of >0.9. Correspondingly, a higher biomarker risk score correlated with decreased frequency of favorable cases within each NCCN category (Figure 3D). Taken together, we conclude that the risk score generated by our 8-biomarker assay offered additional prognostic value for individual patients relative to NCCN risk categories alone. Similar results were obtained when comparing with the D’Amico categories (Figure S5).

To further confirm the benefit of having the 8-biomarker risk score in the context of current risk stratification systems, we performed a Net Reclassification Improvement (NRI) analysis for the favorable and non-favorable categories determined by our 8-biomarker assay, relative to NCCN and D’Amico categories (Figure S6). Using the underlying data shown in Table 3, we found an NRI for NCCN of 0.34 (P<0.00001; 95% CI, 0.20 to 0.48) and for D’Amico of 0.24 (P=0.0001;
95% CI, 0.12 to 0.35). This provides strong evidence that this 8-biomarker assay provides additional discriminatory capability to that offered by current risk stratification systems alone. We also performed Decision Curve Analysis, as this provides another method for characterizing performance of different risk systems and risk cut points (28). Figure S7 provides an example of such an analysis and describes the net benefit for a number of treatment regimes based on the 8-marker model, NCCN, and the combined model NCCN and 8-marker model. Using the illustrated cutpoints, we see that the combined model provides benefit through the span of risk thresholds. The range of incremental net benefit improves from small (0.005 at risk threshold of 20%) to significant (0.06 at a risk threshold of 50%) across the scale of risk thresholds. This further demonstrates the added prognostic information achieved by combining the 8-marker model with a standard of care risk stratification system.

DISCUSSION

Recent studies indicate that long-term survival for patients with organ-confined Gleason 3+4 disease is significantly better than for patients with non-organ-confined disease or for tumor with dominant Gleason pattern 4 or higher,(17, 24, 25) and that deferred therapy for the former group does not significantly change long-term outcome.(29-31) Most risk stratification systems do not discriminate between Gleason 7 biopsies, so typically patients considered candidates for active surveillance belong to ‘very-low-risk’ or ‘low-risk’ groups, which only contain biopsy Gleason score ≤6.(2, 6) However, around 25% of Gleason grade 3+4 biopsies are ‘downgraded’ and a similar percentage of Gleason grade 3+3 biopsies are ‘upgraded’ when comparing with the surgical Gleason scores, primarily owing to biopsy sampling error and pathologist discordance.(11),(15). Therefore, there is a clear unmet clinical need for a molecular evidence-based prognostic assay that can complement the current pathology-based stratification system(s)
to more accurately predict disease aggressiveness of patients with Gleason grade 3+3 and 3+4 biopsies(16).

To address this unmet need, we have embarked on an assay development effort, building on prior work that had led to the identification of 11 biomarker candidates shown to be predictive of both prostate cancer pathology and lethal outcome despite tissue sampling error (22). Of importance is that we have employed a multiplex proteomics imaging platform (23) for quantitative measurements of eight independent biomarkers in intact tissue to generate the risk score (Supplementary Appendix). This in situ approach has the advantage compared with other tests using homogenized tissue (19, 32), as it requires relatively few cancer cells (which is common of the biopsy specimens) and is not sensitive to variations in the ratios of benign tissue relative to tumor tissue.

Here, we report the results demonstrating that we have successfully developed an 8-biomarker biopsy-based prognostic assay that was validated in an independent blinded study and shown to complement current risk stratification systems. Specifically, our study shows that, at a risk score of ≤0.33, our biomarker assay can identify patients with favorable pathology in the very-low- and low-risk NCCN and low-risk D’Amico groups with PPVs of 95%, 81.5%, and 87.2%, respectively, that represents a significant improvement over NCCN or D’Amico stratification system alone. We suggest that information provided by the biomarker assay in addition to existing pathological parameters can help both the physician and patient to consider active surveillance option with added confidence.
Furthermore, our 8-biomarker assay is also able to identify with high confidence patients with non-favorable pathology who are arguably unsuitable for active surveillance. While a patient with a risk score > 0.8 has a PPV of 76.9% for non-favorable pathology, the PPV reaches 100% if the risk score is > 0.9 across all risk groups for both risk stratification systems tested. This shows how this biomarker assay can be used in combination with current stratification system(s) to personalize the risk level for each individual patient, enabling rational and evidence-based decision by both physician and patient. Note that the settings used for clinical decision-making thresholds is a matter of risk tolerance in the patient and physician. Since our assay provides continuous specificity and sensitivity curves, and individualized risk for each patient, other thresholds that reflect different individual risk tolerance could be chosen. For example, if 80% specificity is sufficient for a particular patient, then a risk score of 0.6 or higher would favor treatment over active surveillance.

Because of the heterogeneous and multifocal nature of prostate cancer, prostate biopsies frequently contain only lower-grade morphological features with higher-grade molecular features that are not reflected in morphology. Our 8 biomarkers were specifically selected and evaluated to equally well predict pathology outcome regardless whether they were measured in low grade or high grade regions from the same patient(22). Accordingly, by measuring proteins only from intact tumor regions, our test can accurately and sensitively assess such high grade molecular features in situ, even in tissue samples with variable amounts of tumor versus benign components. This is a significant advantage over previously reported gene expression-based technologies that rely on tissue homogenization.(18, 33-35), and where molecular (genetic) information from higher grade regions is variably diluted by the values of the individual markers.
in intermixed benign tissue regions that inadvertently are part of the tissue homogenate. Genomic health’s “Oncotype DX prostate cancer test” is a 17 gene PCR based assay, similar to the biopsy based test described in this study, that provides Genomic Prostate Score (GPS) for predicting cancer aggressiveness in PCa patients based on the same endpoints we have developed (33, 34). While a stand-alone AUC for the “Oncotype Dx” test has not been reported, the combined AUC for GPS and CAPRA is reported to be 0.67, while the AUC for CAPRA alone was reported as 0.63 (for favorable pathology). Obviously, a direct comparison between these studies is difficult, as a head-to-head study would be required to fully establish relative performance. Myriad’s “Prolaris” test is another gene expression based assay that measures activity of a panel of Cell Cycle Progression (CCP) genes and provides risk of prostate cancer specific progression and mortality(18, 35). While this test was originally developed on prostatectomy samples, it was assessed on needle biopsy material for its ability to predict lethal outcome from prostate cancer. Unfortunately, in this study (19) patients with biopsy Gleason scores of 4+3, 8, 9, and 10 were included and made up ~26% of the study population. Arguably, these patients are in need of aggressive therapy and therefore not included in our intended use, clinical need population. The “Prolaris” CCP score showed a HR of 1.65 in multivariate analysis, and while the majority of the performance of the test was derived from the patients with advanced disease, no specific measure of the performance on patients with biopsy scores of 3+3 or 3+4 was reported. Finally, the “Prolaris” test provides a prostate cancer risk score for patients by combining the CCP score with standard clinicopathological parameters, it is not a stand-alone test (18, 33-35).

In the current study, Gleason-scores and TNM staging of matched post-prostatectomy annotations were used as benchmark to evaluate performance of the 8-biomarker assay, instead
of metastatic disease or death as endpoints. This is due to relatively low grade and early stage of
the target population of our test, rendering it necessary to have very long follow-up of very large
and conservatively managed cohorts in order to be sufficiently powered for such study. Indeed,
only few biopsy studies have had metastatic disease or lethal outcome as endpoints. As
expected, to compensate for the power, these studies have included a significant proportion of
high-risk cases (biopsies with Gleason score ≥4+3) where time to death is much shorter;
however, in such cases, the clinical utility of a complementary molecular test is arguably more
limited(19).

Finally, our methodology was based on a binary endpoint derived from prostate pathology that
was shown in a previous clinical study to correlate with lethal outcome (22). The binary endpoint
was chosen to match clinical practice, in which patients are considered for active surveillance or
aggressive treatment. For future investigations, assays could be developed based on either an
ordered categorical (e.g. low, medium, high risk prostate pathology) or a continuous endpoint by
combining inherently categorical prostate pathology assessment with other clinical variables.
Although development of such models is straightforward, they would inherently be based on
subjective weighting and combination of parameters that would require a new assessment of
correlation with long-term outcome before general acceptance.

In conclusion, this study supports further evaluation of this biopsy-based prognostic biomarker
assay for personalized prognostication of prostate cancer and its impact on therapeutic choice.
The ability to provide differential information for the individual patient relative to current risk
stratification systems, where prognostic values are more limited, makes it a potentially useful addition in practice to improve accuracy of clinical decision.

Author’s contribution

Conception and design: P. Blume-Jensen, F. Saad,


Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): P. Blume-Jensen, E. Giladi, J. Dunyak, L. Coupal, D. Berman, J. Epstein, F. Saad


Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): J. Nardone, E. Giladi, J. Dunyak, A. Kaprelyants, T. Nifong

Study supervision: P. Blume-Jensen, F. Saad

Tables and Figures

Table 1. Summary of the Clinical Patient Cohorts in Both Biomarker Development and Clinical Validation Studies, Including Tumors Scored both at Biopsy (Biopsy Gleason) and after Radical Prostatectomy (Surgical Gleason)

<table>
<thead>
<tr>
<th></th>
<th>Biomarker Assay Development Study</th>
<th>Clinical Validation Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total (N)</td>
<td>381</td>
<td>276</td>
</tr>
<tr>
<td>Age, mean (SD), years</td>
<td>60.6 (7.0)</td>
<td>59.9 (5.7)</td>
</tr>
<tr>
<td>PSA, mean (SD), ng/ml</td>
<td>7.7* (9.6)</td>
<td>6.7 (5.7)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Biopsy Gleason score</th>
<th>N</th>
<th>%</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>3+3</td>
<td>162</td>
<td>42.5</td>
<td>191</td>
<td>69.2</td>
</tr>
<tr>
<td>3+4</td>
<td>191</td>
<td>50.1</td>
<td>68</td>
<td>24.6</td>
</tr>
<tr>
<td>3+4**</td>
<td>28</td>
<td>7.3</td>
<td>17</td>
<td>6.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Surgical Gleason score</th>
<th>N</th>
<th>%</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>3+3=6</td>
<td>138</td>
<td>36.2</td>
<td>80</td>
<td>29</td>
</tr>
<tr>
<td>3+4</td>
<td>97</td>
<td>25.5</td>
<td>150</td>
<td>54.3</td>
</tr>
<tr>
<td>4+3</td>
<td>119</td>
<td>31.2</td>
<td>39</td>
<td>14.1</td>
</tr>
<tr>
<td>8</td>
<td>14</td>
<td>3.7</td>
<td>4</td>
<td>1.5</td>
</tr>
<tr>
<td>9</td>
<td>13</td>
<td>3.4</td>
<td>3</td>
<td>1.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pathologic stage</th>
<th>N</th>
<th>%</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>2</td>
<td>0.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2</td>
<td>269</td>
<td>70.5</td>
<td>168</td>
<td>60.8</td>
</tr>
<tr>
<td>T3a</td>
<td>72†</td>
<td>18.9</td>
<td>91</td>
<td>33</td>
</tr>
<tr>
<td>T3b-c</td>
<td>37</td>
<td>9.7</td>
<td>15</td>
<td>5.5</td>
</tr>
<tr>
<td>T4</td>
<td>1</td>
<td>0.3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Missing</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source Institution</th>
<th>N</th>
<th>%</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>University of Montreal</td>
<td>–</td>
<td>–</td>
<td>276</td>
<td>100</td>
</tr>
<tr>
<td>Urology Austin</td>
<td>147</td>
<td>38.6</td>
<td></td>
<td>–</td>
</tr>
<tr>
<td>Chesapeake Urology</td>
<td>8</td>
<td>2.1</td>
<td></td>
<td>–</td>
</tr>
<tr>
<td>Associates</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cleveland Clinic</td>
<td>71</td>
<td>18.6</td>
<td></td>
<td>–</td>
</tr>
<tr>
<td>Folio Biosciences</td>
<td>36</td>
<td>9.4</td>
<td></td>
<td>–</td>
</tr>
<tr>
<td>Michigan Urology</td>
<td>119</td>
<td>31.2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note that, in the clinical validation study, preoperative PSA was missing for eight patients, and clinical staging was missing for 12 patients.
SD denotes standard deviation.
* Excludes two patients with PSA reported at diagnosis of 791 and 600 ng/ml, which is atypical of newly diagnosed patients.
**Patients with discordant Gleason (3+4/4+3) based on centralized Gleason review were
included. See Methods.
† Includes four samples annotated only as T3.
Table 2. Clinical Validation Study: Prognostic Test Performance of the 8-Biomarker Assay against the two Co-primary Endpoints.

<table>
<thead>
<tr>
<th>Population (N)</th>
<th>Endpoint Definition</th>
<th>AUC (95% CI)</th>
<th>P Value (Bonferroni-Adjusted)</th>
<th>OR Lowest-to-Highest-risk Score Quartile (95% CI)</th>
<th>OR as Point Estimate for Continuous Range of Risk Scores (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(N=274)*</td>
<td>Favorable pathology-Surgical Gleason ≤3+4 and organ-confined (≤T2) vs. nonfavorable- surgical Gleason ≥ 4+3 or non-organ-confined (T3a, T3b, N, or M)</td>
<td>0.68 (0.61 to 0.74)</td>
<td>&lt;0.0001</td>
<td>3.3 (1.8 to 6.1)</td>
<td>20.9 (6.4 to 68.2)</td>
</tr>
<tr>
<td>(N=276)</td>
<td>“GS 6”-Surgical Gleason =3+3 and localized ≤T3a vs. “Non-GS 6”- surgical Gleason ≥3+4 or nonlocalized (T3b, N, or M)</td>
<td>0.65 (0.58 to 0.72)</td>
<td>&lt;0.0001</td>
<td>4.2 (1.9 to 9.3)</td>
<td>12.6 (3.5 to 47.2)</td>
</tr>
</tbody>
</table>

* Two patient samples (of 276) lacked sufficient annotation for this calculation.
Table 3A. Clinical Validation Study. Comparison of Predictive Value of the 8-Biomarker Assay for Favorable Pathology. Green = predictive value for favorable disease based on test Risk Score less than 0.33 in the indicated risk category.

<table>
<thead>
<tr>
<th>Biomarker Assay Risk Score Range</th>
<th>Number of Patients According to Biomarker Assay Scores</th>
<th>%PPV (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>Favorable</td>
</tr>
<tr>
<td>≤0.33</td>
<td>61</td>
<td>51</td>
</tr>
<tr>
<td>0.33 to 0.80</td>
<td>156</td>
<td>91</td>
</tr>
<tr>
<td>&gt;0.80</td>
<td>39</td>
<td>9</td>
</tr>
<tr>
<td>All</td>
<td>256</td>
<td>151</td>
</tr>
</tbody>
</table>
Table 3B. Clinical Validation Study. Comparison of Predictive Value of the 8-Biomarker Assay for Favorable Pathology with NCCN Risk Categories. Green = predictive value for favorable disease based on test Risk Score less than 0.33 in the indicated risk category; blue = predictive value for favorable disease by the standard risk category alone.

<table>
<thead>
<tr>
<th>NCCN</th>
<th>Biomarker Assay Risk Score Range</th>
<th>Number of Patients According to Biomarker Assay Scores and Current Classification Systems Categories</th>
<th>%PPV (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very low by NCCN</td>
<td>–</td>
<td>Total: 66, Favorable: 53, Non-favorable: 13</td>
<td>80.3% (68.7% to 89.1%)</td>
</tr>
<tr>
<td>Very low</td>
<td>≤0.33</td>
<td>Total: 20, Favorable: 19, Non-favorable: 1</td>
<td>95% (75.1% to 99.9%)</td>
</tr>
<tr>
<td>Very low 0.33 to 0.80</td>
<td></td>
<td>Total: 42, Favorable: 33, Non-favorable: 9</td>
<td>78.6% (63.2% to 89.7%)</td>
</tr>
<tr>
<td>Very low &gt;0.80</td>
<td></td>
<td>Total: 4, Favorable: 1, Non-favorable: 3</td>
<td>25% (0.6% to 80.6%)</td>
</tr>
<tr>
<td>Low by NCCN</td>
<td>–</td>
<td>Total: 94, Favorable: 60, Non-favorable: 34</td>
<td>63.8% (53.3% to 73.5%)</td>
</tr>
<tr>
<td>Low ≤0.33</td>
<td></td>
<td>Total: 27, Favorable: 22, Non-favorable: 5</td>
<td>81.5% (61.9% to 93.7%)</td>
</tr>
<tr>
<td>Low 0.33 to 0.80</td>
<td></td>
<td>Total: 59, Favorable: 34, Non-favorable: 25</td>
<td>57.6% (44.1% to 70.4%)</td>
</tr>
<tr>
<td>Low &gt;0.80</td>
<td></td>
<td>Total: 8, Favorable: 4, Non-favorable: 4</td>
<td>50% (15.7% to 84.3%)</td>
</tr>
<tr>
<td>Intermediate by NCCN</td>
<td>–</td>
<td>Total: 88, Favorable: 36, Non-favorable: 52</td>
<td>40.9% (30.5% to 51.9%)</td>
</tr>
<tr>
<td>Intermediate ≤0.33</td>
<td></td>
<td>Total: 12, Favorable: 9, Non-favorable: 3</td>
<td>75% (42.8% to 94.5%)</td>
</tr>
<tr>
<td>Intermediate 0.33 to 0.80</td>
<td></td>
<td>Total: 50, Favorable: 23, Non-favorable: 27</td>
<td>46% (31.8% to 60.7%)</td>
</tr>
<tr>
<td>Intermediate &gt;0.80</td>
<td></td>
<td>Total: 26, Favorable: 4, Non-favorable: 22</td>
<td>15.4% (4.4% to 34.9%)</td>
</tr>
<tr>
<td>High by NCCN</td>
<td>–</td>
<td>Total: 8, Favorable: 2, Non-favorable: 6</td>
<td>25% (3.2% to 65.1%)</td>
</tr>
<tr>
<td>High ≤0.33</td>
<td></td>
<td>Total: 2, Favorable: 1, Non-favorable: 1</td>
<td>50% (1.3% to 98.7%)</td>
</tr>
<tr>
<td>High 0.33 to 0.80</td>
<td></td>
<td>Total: 5, Favorable: 1, Non-favorable: 4</td>
<td>20% (0.5% to 71.6%)</td>
</tr>
<tr>
<td>High &gt;0.80</td>
<td></td>
<td>Total: 1, Favorable: 0, Non-favorable: 1</td>
<td>0% (0% to 97.5%)</td>
</tr>
</tbody>
</table>

CI denotes confidence interval; NCCN denotes National Comprehensive Cancer Network; PPV denotes positive predictive value.
**Figure 1.** A simplified overview of biomarker selection, biomarker assay model development, and validation of biomarker assay model

**Figure 2.** Development of the 8-Biomarker Assay

*Figure 2A.* Odds ratios (with 95% confidence interval) for individual biomarkers chosen for the final assay set. Individual quantitative biomarker measurements were correlated with prostate cancer pathology as an endpoint. Note that effect size has been normalized.

*Figure 2B.* Biomarker frequency utilization in top 10% of multivariate assay sets. Frequency of occurrence in the exhaustive top biomarkers search was used as an additional criterion to choose the ultimate markers for the diagnostic assay.

**Figure 3.** Clinical validation study: performance for prediction of favorable pathology.

Sensitivity and specificity curves (A and B, respectively) may be used to identify appropriate risk classification groups. Risk score distribution relative to NCCN risk classification groups (C and D) showing that the biomarker assay adds significant additional risk information within each NCCN classification level.

*Figure 3A.* Sensitivity as a Function of Biomarker Risk Score.

A sensitivity of 90% (95% CI, 82% to 94%) corresponded to a biomarker assay risk score of 0.33, i.e., Sensitivity = P[risk score>0.33 | non-favorable pathology]. Using this risk score, a patient with non-favorable pathology would have a 10% chance of incorrectly receiving a favorable classification (95% CI, 6% to 18%). This false negative might lead to delayed treatment.
Figure 3B. Specificity as a Function of Biomarker Risk Score.

A specificity of 95% (95% CI, 90% to 98%) in this cohort corresponded to a biomarker assay risk score of 0.8, i.e., Specificity = P[risk score ≤ 0.8| favorable pathology] . Using this risk score threshold, a patient with favorable pathology would have a 5% chance of incorrectly receiving a non-favorable classification (95% CI, 2% to 10%). This false positive might lead to overtreatment.

Figure 3C. Biomarker risk score further stratifies tumor pathology within the existing NCCN classification system. Each rhomboid represents an individual patient case, and favorable (blue) and non-favorable (red) categories refer to final classification of patient cases based on prostatectomy pathology. Note the relative concentration of favorable (blue) cases at biomarker risk score <0.33, and the relative concentration of non-favorable (red) cases at biomarker risk score >0.8. The predictive value (+PV) for favorable pathology was 85% at risk score cut-off <0.33. The predictive value (–PV) for non-favorable pathology was 100% at risk score cut-off >0.9, and 76.9% at risk score >0.8.

Figure 3D. Observed frequency of favorable cases as a function of the biomarker risk score quartile. Increased risk score quartile largely correlates with decreased observed frequency of favorable cases in each NCCN category. Observed frequency of patients with favorable pathology identified by the 8-biomarker assay versus the NCCN stratification alone increases
from 0% to 23.8% at a confidence level of 81%.
References

11 candidate biomarker selection from previous study (Ref: Shipitsin, M., et al., 2014 BJC)

- Uniquely developed Quantitative multiplex imaging (QMPI) platform used
- Selected markers had univariate performance for both aggressiveness and lethal outcome
- Marker performances were resistant to sampling variability

Multicenter biomarker assay model development study
Development cohort size (381)

- Using our QMPI platform, marker assay results compared with prostatectomy pathology to calculate odds ratio (OR)
- Biomarker frequency determined in top multivariate assay models
- Identification of 8 biomarker assay model with predictive value

Clinical Validation of biomarker assay model
Validation cohort size (276)

- Validation of 8 biomarker assay model with identification of defined end points of “favourable” vs “non-favourable” as well as “GS6” vs “non-GS6” pathology in patients
- Biomarker assay “Risk scores” comparison with current risk stratification systems (NCCN and D’Amico) with positive predictive value (PPV)
- Calculation of added benefit using Net Reclassification Index (NRI)
Figure 3
Clinical Cancer Research

Development and Clinical Validation of an in situ Biopsy Based Multi-Marker Assay for Risk Stratification in Prostate Cancer

Peter Blume-Jensen, David Berman, David L. Rimm, et al.

Clin Cancer Res  Published OnlineFirst March 2, 2015.

Updated version  Access the most recent version of this article at:
doi:10.1158/1078-0432.CCR-14-2603

Supplementary Material  Access the most recent supplemental material at:
http://clincancerres.aacrjournals.org/content/suppl/2015/03/04/1078-0432.CCR-14-2603.DC1

Author Manuscript  Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

E-mail alerts  Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions  To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions  To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.