Clinical Proof-of-concept of a Novel Platform Utilizing Biopsy-derived Live Single Cells, Phenotypic Biomarkers, and Machine Learning Toward a Precision Risk Stratification Test for Prostate Cancer Grade Groups 1 and 2 (Gleason 3 + 3 and 3 + 4)

David Albala, Michael S. Manak, Jonathan S. Varsanik, Hani H. Rashid, Vladimir Mouraviev, Stephen M. Zappala, Ene Ette, Naveen Kella, Kimberly M. Rieger-Christ, Grannum R. Sant, and Ashok C. Chander

OBJECTIVE To examine the ability of a novel live primary-cell phenotypic (LPCP) test to predict postsurgical adverse pathology (P-SAP) features and risk stratify patients based on SAP features in a blinded study utilizing radical prostatectomy (RP) surgical specimens.

METHODS Two hundred fifty-one men undergoing RP were enrolled in a prospective, multicenter (10), and proof-of-concept study in the United States. Fresh prostate samples were taken from known areas of cancer in the operating room immediately after RP. Samples were shipped and tested at a central laboratory. Utilizing the LPCP test, a suite of phenotypic biomarkers was analyzed and quantified using objective machine vision software. Biomarkers were objectively ranked via machine learning-derived statistical algorithms (MLDSA) to predict postsurgical adverse pathological features. Sensitivity and specificity were determined by comparing blinded predictions and unblinded RP surgical pathology reports, training MLDSAs on 70% of biopsy cells and testing MLDSAs on the remaining 30% of biopsy cells across the tested patient population.

RESULTS The LPCP test predicted adverse pathologies post-RP with area under the curve (AUC) via receiver operating characteristics analysis of greater than 0.80 and distinguished between Prostate Cancer Grade Groups 1, 2, and 3/Gleason Scores 3 + 3, 3 + 4, and 4 + 3. Further, LPCP-derived biomarker scores predicted Gleason pattern, stage, and adverse pathology with high precision—AUCs > 0.80.

CONCLUSION Using MLDSA-derived phenotypic biomarker scores, the LPCP test successfully risk stratified Prostate Cancer Grade Groups 1, 2, and 3 (Gleason 3 + 3 and 7) into distinct subgroups predicted to have surgical adverse pathologies or not with high performance (>0.85 AUC). UROLOGY 124: 198–206, 2019. © 2018 Elsevier Inc.
Historically, Gleason score (GS) 7 is the most common Prostate cancer (PCa) diagnosis next to GS 6 (49% of cases); in a recent study of >1800 PCa patients, 39% were diagnosed as GS 7. A study of 320 GS 7 patients published in 2006 found 75% had the less aggressive form and were classified as (3 + 4) instead of (4 + 3) postprostatectomy. Improved risk stratification (RS) of GS 6 and 7 by differentiating indolent from aggressive GS 6 and GS 7 cancers may reduce rates of overtreatment. This observation is underscored by the new guidelines that the International Society of Urological Pathology recently issued in 2014. The Prostate Cancer Grading Group (PCGG) 1-5 scoring system defines group 1 as GS 6, grade group 2 as GS 3 + 4 = 7, grade group 3 as GS 4 + 3 = 7, grade group 4 as GS 8, and grade group 5 as GS 9-10. While the new PCGGs have demonstrated the ability to more accurately grade stratify than the current Gleason system, the ability to predict postsurgical adverse pathology (P-SAP) features with precision (>0.85%) receiver operator characteristic: area under the curve (AUC), before surgery, for low-risk and indolent prostate cancers (Gleason 3 + 3/PCGG 1 and Gleason 3 + 4/PCGG 2), remains the "holy grail" of prostate risk-stratification. Toward that aim, and given that P-SAP features remain the "gold-standard" of clinical diagnosis, prediction of P-SAP features from radical prostatectomy (RP) specimens has been the basis of validation of a number of new biopsy-based immunohistochemistry, molecular, and genomic biomarker tests that aim to ameliorate the problem of PCa patient RS.

Currently, the primary means to predict P-SAP features and risk stratify patients to guide definitive treatment decisions are the GS, the PCGG scores, as well as genomic tests. Importantly, current methods utilize fixed-biopsy tissue that allows for single time point biomarker measurements. Further, the GS and PCGGs are both based on architectural glandular patterns of the tumor rather than individual cellular characteristics and do not include live-cell phenotypic and biological attributes of individual cancer cells (eg, actin dynamics, cell motility, subcellular protein localization, etc.). Integrating information from a complimentary live-cell prognostic test that leverages innovations in molecular, cellular, and functional dynamic behavioral approaches may improve RS of PCa. Recently, we described a novel patient RS platform capable of predicting P-SAP features—STRAT-AP—that leverages the ability to rapidly culture prostate biopsy cells to measure dynamic and fixed molecular and cellular phenotypic biomarkers with the potential of predicting SAP features. This platform stratified low- and intermediate-risk patients with high sensitivity and specificity in a cohort of ~60 men. The current study (Fig. 1) attempts to validate and demonstrate clinical proof-of-concept of a live primary-cell phenotypic (LPCP) test based on the STRAT-AP technology (Fig. 2) in a sample set of 251 patients by predicting specific and clinically actionable adverse pathology features as well as risk stratifying men based on predicted post-RP GS/PCGG and number of P-SAP features.

In the following study (Fig. 1) we present data that supports the clinical proof-of-concept of the LPCP test (Fig. 1D-F) and its ability to apply novel machine vision to quantify LPCP biomarkers (Fig. 2) and machine learning (Fig. 3) to: (1) predict specific SAP features, (Fig. 3) (2) generate clinical scores capable of risk stratifying patients based on PCGG and GS and number of P-SAP features (Fig. 4) as well as (3) predict biochemical recurrence (BCR) (Supplemental Figure 5), (4) predict stage, (Supplemental Figure 7), and (5) predict predominant postsurgical GS/PCGG, (Supplemental Figure 7) with AUCs>0.80.

MATERIALS AND METHODS

Ten sites (5 university hospitals, 3 large urology group practices, and 2 bio-banks) provided fresh prostate samples after Institutional Review Board approval (Fig. 1A). Samples were obtained from the fresh RP specimens via needle core biopsy or excision from (1) known areas of prostate cancer and (2) well away from the known cancer area, as determined by visual inspection of the prostatectomy sample in the operating room and prior diagnostic biopsy results. Areas of prostate cancer were determined by diagnostic biopsy results and confirmed by a board certified pathologist. Samples presented with the following postsurgical GS percentage distribution of the total sample size in the cohort: PCGG 1/Gleason 3 + 3 (13.87%), PCGG 2/Gleason 7 (3 + 4) (49.16%), PCGG 3/Gleason 7 (4 + 3) (25.21%), PCGG 4/Gleason 8 (5.04%), PCGG 5/Gleason 9 (5.04%), PCGG 5/Gleason 10 (1.68%) and tumor stage of pT2a (7.14%), pT2b (0.84%), pT2c (53.78%), pT3a (14.71%), pT3b (8.4%), pT3c (0.42%) (Fig. 1 B-C).

Cell culturing methods are described in reference and in Supplemental information. Phenotypic biomarkers were measured using differential interference contrast and fluorescence microscopy and described in reference and Supplemental information. Machine vision based measurement of phenotypic biomarkers is described in reference and Supplemental information.

Statistical analysis algorithms (Fig. 3) were used to analyze the data generated from the machine vision image analysis algorithms. Data were independently validated by an external third party statistical analysis group - Anoixis, LLC (Natick, MA).

Machine learning-derived algorithms then ranked individual biomarkers in an unbiased/nonsubjective manner on their ability to predict a given group-of or specific adverse pathology state(s) to determine the exact formal machine learning-derived statistical algorithm (MLDSA) used to for clinical predictions. Machine learning-derived algorithm-mediated biomarker analysis is described in reference and Supplemental information.

RESULTS

Patient Characteristics, RP Analysis, and Reported Postsurgical Adverse Pathology Features

Two hundred fifty-one RP samples were collected in a blinded fashion with Institutional Review Board approval
Figure 1. Patient sample postsurgical adverse pathology distribution: Gleason score, Prostate Cancer Grade Group, stage as well as sample selection criteria and live primary-cell phenotypic (LPCP) test biomarker measurement and analysis workflow. (A) The cohort distribution is shown for unblinded postsurgical Gleason Score and stage of patient samples. (B) Cohort distribution based on tumor stage. (C) Cohort distribution based on Gleason score and PCGG. (D) Sample collection and sample transport to central laboratory workflow. (E) Flow diagram highlighting processing of live-primary samples and sample. (F) Phenotypic, (cellular and molecular) biomarkers are measured on live and subsequently fixed patient samples. The flow diagram outlines the process of the LPCP test’s biomarker measurement, algorithmic analysis, and generation of predictive measurements. Machine vision image analysis measure biomarkers from the images collected via live- and subsequently fixed-primary samples. Machine learning derived statistical algorithms are employed to characterize cellular biomarkers with single-cell resolution. Abbreviations: PCGG, prostate cancer grade group. (Color version available online.)
and exhibited the following distribution for GS/PCGG: 14%—GS 3 + 3/PCGG 1, 49%—Gleason 3 + 4/PCGG 2, 25%—GS 4 + 3/PCGG 3, 5%—GS 8/PCGG 4, and 7%—GS 9 (GS10/PCGG 5). Patients were males between the ages of 40 and 80 years old of any race that did not have a history of infectious disease or were previously treated with chemotherapy (Fig. 1A-C).

Of the 251 samples, only 238 samples were successfully cultured as 13 samples exhibited yeast or bacterial infection during collection or transport and were unable to be cultured. A total of 237 samples were successfully cultured and analyzed after a machine vision-based biomarker quantification error occurred during the analysis process of one sample.

**LPCP BIOMARKER MEASUREMENT USING MACHINE VISION AND ANALYSIS VIA MACHINE LEARNING DERIVED ALGORITHMS**

Based on the STRAT-AP technology, developed by Lanyx Diagnostics (Beverly, MA),15 the LPCP test analyzes primary biopsy tissue biomarkers using both live-cell and fixed-cell imaging techniques to generate both dynamic and static phenotypic biomarker measurements that can be input into machine learning algorithms to predict P-SAP features as well as other clinical end points such as BCR (Fig. 2A). Each patient’s live cells are imaged to assess dynamic cellular and molecular biomarkers including, but not limited to, adhesion, tortuosity, cell motility, membrane dynamics, and actin dynamics (Fig. 2B and C)—indirect measures of cell–extra-cellular matrix (ECM) interactions, cellular mechanics, and signal pathway activation responsible for motility and membrane dynamics. Fixed-biomarkers are subsequently analyzed utilizing immunofluorescence to assess specific protein localization and modification states enabling the interrogation of specific signaling pathways (Fig. 2D and E).

Machine vision software tracks individual cells and correlates both live-cell and fixed-cell LPCP biomarker measurements for each corresponding cell, with single cell resolution, to achieve the quantification of cellular and molecular phenotypic primary and aggregate biomarkers for each cell (Fig. 2F). The LPCP machine learning algorithms, based on random forest, decision tree analysis15 are trained to develop MLDSAs that can predict if a patient is positive or negative for a given P-SAP feature (Fig. 3). MLDSAs then generate threshold values that determine if a patient is positive or negative for a specific or group of P-SAP features allowing prediction performance to be assessed.

**Clinical Workflow and Prediction of Specific Postsurgical Adverse Pathology Features, Postsurgical Stage, Postsurgical GS/PCGG, and BCR**

The LPCP test was designed as a laboratory-developed test that could be run in any CLIA laboratory. Specifically, patient sample characteristics and acquisition were optimized to fit seamlessly in the prostate biopsy workflow of the urologist. Figure 3A summarizes the current surgical workflow for histopathological analysis and postsurgical GS/PCGG reporting compared to the LPCP test workflow (Fig. 3B).

Specific postsurgical adverse pathways (ie, EP, positive surgical margins [PSM], and seminal vesicle invasion) were predicted and compared to unblinded postsurgical report findings. The performance (AUCs > 0.80) of these predictions is summarized in Figure 3C and ROC curves are presented in Supplemental Figure 3. Specifically, prediction of P-SAP features such as PSM, extra-prostatic extension (EPE), seminal vesicle invasion, and lymph node positive exhibited AUCs > 0.85. Importantly, when field samples containing nontumor or field, tissue was analyzed on MLDSAs trained on tumor/cancer tissue the MLDSAs demonstrated suboptimal AUCs of 0.57-0.74 for specific P-SAP features (PSM, EPE, Peri-neural invasion (PNI), lymph node positive, and lympho-vascular invasion (LVI)) confirming that the cancer trained MLDSAs are specific to tumor tissue and not applicable to nontumor cells (Supplemental Figure 3). Interestingly, the LPCP test was also able to predict BCR (Supplemental Figure 5) as well as postsurgical stage and postsurgical GS/PCGG with AUCs > 0.80 (Supplemental Figure 7). Similar to the prediction of specific P-SAP features, the tumor-trained MLDSAs were specific to tumor-samples and performed suboptimally when applied to nontumor/field tissue supporting the idea that tumor-based MLDSAs are specific to tumor samples and not random sampling of the prostate (Supplemental Figure 7).
Figure 3. Quantified biomarkers are input into random forest decision trees to generate Machine Learning Derived Statistical Algorithms (MLDSAs) that subsequently generate predictions of postsurgical adverse pathology features. (A) Biomarker measurements, quantified from machine vision analysis are recorded in a database that is input into machine learning algorithms to train, and subsequently test, Machine Learning Derived Statistical Algorithms (MLDSAs). (B) MLDSAs are based on random forest decision tree analysis. (C) MLDSAs predict generate predictions of clinically relevant postsurgical adverse pathology features. Clinical workflow of Live-Primary Cell Phenotypic test to standard histology/Gleason Score/Prostate Cancer Grading Group workflow and predictive performance for post-surgical adverse pathology features. (D) Clinical workflow comparison of histopathology (Gleason/PCGG) analysis with the LPCP test. (E) The LPCP test requires only one biopsy compared to 12 and takes 72 hours to return results after being processed in a central laboratory via high-content microscopy, machine vision and machine learning-derived statistical algorithms (MLDSAs). (F) Statistical performance table summarizing the LPCP test’s ability to predict specific postsurgical adverse pathology features. Abbreviations: LPCP, live-primary cell phenotypic; PCGG, prostate cancer grade group. (Color version available online.)
Predictive Performance of MAPP, and LAPP Scores (Adverse Pathology Scores) and Enhanced RS of GS 3 + 3/PCGG 1 and GS 3 + 4/PCGG 2 Patients Based on LAPP and MAPP Scores

Toward enhanced patient risk-stratification, P-SAP scores were developed and defined as: Local Adverse Pathology Potential (LAPP) score (prediction of EPE, seminal vesicle invasion, and positive surgical margins), and Metastatic Adverse Pathology Potential (MAPP) score (prediction of only perineural invasion, lympho-vascular invasion, and lymph node involvement). Both the LAPP and MAPP score incorporate features that are used in clinical practice. Given such a definition, the data support that LAPP and MAPP enable a clinically meaningful identification of normal vs malignant tissue and enhanced risk-stratification of indolent and aggressive patients within GS 3 + 3/PCGG 1 and GS 3 + 4/PCGG 2. To demonstrate RS based on adverse pathologies at the time of RP, a plot of LAPP vs MAPP (Fig. 4) of a random selection of 100 patient data points (representative set of data of full data set [Supplemental Figure 4]) demonstrates the ability of the diagnostic platform to stratify samples based on the number of postsurgical adverse pathological features noted on the unblended, respective RP surgical pathology reports (Fig. 4 and Supplemental Figure 4). The plots highlight how the thresholds produced by the MLDSAs can be used to risk stratify patients that have similar postsurgical GSs/PCGGs yet end up having different associated postsurgical adverse pathologies for example, distinguishing a GS 3 + 3/PCGG 1 patient with SAP features from a GS 3 + 3/PCGG 1 patient without P-SAP features.

DISCUSSION

In this study we utilized a first-in-class LPCP test to measure a suite of novel phenotypic and dynamic biomarkers in live prostate cells with single cell resolution. This LPCP test was assessed for its ability to (1) predict specific P-SAP features and (2) risk stratify GS shows how these patient samples can be risk stratified using the LPCP test scoring metrics — LAPP and MAPP. (C) LAPP and MAPP scatter plot comparing Gleason 3 + 3/PCGG 1 samples to Gleason 8/PCGG 4 or higher samples confirms that these patient samples can be risk stratified using the LPCP based scoring metrics of LAPP and MAPP. Importantly, samples with no adverse pathologies cluster below the threshold scores in the lower left quadrant regardless of Gleason score. Samples with more adverse pathologies cluster toward the right and above the threshold score. In each case the level of detail provided by the LPCP’s scoring metrics supplements the Gleason score/PCGG. Abbreviations: AP, post surgical adverse pathology; LPCPs, live primary-cell phenotypic; MLDSAs, machine learning derived statistical algorithms; PCGG, prostate cancer grade group. (Color version available online.)
3 + 3/PCGG1 and GS 3 + 4/PCGG 2 patients for adverse pathology to demonstrate the clinical proof-of-concept of a prostate cancer RS test.

The primary limitation of this study is that the same patient samples were used as training and test sets. That said, given that the LPCP test measures biomarkers with single-cell resolution, each training/test data point can be viewed, and thought of, as a single cell which enables the randomized grouping of 70% of cells from the patient population to train MLDSAs and subsequent blinded utilization of the remaining distinct 30% of cells as the test set. Importantly, a major limitation of the current study is that test data sets (single cell data) are taken from the same samples that were included in our test set and work is being undertaken to train this model on a wider set of samples, expanding the range of samples for which a more robust prediction can be generated based on completely different training and test sets. Further, this study only collected patient’s samples from 10 different sites within the United States and only analyzed 237 patient samples to generate sensitivity and specificity measurements in a short-term, randomized prospective study using blinded P-SAP results after analyzing RP surgical samples. Additionally, another limitation of the presented study is that clinical patient data, such as presurgical stage and GS, age, and race, were not incorporated into the MLDSAs. Finally, given our observation that tumor-trained MLDSAs were unable to predict specific P-SAP features when analyzing field/nontumor samples, the LPCP test requires tumor rich sample to provide accurate predictions. Results from a study to train new MLDSAs on field, or nontumor, are forthcoming.

New and emerging molecular and genomic tests aimed at risk stratifying patients have used the prediction of adverse pathology as the end point of their respective proof-of-concept studies followed by subsequent long-term outcome studies. This proof-of-concept study demonstrated that the LPCP test and its suite of phenotypic biomarkers15 improved upon prediction of P-SAP features and existing PCa RS measures. Analysis of 237 patients yielded strong performance with the ability to predict if a GS 3 + 3/PCGG 1 or GS 3 + 4/PCGG 2 patient will experience P-SAP (with AUC>0.80) without the use of other patient data such as GS, patient age, or PSA level. Addition of these demographic and clinical data may in fact further increase the performance of LPCP and is the subject of continued investigation. The LPCP test’s strong performance may be attributed to not only the measurement of biomarkers with single-cell resolution, but the ability to measure dynamic biomarkers and cell behavior objectively with machine vision (Fig. 2) as well as unbiased biomarker ranking by machine learning algorithms (Fig. 3) that predict P-SAP features and risk stratify GS 3 + 3/PCGG 1 and GS 3 + 4/PCGG2 patients (Figs. 2C and 4).15

Interestingly, comparison of predictive performance when using biopsy tissue from noncancerous tissue demonstrated that LAPP, and MAPP based predictions are specific to cancerous tissue (Supplemental information, Supplemental Figures 7 and 8).

Scores derived from phenotypic biomarkers assessing tumor aggressiveness (ie, LAPP), metastatic invasion potential (ie, MAPP), or combined as General Adverse Pathology Potential (GAPP) (Supplemental Figure 3) may then provide objective and actionable insight for treatment selection for GS 3 + 3/PCGG 1 and GS 3 + 4/PCGG 2 patients whether it be active surveillance or definitive therapy with surgical intervention or radiotherapy.

This novel live-cell phenotypic diagnostic platform provides unique, and previously inaccessible, cellular and molecular biomarker information to assess temporal and spatial cellular and subcellular features that predict pathobiological behavior and adverse pathology. The phenotypic biomarkers measured using the LPCP test resulted in statistically significant predictions of P-SAP features in RP specimens. The LPCP test, as a member of a new class of diagnostic/prognostic tests, generated clinically predictive scores: LAPP, and MAPP, for the RS of PCa, and has the potential to improve personalized prognosis and treatment guidance in men with newly diagnosed prostate cancer.

The strong performance demonstrated by this novel live-cell phenotypic biomarker test in predicting adverse pathologies at the time of RP supports a similar follow-up study from patient biopsy samples prior to RP. A multi-center prostate needle biopsy is planned to confirm the findings of the presented proof of concept study that utilized prostate tissue from RP specimens. This will further support the LPCP test’s ability to better understand the dynamic and heterogeneous biology of PCa toward precision RS and aid in personalized treatment decision-making.

Acknowledgment. In addition to Lahey Hospital and Medical Center, American Medical Professionals of New York, and Urology Place of San Antonio, tissue samples were provided by the NCI Cooperative Human Tissue Network (CHTN). Authors respectfully acknowledge: Dr. Hui-May Chu for thoughtful review of the manuscript and Dr. Pravin Chaturvedi, Mr. Robert Gottlieb, and Ms. Mary Beth Lismam for discussion as well as Dr. Mani Foroohar, and Dr. Travis B Sullivan. MSM acknowledges Jenn A. Manak, Chloe A. Manak, Gabriel H. Manak, Paul L. Manak, Linda L. Manak, and Paul W. Manak for support. JSV acknowledges Sarah H. Varsanik for support. ACC acknowledges Coimbatore S. Chandrasekaran, Achamma C. Chandrasekaran, Angela B. Cravens Chander, and Indira E.A. Chander for support.

SUPPLEMENTARY MATERIALS

Supplementary material associated with this article can be found, in the online version, at https://doi.org/10.1016/j.urology.2018.06.068.
References


