Gene Expression Profiling and Protein Biomarkers for Prostate Cancer Management

Policy #: 00403
Original Effective Date: 02/19/2014
Current Effective Date: 06/21/2017

Applies to all products administered or underwritten by Blue Cross and Blue Shield of Louisiana and its subsidiary, HMO Louisiana, Inc. (collectively referred to as the “Company”), unless otherwise provided in the applicable contract. Medical technology is constantly evolving, and we reserve the right to review and update Medical Policy periodically.

Note: Gene-Based Tests for Screening, Detection, and/or Management of Prostate Cancer is addressed separately in medical policy 00272.

Services Are Considered Investigational
Coverage is not available for investigational medical treatments or procedures, drugs, devices or biological products.

Based on review of available data, the Company considers gene expression analysis and protein biomarker to guide management of prostate cancer in all situations to be investigational.*

Background/Overview
Gene expression profile analysis and protein biomarkers have been proposed as a means to risk-stratify patients with prostate cancer to guide treatment decisions. These tests are intended to be used either on prostate needle-biopsy tissue to guide management decisions regarding active surveillance versus therapeutic intervention, or after radical prostatectomy (RP) to guide radiotherapy (RT) use. Two gene expression profiling tests, Prolaris® and Oncotype Dx® Prostate, are intended to be used in combination with accepted clinical criteria (Gleason score, prostate-specific antigen [PSA], clinical stage) to stratify needle biopsy–diagnosed localized prostate cancer according to biological aggressiveness, and direct initial patient management. The Promark™ protein biomarker test uses immunofluorescence and automated quantitative images in intact biopsy tissue to risk stratify patients to active surveillance or therapeutic intervention.

Prostate cancer is the second most common cancer diagnosed among men in the U.S. According to the National Cancer Institute (NCI), nearly 180,000 new cases are expected to be diagnosed in the United States in 2016 and are associated with approximately 26,000 deaths. Autopsy studies in the pre-PSA screening era have identified incidental cancerous foci in 30% of men 50 years of age, with incidence reaching 75% at age 80 years. However, between 1975 and 1991 prostate cancer mortality rose and subsequently dropped 39% by 2007. The rise in mortality is unexplained but is suggested to be due to how cause of death was assigned. Regarding the subsequent decline, a number of potential explanations have been suggested as underlying reasons: improvements in treatment and screening, changes in assigning causes of death, and risk of cardiovascular death among men with prostate cancer treated with hormonal therapy.

FDA or Other Governmental Regulatory Approval
U.S. Food and Drug Administration (FDA)
Clinical laboratories may develop and validate tests in-house and market them as a laboratory service; laboratory-developed tests (LDTs) must meet the general regulatory standards of the Clinical Laboratory

©2017 Blue Cross and Blue Shield of Louisiana
An independent licensee of the Blue Cross and Blue Shield Association
No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, or otherwise, without permission from Blue Cross and Blue Shield of Louisiana.
Page 1 of 33
Gene Expression Profiling and Protein Biomarkers for Prostate Cancer Management

Policy #       00403
Original Effective Date: 02/19/2014
Current Effective Date: 06/21/2017

Improvement Act (CLIA). Prolaris, Oncotype Dx Prostate and Decipher gene expression profiling, and the ProMark protein biomarker test are available under the auspices of CLIA. Laboratories that offer LDTs must be licensed by CLIA for high-complexity testing. To date, the U.S. FDA has chosen not to require any regulatory review of this test.

In November 2015, the FDA’s Office of Public Health Strategy and Analysis published a document on public health evidence for FDA oversight of LDTs. FDA argued that many tests need more FDA oversight than the regulatory requirements of CLIA. CLIA standards relate to laboratory operations, but do not address inaccuracies or unreliability of specific tests. Prolaris is among the 20 case studies in the document cited as needing FDA oversight. The document asserted that patients are potentially receiving inappropriate prostate cancer care because there is no evidence that results from the test meaningfully improve clinical outcomes.

Centers for Medicare and Medicaid Services (CMS)
Palmetto GBA, a local carrier, issued “limited coverage” determinations under the auspices of a “Coverage with Data Development” mechanism for the following tests (date effective): Prolaris (03/02/15), Decipher (03/02/15), Oncotype DX Prostate (10/05/15), and ProMark (10/10/2016).

Rationale/Source
This policy was based initially on a 2014 Technology Evaluation Center (TEC) Assessment addressing disease detected on needle biopsy, which has been supplemented by a 2015 TEC Assessment addressing high-risk disease postprostatectomy. The most recent MEDLINE database search was through October 4, 2016; publications were also submitted for consideration by test suppliers.

Full-length publications were sought that described the analytic validity (technical performance), clinical validity (prognostic accuracy), and clinical utility (accurately identifying men experiencing improved health outcomes by avoiding treatment or undergoing more appropriate therapies) of Prolaris, Oncotype DX Prostate, the ProMark protein biomarker test, and Decipher gene expression profiling.

The level of evidence (LOE) will be evaluated using the Simon et al framework for study classification and LOE for prognostic studies using archived specimens. Category A studies are prospective, randomized trials designed to evaluate prognostic markers; 1 such study would establish LOE 1. Category B studies are prospective trials designed for another purpose with retrospective analysis of archives sample that is prospectively described (“prospective-retrospective” studies); 2 or more such studies are required for LOE 1. Category C studies are prospective observational registry studies with treatment and follow-up not dictated. As noted by Simon et al, studies considered category C are LOE III but “may be validated to LOE II if 2 or more subsequent studies provide similar results. However, it is unlikely that category C studies would ever be sufficient to change practice, except under particularly compelling circumstances.” Category D studies are retrospective in design and represent LOE IV and V.
INITIAL MANAGEMENT DECISION: ACTIVE SURVEILLANCE VS THERAPEUTIC INTERVENTION
Clinical Context and Test Purpose
In men newly diagnosed with clinically localized prostate cancer, the purpose of gene expression profiling and protein biomarkers tests is to inform a decision whether to undergo immediate therapy versus forego immediate therapy and begin active surveillance.

Localized prostate cancers may appear very similar clinically at diagnosis. However, they often exhibit diverse risk of progression that may not be captured by clinical risk categories (eg, D’Amico criteria) or prognostic tools based on clinical findings, including PSA titers, Gleason grade, or tumor stage. In studies of conservative management, the risk of localized disease progression based on prostate cancer–specific survival rates at 10 years may range from 15% to 20% to perhaps 27% at 20-year follow-up. Among elderly men (ages ≥70 years) with low-risk disease, comorbidities typically supervene as a cause of death; these men will die with prostate cancer present, rather than from the cancer itself. Other very similar appearing low-risk tumors may progress unexpectedly rapidly, quickly disseminating and becoming incurable.

The divergent behavior of localized prostate cancers creates uncertainty whether to treat immediately. A patient may choose potentially curative treatment upfront. Surgery including RP or external-beam radiotherapy (EBRT) is most commonly used to treat patients with localized prostate cancer. Complications most commonly reported with RP or EBRT and with the greatest variability are incontinence (0%-73%) and other genitourinary toxicities (irritative and obstructive symptoms); hematuria (typically ≤5%); gastrointestinal and bowel toxicity, including nausea and loose stools (25%-50%); proctopathy, including rectal pain and bleeding (10%-39%); and erectile dysfunction, including impotence (50%-90%).

American Urological Association (AUA) guidelines have suggested patients with low- and intermediate-risk disease may opt for “active surveillance,” taking into account patient age, patient preferences, and health conditions related to urinary, sexual, and bowel function. With this approach, the patient will forgo immediate therapy and continue regular monitoring until signs or symptoms of disease progression are evident, at which point curative treatment is instituted. In the Prostate Testing for Cancer and Treatment (ProtecT) trial, active surveillance, RP, and EBRT for the treatment of clinically localized prostate cancer were compared in 1643 men who were identified through PSA testing. At a median of 10-year follow-up, prostate cancer–specific mortality was low and similar across the 3 treatment groups: 1.5 (95% CI, 0.7 to 3.0) deaths per 1000 person-years in active surveillance, 0.9 (95% CI, 0.4 to 2.2) per 1000 person-years in the surgery group, and 0.7 (95% CI, 0.3 to 2.0) per 1000 person-years in the RT group. Surgery and RT were associated with lower incidences of disease progression and metastases compared to active surveillance. Approximately 55% of men in the active surveillance group had received a radical treatment by the end of follow-up. Earlier results from the Prostate Cancer Intervention versus Observation Trial (PIVOT) also concluded that RP did not prolong survival compared to observation through 12 years of follow-up. An observational study comparing sexual function of men with low-risk prostate cancer who chose active surveillance versus men who received RT or RP found those who chose active surveillance were more often sexually active than similar men who received RP. In a report of quality of life for men in the
Scandinavian Prostate Cancer Group Study Number 4 (SPCG-4), after a median follow-up of more than 12 years, distress caused by treatment-related side effects was reported significantly more often by men assigned to RP than by men assigned to watchful waiting.

Given the unpredictable behavior of early prostate cancer additional prognostic methods to biologically stratify this disease are needed.

The first question addressed in this evidence review is: Does gene expression profiling, compared with clinicopathologic risk stratification or when used with clinicopathologic risk stratification, improve outcomes in newly diagnosed men with clinically localized prostate cancer?

The following PICOTS were used to select literature to inform this review.

Patients
The relevant population of interest is individuals with newly diagnosed, localized prostate cancer, who have not undergone treatment for prostate cancer, and who are deciding between therapeutic intervention or active surveillance.

Intervention
Gene expression profiling refers to analysis of mRNA expression levels of many genes simultaneously in a tumor specimen, and protein biomarkers. Two gene expression profiling tests and 1 protein biomarker test are intended to biologically stratify prostate cancers diagnosed on prostate needle biopsy: Prolaris (Myriad Genetics, Salt Lake City, UT) and Oncotype DX Prostate Cancer Assay (Genomic Health, Redwood City, CA) are gene expression profiling tests that use archived tumor specimens as the mRNA source, reverse transcriptase polymerase chain reaction amplification, and the TaqMan low-density array platform (Applied Biosystems, Foster City, CA). A protein biomarker test, ProMark (Metamark Genetics, Cambridge, MA), is an automated quantitative imaging method to measure protein biomarkers by immunofluorescent staining in defined areas in intact formalin-fixed paraffin-embedded biopsy tissue, in order to provide independent prognostic information to aid in the stratification of patients with prostate cancer to active surveillance or therapy.

Comparator
Clinicopathologic risk stratification is currently being used to make decisions about prostate cancer management. Clinical characteristics (eg, stage, biopsy Gleason grade, serum PSA) and demographic characteristics (eg, as age, life expectancy) are combined to classify men according to risk. National Comprehensive Cancer Network (NCCN) and AUA provide treatment recommendations based on risk stratification. The Kattan et al (2003) nomogram was developed to predict risk of indolent cancer in a low-risk population considering active surveillance. The Cancer of the Prostate Risk Assessment (CAPRA) is a pretreatment nomogram that provides risk prediction of outcomes following RP developed from a cohort of RP patients.
Outcomes
Beneficial outcomes resulting from a true test result are prolonged survival, improved quality of life, and reduction in unnecessary treatment-related adverse effects. Harmful outcomes resulting from a false test result are recurrence, metastases or death, and unnecessary treatments. The outcomes of interest are listed in Table 1.

Table 1. Outcomes of Interest for Individuals With Newly Diagnosed, Localized Prostate Cancer

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall survival</td>
<td>10-year survival</td>
</tr>
<tr>
<td>Disease-specific survival</td>
<td>10-year prostate cancer – free survival; 10-year prostate cancer death rate; 10-year recurrence rate</td>
</tr>
<tr>
<td>Quality of life</td>
<td>See Chen et al (2014) for NCI-recommended health-related quality of life measures for localized prostate cancer</td>
</tr>
<tr>
<td>Treatment-related morbidity</td>
<td>Adverse effects of radiotherapy or radical prostatectomy</td>
</tr>
</tbody>
</table>

NCI: National Cancer Institute.

Time
Ten-year outcomes are of interest due to the prolonged natural history of localized prostate cancer.

Setting
Decisions about management of localized prostate cancer are generally made by patients and urologists in the secondary or tertiary care setting.

Prolaris
Prolaris is used to quantify expression levels of 31 cell cycle progression (CCP) genes and 15 housekeeper genes to generate a CCP score. Oncotype DX Prostate is used to quantify expression levels of 12 cancer-related and 5 reference genes to generate a Genomic Prostate Score (GPS). In the final analysis, the CCP score or GPS is combined in proprietary algorithms with clinical risk criteria (PSA, Gleason grade, tumor stage) to generate new risk categories (ie, reclassification) intended to reflect biologic indolence or aggressiveness of individual lesions, and thus inform management decisions. This section will review Prolaris for initial management decisions in newly diagnosed, localized cancer. Prolaris for management after RP will be discussed in the following section.

Analytic Validity
Although there is no reference standard for gene expression profiling tests, other measures of technical performance are relevant and include reproducibility, tissue-sample adequacy, potential batch effects, and test-set bias. Warf et al (2015) evaluated the precision of the CCP score using 6 formalin-fixed, paraffin-embedded (FFPE) biopsy (3 replicate scores) and 12 FFPE RP (4-6 replicate scores) specimens. Overall precision was estimated from replicate samples, intended to reflect combined variation from tissue dissection through gene expression. Across replicate samples, the standard deviation of the CCP score was 0.1 (95% confidence interval [CI], 0.98 to 0.13). After 8 weeks of sample storage, results were similar. In 2013, Myriad Genetics reported 95.3% of samples were adequate to produce a CCP score. Information is available on the performance of the TaqMan array platform used in Prolaris and Oncotype Dx Prostate through the MicroArray Quality Control (MAQC) project. In the MAQC project, which was initiated and led by
Gene Expression Profiling and Protein Biomarkers for Prostate Cancer Management

Policy #       00403
Original Effective Date: 02/19/2014
Current Effective Date: 06/21/2017

FDA scientists, expression data on 4 titration pools from 2 distinct reference RNA samples were generated at multiple test sites on 7 microarray-based and 3 alternative technology platforms, including TaqMan. According to the investigators, the results provide a framework to assess the potential of array technologies as a tool to provide reliable gene expression data for clinical and regulatory purposes. The results showed very similar performance across platforms, with a median coefficient of variation of 5% to 15% for the quantitative signal and 80% to 95% concordance for the qualitative detection call between sample replicates.

Clinical Validity
Two studies reporting clinical validity related to newly diagnosed men with clinically localized prostate cancer were included as outlined in Table 2.

Table 2. Studies Reporting Clinical Validity of Prolaris for informing initial management decisions

<table>
<thead>
<tr>
<th>Study</th>
<th>Design</th>
<th>Outcome</th>
<th>Dates</th>
<th>Sites</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cuzick et al (2015)</td>
<td>Retrospective cohort from prospective registry (Simon category C)</td>
<td>PC death</td>
<td>1990-2003</td>
<td>3 UK registries(^a)</td>
</tr>
</tbody>
</table>

PC: prostate cancer.
\(^a\) No overlap in population with Cuzick et al (2012).

Cuzick et al (2012) examined the Prolaris prognostic value for prostate cancer death in a conservatively managed needle biopsy cohort. Cell cycle expression data were read blind to all other data. Patients were identified from 6 cancer registries in Great Britain and were included if they had clinically localized prostate cancer diagnosed by needle biopsy between 1990 through 1996; were younger than 76 years at diagnosis; had a baseline prostate-specific antigen (PSA) measurement; and were conservatively managed. Potentially eligible patients who underwent RP, died, showed evidence of metastatic disease within 6 months of diagnosis, or received hormone therapy before diagnostic biopsy were excluded. The original biopsy specimens were retrieved and centrally reviewed by a panel of expert urologic pathologists to confirm the diagnosis and, where necessary, to reassign Gleason scores. Of 776 patients diagnosed by needle biopsy and for which a section was available to review histology, needle biopsies were retrieved for 527 (68%), 442 (84%) of which had adequate material to assay. From the 442 samples, 349 (79%) produced a CCP score and had complete baseline and follow-up information, representing 45% of 776 patients initially identified. The median follow-up time was 11.8 years. A total of 90 deaths from prostate cancer occurred within 2799 person-years.

The primary, unadjusted analysis found a 1-unit increase in CCP score associated with a 2-fold increase (hazard ratio [HR], 2.02; 95% CI, 1.62 to 2.53) in the risk of dying from prostate cancer (see Table 2). In a multivariate model including CCP, Gleason score, and PSA level, the adjusted hazard ratio for a 1-unit increase in CCP score was 1.65 (95% CI, 1.31 to 2.09). Death rates were low in this group and larger cohorts are required to assess fully the value of the CCP combined score. Kaplan-Meier analyses of 10-year risk of prostate cancer death stratified by CCP score groupings are shown in Table 3. Confidence intervals were not reported. Cuzick et al (2012) did not explain the apparent substantial difference in
mortality rates among patients in the 0 ≤ CCP ≤ 2 grouping (range, 19.3%-21.1%) and those in the 2< CCP ≤ 3 and >3 groupings (range, 48.2%-74.9%). The difference may reflect clinical criteria (eg, proportions of lower vs higher Gleason grade cancers, respectively). Measures that would suggest improved discriminatory ability (eg, area under the curve [AUC] or reclassification) were not reported in Cuzick et al (2012). The authors did not provide evidence that the test could correctly reclassify men initially at high risk to lower risk to avoid overtreatment, or conversely, correctly reclassify those initially at low risk to high risk to avoid undertreatment.

Cuzick et al (2015) examined 3 U.K. cancer registries from 1990 to 2003 to identify men with prostate cancer who were conservatively managed following needle biopsy, with follow-up through December 2012. The authors stated that the samples did not overlap with Cuzick et al (2012). Men were excluded if they had undergone RP or RT within 6 months of diagnosis. A combination of the CCP and CAPRA scores was used to predict prostate cancer death. There were 989 men who fit eligibility criteria; CCP scores were calculable for 761 (77%) and combined CCP and clinical variables were available for 585 (59%). Median age at diagnosis was 70.8 years and median follow-up was 9.5 years. The prostate cancer mortality rate was 17% (n=100), with 29% (n=168) dying from competing causes. Higher CCP scores were associated with increased 10-year risk of prostate cancer mortality (see Table 4): 7% (CCP score <0), 15% (CCP score 0-1), 36% (CCP score 1-2), 59% (CCP score >2). A 1-unit increase in CCP was associated with a crude hazard ratio for death of 2.08 (95% CI, 1.76 to 2.46) and when adjusted for CAPRA score yielded a hazard ratio of 1.76 (95% CI, 1.47 to 2.14). For the combined CAPRA/CCP score, the hazard ratio for 10-year prostate cancer mortality increased to 2.17 (95% CI, 1.83 to 2.57). The C statistic for the CAPRA score was 0.74; adding the CCP score increased the C statistic to 0.78 (no confidence intervals for the C statistic were reported). Treatment changes after 6 months were documented in only part of 1 of the 3 cohorts; at 24 months, 45% of the men in this cohort had undergone RT or prostatectomy.

### Table 3. Univariate and Multivariate Association Between CCP and Death From Prostate Cancer in the Cuzick (2012) and Cuzick (2015) Validation Studies

<table>
<thead>
<tr>
<th>Study</th>
<th>N</th>
<th>Unadjusted Hazard Ratio (95% Confidence Interval)</th>
<th>Multivariate Hazard Ratio (95% Confidence Interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cuzick et al (2012)</td>
<td>349</td>
<td>2.02 (1.62 to 2.53)</td>
<td>1.65 (1.31 to 2.09)&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Cuzick et al (2015)</td>
<td>585</td>
<td>2.08 (1.76 to 2.46)</td>
<td>1.76 (1.47 to 2.14)&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup> Adjusted for Gleason score and prostate-specific cancer level.

<sup>b</sup> Adjusted for Cancer of the Prostate Risk Assessment.

### Table 4. Kaplan-Meier Estimates of Prostate Cancer Death at 10 Years by CCP Score Groupings in the Cuzick (2012) and Cuzick (2015) Validation Studies

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>10-Year Death Rate, %&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>≤0</td>
<td>36</td>
<td>19.3</td>
</tr>
<tr>
<td>0 to ≤1</td>
<td>133</td>
<td>19.8</td>
</tr>
<tr>
<td>1 to ≤2</td>
<td>114</td>
<td>21.1</td>
</tr>
<tr>
<td>2 to ≤3</td>
<td>50</td>
<td>48.2</td>
</tr>
<tr>
<td>&gt;3</td>
<td>16</td>
<td>74.9</td>
</tr>
</tbody>
</table>

<sup>a</sup> No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, or otherwise, without permission from Blue Cross and Blue Shield of Louisiana.
Gene Expression Profiling and Protein Biomarkers for Prostate Cancer Management

Systematic Reviews
In 2016, results of a systematic review and meta-analysis supported by the manufacturer were reported. Published and unpublished studies of prognostic validity or clinical utility of CCP testing were eligible for inclusion. Seven published studies were identified; 5 were clinical validity studies. Two were reviewed in the previous paragraphs and the remaining validity studies will be reviewed in a subsequent section on post-RP management. The other 2 “utility” studies are discussed in the following section. Two validity studies reported outcomes for disease-specific mortality but of the 2 only the Cuzick et al (2012) included newly diagnosed patients, so the pooled outcome is not of relevance in this section.

Clinical Utility
We identified no studies to directly support the clinical utility of Prolaris. Three decision-impact studies assessed the potential impact of Prolaris on physicians’ treatment decisions in patients. The authors of each study (Crawford et al, 2014; Shore et al, 2014; Shore et al, 2016) have suggested that their findings support the “clinical utility” of the test, based on whether the results would lead to a change in treatment. Pathology results were not reported for these studies. Given the lack of established clinical validity and no reported outcomes, it is not known whether any treatment changes were clinically appropriate.

Section Summary: Prolaris
Analytic validity of gene expression analysis for prostate cancer management using Prolaris was reported by Warf et al (2015) and supported by results from the MAQC project.

In a cohort of men conservatively managed following needle biopsy, Cuzick et al (2012) suggested that the CCP score alone was more prognostic than either PSA or Gleason score for tumor-specific mortality at 10-year follow-up based on hazard ratios. Comparison to CAPRA was not provided in Cuzick et al (2012). Cuzick et al (2015) found that discrimination improved somewhat by adding the CCP score to the CAPRA score, as reflected in the C statistic. Validation studies were Simon category C.

No direct evidence is available to support the clinical utility of Prolaris for improving net outcomes of patients with localized prostate cancer. The indirect chain of evidence is also incomplete. Simon category C studies are not sufficient to determine with confidence prognosis of CCP score or whether there is incremental improvement when combined with CAPRA. Ten-year prostate cancer-specific survival outcomes from a Simon category A, or multiple, independent Simon category B studies are needed. Overall the 10 year mortality is low for localized prostate cancer. Therefore to improve the net outcome, a biomarker test would need to lead to large improvements in risk stratification compared to existing tools such as CAPRA.

Oncotype DX Prostate
The Oncotype Dx Prostate assay includes 5 reference genes and 12 cancer genes that represent 4 molecular pathways of prostate cancer oncogenesis: androgen receptor, cellular organization, stromal
Gene Expression Profiling and Protein Biomarkers for Prostate Cancer Management

Policy #       00403
Original Effective Date: 02/19/2014
Current Effective Date: 06/21/2017

response, and proliferation. The assay results are combined to produce a GPS, which ranges from 0 to 100. Higher GPS scores indicate more risk.

Analytic Validity
Knezevic et al (2013) reported on the analytic validity of Oncotype DX Prostate. Estimates of analytic precision and reproducibility were derived from analysis of RNA prepared from 10 microdissected prostate tumor samples obtained by needle biopsy. Individual Gleason scores were assigned using the 2005 International Society of Urological Pathology consensus guidelines.

The results showed that the assay could accurately measure expression of the 12 cancer-related and 5 reference genes over a range of absolute RNA inputs (0.005-320 ng); the limit of detection in a sample was 0.5 ng/μL. The analytic accuracy showed average variation of less than 9.7% across all samples at RNA inputs typical of needle-biopsy specimens. The amplification efficiency for the 17 genes in the test ranged from 88% to 100%, with a median (SD) of 93% (6%) for all 17 genes in the assay. Analytic precision was assessed by examining variability between replicate results obtained using the same mRNA input. Reproducibility was measured by calculating both within- and between-mRNA input variation. A low input level of mRNA 5 ng was used to reflect the lowest 2.5 percentile of a tumor sample of 0.023 cm³. When converted to Genomic Prostate Score (GPS) units (unit measure for reporting test results), the standard deviation for analytic precision was 1.86 GPS units (95% CI, 1.60 to 2.20) on the 100-unit scale. The standard deviation for reproducibility was 2.11 GPS units (95% CI, 1.83 to 2.50) on the 100-point scale.

Clinical Validity
Three studies reporting clinical validity were included as outlined in Table 5. One publication by Klein et al (2014) compiled results for 3 cohorts: 2 in test development including a contemporary (1997-2011) group of patients in a prostatectomy study (N=441; Cleveland Clinic database, 1987-2004) and a biopsy study (N=167; Cleveland Clinic database, 1998-2007); and 1 independent clinical validation study cohort (N=395; UCSF Database, 1998-2011). A second study, Cullen et al (2015), evaluated men with National Comprehensive Cancer Network (NCCN) clinically very low to intermediate risk undergoing prostatectomy. The third study, Whalen et al (2016), evaluated men in a clinical practice setting.

<table>
<thead>
<tr>
<th>Study</th>
<th>Design</th>
<th>Outcome</th>
<th>Dates</th>
<th>Sites</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Klein et al</td>
<td>Case cohort from prospective registry (Simon</td>
<td>Adverse pathology at</td>
<td>1998-2011</td>
<td>UCSF</td>
<td>395</td>
</tr>
<tr>
<td>(2014)</td>
<td>cancer category C)⁹</td>
<td>RP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cullen et al</td>
<td>Retrospective cohort from prospective</td>
<td>Adverse pathology at</td>
<td>1990-2011</td>
<td>U.S. military</td>
<td>382</td>
</tr>
<tr>
<td>(2015)</td>
<td>longitudinal study (Simon category C)⁹</td>
<td>RP, BCR</td>
<td></td>
<td>centers</td>
<td></td>
</tr>
<tr>
<td>Whalen et al</td>
<td>Prospective observational cohort (Simon</td>
<td>Adverse pathology at</td>
<td>2013-2014</td>
<td>Mount Sinai</td>
<td>50</td>
</tr>
<tr>
<td>(2016)</td>
<td>category C)⁹</td>
<td>RP</td>
<td></td>
<td>Hospital</td>
<td></td>
</tr>
</tbody>
</table>

BCR: biochemical recurrence; RP: radical prostatectomy; UCSF: University of California, San Francisco.

Results from the Klein et al (2014) clinical validation study and prostatectomy study provided information on the potential clinical validity of this test. The cohorts included men with a mix of low to low-intermediate
clinical risk characteristics using NCCN or AUA criteria. Patients included in the validation and prostatectomy studies would be considered (a) eligible for active surveillance based on clinical and pathologic findings and (b) representative of patients in contemporary clinical practice. However, all patients elected RP within 6 months of their initial diagnostic biopsies.

The Klein et al (2014) clinical validation study (see Table 5) was designed to evaluate the ability of Oncotype DX Prostate to predict tumor pathology in needle-biopsy specimens. It was prospectively designed, used masked review of prostatectomy pathology results, and as such met the REMARK (Reporting Recommendations for Tumor Marker Prognostic Studies) guidelines for biomarker validation. In the prostatectomy study, all patients with clinical recurrence (local recurrence or distant metastasis) were selected, together with a random sample of those who did not recur, using a stratified cohort sampling method (case-cohort design) to construct a 1:3 ratio of recurrent to nonrecurrent patients. The prespecified primary end point of the validation study was the ability of the GPS to predict the likelihood of favorable pathology in the needle-biopsy specimen. Favorable pathology was defined as freedom from high-grade or non-organ-confined disease. In the prostatectomy study, the ability of the GPS to further stratify patients within AUA groupings was related to clinical recurrence-free interval in regression-to-the-mean estimated survival curves. The Klein et al (2014) validation study results showed that the GPS could refine stratification of patients within specific NCCN criteria groupings, as summarized in Table 6. Proportions were estimated from a plot of GPS versus the percent likelihood of favorable pathology. These findings suggest that a lower GPS could reclassify the likelihood of favorable pathology (ie, less biologically aggressive disease) upward (ie, a potentially lower risk of progression), and vice versa within each clinical stratum. For example, among patients in the cohort classified by NCCN criteria as low risk, the mean likelihood of favorable pathology in a tumor biopsy was about 76%, with 24% then having unfavorable pathology. With the GPS, the estimated likelihood of favorable tumor pathology was broadened, ranging from 55% to 86%, conversely reflecting a 45% to 14% likelihood of adverse pathology, respectively.

Table 6. Reclassification of Prostate Cancer Risk Categories With Oncotype DX Prostate From the Klein (2014) Validation Study

<table>
<thead>
<tr>
<th>NCCN Risk Level</th>
<th>Estimated Mean Likelihood of Favorable Tumor Pathology</th>
<th>Estimated Corresponding GPS, Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCCN Criteria</td>
<td>GPS + NCCN, Range</td>
<td></td>
</tr>
<tr>
<td>Very low</td>
<td>&lt;84%</td>
<td>63%-91%</td>
</tr>
<tr>
<td>Low</td>
<td>&lt;76%</td>
<td>55%-86%</td>
</tr>
<tr>
<td>Intermediate</td>
<td>&lt;56%</td>
<td>29%-75%</td>
</tr>
</tbody>
</table>

GPS: Genomic Prostate Score; NCCN: National Comprehensive Cancer Network.

In effect, the risk of adverse tumor pathology indicated by the GPS could be nearly halved (24%-14%) at 1 extreme, or nearly doubled (24%-45%) at the other, but the actual number of patients correctly or incorrectly reclassified between all 3 categories cannot be ascertained from the data provided. The results suggest that the combination of GPS plus clinical criteria can reclassify patients on an individual basis within established clinical risk categories. However, whether these findings support a conclusion that the GPS could predict the biological aggressiveness of a tumor—hence its propensity to progress—or disease-specific survival based solely on the level of pathology in a biopsy specimen is unclear. Moreover, extrapolation of this
Gene Expression Profiling and Protein Biomarkers for Prostate Cancer Management

Policy #       00403
Original Effective Date: 02/19/2014
Current Effective Date: 06/21/2017

evidence to a true active surveillance population, for which the majority in the study would be otherwise eligible, is difficult because all patients had elective RP within 6 months of diagnostic biopsy.

The Klein et al (2014) prostatectomy study, although used to identify genes to include in the GPS, provided estimates of clinical recurrence rates stratified by AUA criteria compared with rates after further stratification according to the GPS from the validation study. The survival curves for clinical recurrence reached a duration of nearly 18 years based on the dates individuals in the cohort were entered into the database (1987-2004). The reclassifications are summarized in Table 7. The GPS groups are grouped by tertiles defined in the overall study. Absolute rates and precision estimates of clinical recurrence by GPS low-, intermediate-, and high-risk groups were not reported. In the NCCN intermediate group, eg, the 10-year recurrence rate among RP patients was 9.6%. When the GPS was used in the analysis, the 10-year recurrence rate fell to as low as 2.8% (71% reduction) among patients in the low GPS group and to 5.1% (47% reduction) in the intermediate GPS group, but rose to 14.3% (49% increase) in the high GPS group. These data suggest the GPS can reclassify patient risk of recurrence based on a specimen obtained at biopsy. However, the findings do not necessarily reflect a clinical scenario of predicting disease progression in untreated patients under active surveillance.

Table 7. Reclassification of Prostate Cancer 10-Year Clinical Recurrence Risk With Oncotype DX Prostate From the Klein et al (2014) Prostatectomy Study

<table>
<thead>
<tr>
<th>Overall 10-Year Risk (AUA Risk Level)</th>
<th>10-Year Risk (GPS Low Group)</th>
<th>10-Year Risk (GPS Intermediate Group)</th>
<th>10-Year Risk (GPS High Group)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.4% (low)</td>
<td>2.0%</td>
<td>3.4%</td>
<td>7.0%</td>
</tr>
<tr>
<td>9.6% (intermediate)</td>
<td>2.8%</td>
<td>5.1%</td>
<td>14.3%</td>
</tr>
<tr>
<td>18.2% (high)</td>
<td>6.2%</td>
<td>9.2%</td>
<td>28.6%</td>
</tr>
</tbody>
</table>

AUA: American Urological Association; GPS: Genomic Prostate Score.

A retrospective cohort study by Cullen et al (2015) included men with NCCN-defined very low through intermediate-risk prostate cancer undergoing RP within 6 months of diagnosis. The sample was obtained from men enrolled in the Center for Prostate Disease Research longitudinal study at 2 U.S. military medical centers. A Gleason score of 4 or 5 with non-organ-confined disease was considered adverse pathology. Biopsies were available for 500 (57.9%) of 864 eligible patients; 382 (44.2% of eligible) with both adequate tissue for gene expression analysis and available RP pathology were included in the analysis. Selected patients were older (61.0 years vs 59.7 years, p=0.013) and had both higher Gleason scores (p<0.001) and NCCN risk classification (29.8% vs 32.9% intermediate, p=0.035). Median follow-up was 5.2 years and biochemical recurrence (BCR) occurred in 62 (15.4%). Adverse pathology was noted in 163 (34%) men. In an analysis adjusted for baseline characteristics, the GPS was associated with BCR-free survival (HR=2.73 for each 20-point increase; 95% CI, 1.84 to 3.96) (see Table 8). Similarly, the GPS was associated with adverse pathology following RP (HR=3.23 per 20-point increase; 95% CI, 2.14 to 4.97). The GPS improved the C statistic for adverse pathology over NCCN risk alone from 0.63 to 0.72 (confidence intervals not reported). Comparisons with other predictors such as CAPRA or Gleason score alone were not reported. Study implications are also limited by the low proportion of eligible men in the analysis and differences between excluded and included men.
Gene Expression Profiling and Protein Biomarkers for Prostate Cancer Management

Whalen et al (2016) prospectively evaluated the correlation of GPS with final pathology at RP in a clinical practice setting. Eligible men were 50 years of age and older with more than 10 years of life expectancy, PSA levels of 20 ng/mL or less, stage cT1c-cT2c newly diagnosed, untreated prostate cancer, and who met NCCN classifications as very low risk, low risk, or low-intermediate risk. Men were enrolled from May 2013 to August 2014 at an academic medical center. After initial review at the institution, Genomic Health further reviewed biopsy samples to assign Gleason score and tumor length. Samples with Gleason grade discrepancy between initial and central review were excluded from analyses. Clinicians were blinded to GPS when counseling patients on management with active surveillance versus definitive treatment. Genomic Health reclassified patients’ cancers as “less favorable,” “consistent with,” or “more favorable” than what would have been predicted by their NCCN risk group. The primary outcome was adverse pathology at RP defined as any pT3 stage and primary Gleason grade of 4 or any pattern 5. Fifty patients had RP pathology and the reclassification results for these participants are discussed here; 21 (42%) met the definition of adverse pathology. The NCCN risk classification categorized 2 (4%) patients as very low risk, 34 (68%) as low risk, and 14 (28%) as low-intermediate risk. Twenty-three (46%) of patients were reclassified using GPS and the percentage with adverse pathology for the reclassification is shown in Table 9 as derived from data provided in the text. Confidence intervals were not provided.

Table 8. Univariate and Multivariate Association Between GPS and Outcomes in the Klein (2014) Validation Study, Cullen (2015), and Whalen (2016)

<table>
<thead>
<tr>
<th>Study</th>
<th>Outcome</th>
<th>N</th>
<th>Unadjusted</th>
<th>Multivariate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Ratio (95% CI)</td>
<td>Ratio (95% CI)</td>
</tr>
<tr>
<td>Klein (2014) validation study</td>
<td>Adverse pathology</td>
<td>395</td>
<td>OR=2.1 (1.4 to 3.2)</td>
<td>1.9 (1.3 to 2.8)</td>
</tr>
<tr>
<td>Cullen (2015)</td>
<td>BCR</td>
<td>392</td>
<td>HR=2.9 (2.0 to 4.2)</td>
<td>2.7 (1.8 to 3.8)</td>
</tr>
<tr>
<td>Whalen (2016)</td>
<td>Adverse pathology</td>
<td>50</td>
<td>NR</td>
<td>OR=1.4 (NR)</td>
</tr>
</tbody>
</table>

BCR: biochemical recurrence; CI: confidence interval; GPS: Genomic Prostate Score; HR: hazard ratio; NCCN: National Comprehensive Cancer Network; NR: not reported; OR: odds ratio.

a Per 20-point increase in GPS; adjusted for NCCN risk group.
b Per 20-point increase in GPS; adjusted for NCCN risk group and medical center.
c As a continuous variable, adjusted for age, prostate-specific antigen, clinical Gleason score, NCCN risk category.

Table 9. Risk of Adverse Pathology With Oncotype DX Prostate From Whalen et al (2016)

<table>
<thead>
<tr>
<th>Overall AP Risk, % (NCCN Risk Level)</th>
<th>AP Risk, n (%) (GPS Less Favorable Group; n=5)</th>
<th>AP Risk, n (%) (GPS Consistent With Group; n=29)</th>
<th>AP Risk, n (%) (GPS More Favorable Group; n=18)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 (very-low; n=2)</td>
<td>--</td>
<td>0</td>
<td>--</td>
</tr>
<tr>
<td>32% (low; n=34)</td>
<td>5 (100%)</td>
<td>6 (21%)</td>
<td>0</td>
</tr>
<tr>
<td>71% (low-intermediate; n=14)</td>
<td>--</td>
<td>10 (34%)</td>
<td>0</td>
</tr>
</tbody>
</table>

AP: adverse pathology; NCCN: National Comprehensive Cancer Network; GPS: Genomic Prostate Score.

Systematic Reviews

In 2016, Brand et al combined the Klein et al (2014) and Cullen et al (2015) studies using a patient-specific meta-analysis. The GPS was compared to the CAPRA score, NCCN risk group, and AUA/EAU risk group. The authors tested whether the GPS added predictive value for the likelihood of favorable pathology above the clinical risk assessment tools. The model including the GPS and CAPRA score provided the best risk
discrimination; the AUC improved from 0.68 to 0.73 by adding the GPS to the CAPRA score. The AUC improved from 0.64 to 0.70 by adding the GPS to the NCCN risk group. The improvements were reported to be significant but the confidence intervals for AUC were not provided.

Clinical Utility
Klein et al reported a decision-curve analysis that they have proposed reflects the clinical utility of Oncotype DX Prostate. In this analysis, they compared the predictive impact of the GPS in combination with the CAPRA validated tool to the CAPRA score alone on the net benefit for the outcomes of patients with high-grade disease (Gleason score >4+3), high-stage disease, and combined high-grade and high-stage disease. They reported that over a range of threshold probabilities for implementing treatment, “…incorporation of the GPS would be expected to lead to fewer treatments of patients who have favorable pathology at prostatectomy without increasing the number of patients with adverse pathology left untreated.” For example, at a threshold risk of 40% (eg, a man weighing the harms of prostatectomy vs the benefit of active surveillance at 4:6), the test could identify 2 per 100 men with high-grade or high-stage disease at a fixed false-positive rate, compared with using the CAPRA score alone. Thus, an individual patient could use the findings to assess his balance of benefits and harms (net benefit) when weighing the choice to proceed immediately to curative RP with its attendant adverse sequelae, or deciding to enter an active surveillance program. The latter would have an immediate benefit realized by forgoing RP, but might be associated with greater downstream risks of disease progression and subsequent therapies. However, no confidence intervals were presented for the decision-curve analysis.

Finally, Badani et al (2015) prospectively evaluated the decision impact of obtaining a GPS in men with NCCN-defined very low to intermediate-risk cancers. Following test results, recommendations for active surveillance increased from 41% to 51%. Actual treatments received and accuracy of predicted outcomes were not assessed, thereby limiting implications of the study. The study was supported by Genomic Health and all authors reported financial or other relationships with the funder.

Section Summary: Oncotype DX Prostate
The study by Knezevic et al has provided sufficient evidence to establish the analytic validity of Oncotype DX Prostate.

The evidence from 3 studies on clinical validity for Oncotype DX Prostate has suggested the GPS can reclassify a patient’s risk of recurrence based on a biopsy specimen. However, whether these findings support a conclusion that the GPS could predict the biological aggressiveness of a tumor or disease-specific survival, based solely on the level of pathology in a biopsy specimen is unclear. Moreover, generalizing this evidence to a true active surveillance population, for which most in the study would be otherwise eligible, is difficult because all patients had elective RP. Thus the findings do not reflect a clinical scenario of predicting risk of 10-year distant recurrence in untreated patients under active surveillance. The publications also lack precision estimates for important variables such as risk estimates for recurrence or AUC estimates. All validation studies are Simon category C.

No direct evidence of clinical utility was found. The indirect chain of evidence is also incomplete. Klein’s decision-curve analyses has suggested the potential for the combined GPS and CAPRA data to help
patients make decisions based on relative risks associated with immediate treatment or deferred treatment (i.e., active surveillance). This would reflect the clinical utility of the test. However, it is difficult to ascribe possible clinical utility of Oncotype DX Prostate in active surveillance because all patients regardless of clinical criteria elected RP within 6 months of diagnostic biopsy. Moreover, the validity of using tumor pathology as a surrogate for risk of progression and cancer-specific death is unclear. Reports from validation studies lack precision estimates for important variables such as risk estimates for recurrence or AUC estimates. All validation studies were Simon category C. Simon category C studies are not sufficient to determine with confidence the prognosis of low-risk individuals. Ten-year outcomes from a Simon category A or multiple, independent Simon category B studies are needed.

ProMark Protein Biomarker Test
The ProMark assay includes 8 biomarkers that predict prostate pathology aggressiveness and lethal outcomes: DERL1, PDSS2, pS6, YBX1, HSPA9, FUS, SMAD4, and CUL2. The assay results are combined using predefined coefficients for each marker from a logistic regression model to calculate a risk score. The risk score is continuous number between 0 and 1, which estimates the probability of "non-GS 6" pathology.

Analytic Validity
Shipitsin et al reported on the analytic validity of the automated quantitative multiplex immunofluorescence in situ imaging approach assessing: the ability of the test to quantitate markers in a defined region of interest (tumor vs surrounding benign), tissue quality control, assay staining format and reproducibility. To evaluate tissue sample quality, they assessed the staining intensities of several protein markers in benign tissue and using these, categorized prostate cancer tissue blocks into 4 quality groups, of which the best 2 groups were used to generate tumor microarray blocks; 508 prostatectomy specimens were used and of these, 418 passed quality testing and were used for the tumor microarray blocks. For intraexperiment reproducibility, 2 consecutive sections from a prostate tumor test microarray block were stained in the same experiment and scatter plots compared the mean values of the staining intensities; signals from consecutive sections showed $R^2$ correlation values above 0.9 and differences in absolute values typically less than 10%.

Clinical Validity
Blume-Jensen et al (2015) reported on a study of 381 biopsies matched to prostatectomy specimens used to develop an 8-biomarker proteomic assay to predict prostate final pathology on prostatectomy specimen using risk scores.

Biomarker risk scores were defined as favorable if less than or equal to 0.33 and unfavorable if greater than 0.80 with a possible range between 0 and 1 based on false-negative and false-positive rates of 10% and 5%, respectively. The risk score generated for each patient was compared with 2 current risk stratification systems, NCCN guideline categories and the D'Amico system. Results from the study showed that, at a risk score of less than or equal to 0.33, the predictive value of the assay for favorable pathology in very low- and low-risk NCCN and low-risk D'Amico groups were 95%, 81.5%, and 87.2%, respectively, while the NCCN and D'Amico risk classification groups alone had predictive values of 80.3%, 63.8%, and 70.6%, respectively. The positive predictive value for identifying favorable disease with a risk score of less than or equal to 0.33 was 83.6% (specificity, 90%). At a risk score of greater than 0.80, 77% had unfavorable disease. Overall, 39% of the patients in the study had risk scores less than or equal to 0.33 or
Gene Expression Profiling and Protein Biomarkers for Prostate Cancer Management

Policy # 00403
Original Effective Date: 02/19/2014
Current Effective Date: 06/21/2017

greater than 0.8, 81% or which were correctly identified with the 8-biomarker assay. Of the patients with intermediate risk scores (>0.33 to ≤0.8), 58.3% had favorable disease.

The performance of the assay was evaluated on a second blinded study of 276 cases to validate the assay's ability to distinguish “favorable” pathology (defined as Gleason score on prostatectomy less than or equal to 3+4 and organ-confined disease) versus “nonfavorable” pathology (defined as Gleason score on prostatectomy greater than or equal to 4+3 or non-organ-defined disease). The second validation study separated favorable from nonfavorable pathology (AUC=0.68; 95% CI, 0.61 to 0.74).

Table 10. Study Reporting Clinical Validity of ProMark

<table>
<thead>
<tr>
<th>Study</th>
<th>Design</th>
<th>Outcome</th>
<th>Site</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blume-Jensen et al</td>
<td>Retrospective cohort</td>
<td>Favorable pathology at RP</td>
<td>Montreal, QC</td>
<td>276</td>
</tr>
</tbody>
</table>

Clinical Utility
No published studies on the clinical utility of the ProMark test were identified.

Section Summary: ProMark Protein Biomarker Test
Data are insufficient to establish the analytic and clinical validity and clinical utility of the ProMark test.

MANAGEMENT DECISION AFTER RP
Clinical Context and Test Purpose
The purpose of gene expression profiling and protein biomarkers tests in patients who have prostate cancer and who have undergone RP is to inform a management decision.

For example, the optimal timing of RT after RP is a debate. Adjuvant RT may maximize cancer control outcomes; however, salvage RT can minimize overtreatment and still lead to acceptable oncologic outcomes. Several analyses have shown conflicting conclusions whether adjuvant RT is favored over salvage RT (with salvage RT typically being initiated at a post-RP PSA level of 0.3 to 0.6 ng/mL).

The second question addressed in this evidence review is: Does gene expression profiling or tests of protein biomarkers, compared with clinicopathologic risk stratification or when used with clinicopathologic risk stratification, improve outcomes in men following RP?

The following PICOTS were used to select literature to inform this review.

Patients
The relevant population of interest is individuals who have undergone RP treatment for prostate cancer, and who are deciding on subsequent management such as adjuvant RT versus no adjuvant RT.

Intervention
Prolaris, described in the previous section, is also intended to classify low-to-intermediate risk individuals who have undergone RP.
Decipher is a tissue-based tumor 22-biomarker gene expression profiling test intended to classify high-risk individuals who have undergone RP.

Comparator
Clinicopathologic risk stratification is currently being used to make decisions about prostate cancer management following RP. Clinical characteristics (eg, stage, biopsy Gleason grade, serum PSA, surgical margin, disease involvement) and demographic characteristics (eg, age, life expectancy) are combined to classify men according to risk. As described previously, NCCN and AUA provide risk-stratification guidelines. The Stephenson nomogram and Cancer of the Prostate Risk Assessment‒Surgical (CAPRA-S) nomogram can be used to predict outcomes after RP.

Outcomes
Beneficial outcomes resulting from a true test result are prolonged survival, improved quality of life and reduction in unnecessary treatment-related adverse effects. Harmful outcomes resulting from a false test result are recurrence, metastases or death, and unnecessary treatments. The outcomes of interest are listed in Table 11.

<table>
<thead>
<tr>
<th>Table 11. Outcomes of Interest for Individuals After Radical Prostatectomy</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Outcome</strong></td>
</tr>
<tr>
<td>Overall survival</td>
</tr>
<tr>
<td>Disease-specific survival</td>
</tr>
<tr>
<td>Treatment-related morbidity</td>
</tr>
</tbody>
</table>

NCI: National Cancer Institute.
Gene Expression Profiling and Protein Biomarkers for Prostate Cancer Management

Policy # 00403
Original Effective Date: 02/19/2014
Current Effective Date: 06/21/2017

Time
Ten-year outcomes are of interest due to the prolonged natural history of prostate cancer and low number of events observed.

Setting
Decisions about management of prostate cancer following RP are generally made by patients and urologists in the secondary or tertiary care setting.

Prolaris
Prolaris used for initial management decisions was described in the previous section. This section will review Prolaris for management after RP.

Analytic Validity
The analytic validity of Prolaris was described in the previous section.

Clinical Validity
Four studies reporting clinical validity in the post-RP management setting were included as outlined in Table 12. Three of these studies—Cuzick et al (2011), Cooperberg et al (2013) and Bishoff et al (2014)—reported on post-RP patients. Koch et al (2016) reported on post-RP patients with BCR. Freedland et al (2013) reported on post-RT patients but is included in this section for completeness.

Table 12. Studies Reporting Clinical Validity of Prolaris for post-RP or post-RT management

<table>
<thead>
<tr>
<th>Study</th>
<th>Design*</th>
<th>Outcome</th>
<th>Dates</th>
<th>Sites</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Postprostatectomy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cuzick et al (2011)</td>
<td>Retrospective cohort from prospective registry (Simon category C)</td>
<td>BCR (median follow-up, 9.4 y)</td>
<td>1985-1995</td>
<td>Scott and White Clinic</td>
<td>366</td>
</tr>
<tr>
<td>Cooperberg et al (2013)</td>
<td>Retrospective cohort from prospective registry (Simon category C)</td>
<td>BCR (median follow-up, 7 y)</td>
<td>1994-2011</td>
<td>UCSF Registry</td>
<td>413</td>
</tr>
<tr>
<td>Bishoff et al (2014)</td>
<td>Retrospective cohort from medical records (Simon category D)</td>
<td>BCR (median follow-up, 5 y, 7 y, NR for 3 cohorts)</td>
<td>2005-2006</td>
<td>Martini Clinic</td>
<td>283</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Durham VAMC</td>
<td>176</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Intermountain Healthcare</td>
<td>123</td>
</tr>
<tr>
<td>Koch et al (2016)</td>
<td>Retrospective cohort from medical records (Simon category D)</td>
<td>Systemic disease (median follow-up, 9.4 y)</td>
<td>1995-2010</td>
<td>Indiana University SOM</td>
<td>47</td>
</tr>
<tr>
<td>After external-beam radiotherapy</td>
<td>Retrospective cohort, source unclear (Simon category D)</td>
<td>BCR</td>
<td>1991-2006</td>
<td>Durham VAMC</td>
<td>141</td>
</tr>
</tbody>
</table>

BCR: biochemical recurrence; NR: not reported; PC: prostate cancer; SOM: School of Medicine; UCSF: University of California, San Francisco; VAMC: Veterans Affairs Medical Center.

Cuzick et al (2011) examined the potential use of the Prolaris CCP test combined with a clinical score following RP, using a retrospective cohort of archived samples from a tumor registry. The study also
included a cohort of men with localized prostate cancer detected from specimens obtained during transurethral resection of the prostate, which is not a population of interest here, and so has not been described. Men conservatively managed after RP between 1985 and 1995 were identified from a tumor registry (n=366 with CCP scores, Scott and White Clinic, in Texas). The primary end point was time to BCR and the secondary end point was prostate cancer death. Myriad Genetics assessed CCP scores blindly. The median age of patients was 68 years and the median follow-up 9.4 years. Gleason scores were 7 or lower in 96%, but margins were positive in 68%. Cancers were clinically staged as T3 in 34%; following RP, 64% was judged pathologic stage T3. CCP score was associated with BCR (adjusted HR=1.77; 95% CI, 1.40 to 2.22) (see Table 13). Analyses of prostate cancer deaths in the RP cohort were problematic, owing to only 12 (3%) deaths. The clinical score included PSA, stage, positive surgical margins, and Gleason score. The model was optimized using stepwise variable selection (eg, a development model). The AUC for BCR within 5 years in the RP cohort was 0.825 for the clinical score and 0.842 for the combined clinical/CCP score. The discriminatory ability of the clinical score is noteworthy. Although the CCP increased the AUC by 2%, whether that improvement is clinically useful is unclear because we lack reclassification data and analysis of net benefit.

Cooperberg et al (2013) sought to evaluate the CCP score in a RP cohort and the incremental improvement over the Cancer of the Prostate Risk Assessment Postsurgical (CAPRA-S) score for predicting BCR using a prospective-retrospective design (conforming to a PRoBE study design). A prognostic model was developed from the RP cohort described by Cuzick et al (2011). The validation cohort was obtained from patients identified from the University of California, San Francisco (UCSF) Urologic Oncology Database. Tissue sufficient to obtain a CCP score was available for 413 men (69% of the 600 eligible samples). Both UCSF and Myriad Genetics performed statistical analyses. In the validation cohort, 95% had Gleason scores of 7 or lower, 16% of samples had positive margins, 4% had seminal vesicle invasion, and 23% had extracapsular extension. BCR occurred in 82 (19.9%) men. The unadjusted hazard ratio for BCR increased by 2.1 (95% CI, 1.6 to 2.9) per unit increase in CCP score (see Table 13). A predictive model for the combined CCP/CAPRA-S score developed in the Cuzick et al (2011) RP cohort applied to the UCSF cohort obtained an AUC for BCR with CAPRA-S alone of 0.73, increasing to 0.77 for the combined CCP/CAPRA-S score.

Bishoff et al (2014) examined the prognostic ability of the CCP score in 3 cohorts: the Martini Clinic (n=283, simulated biopsies from FFPE RP specimen), Durham Veterans Affairs Medical Center (n=176, diagnostic biopsies), and Intermountain Healthcare (n=123, diagnostic biopsies).66 The combined analysis included all 582 patients. Gleason scores were 7 or lower in 93% of men. In the combined cohorts, a unit increase in the CCP score increased the adjusted hazard ratio for BCR by 1.47 (95% CI, 1.23 to 1.76) (see Table 13). Metastatic events (n=12) were too few to draw conclusions. Although the CCP score was associated with increased risk of BCR, the small number of metastatic events does not permit examining whether the CCP score provides improved discrimination over clinicopathologic variables.

Koch et al (2016) evaluated whether the CCP score could discriminate between systemic disease and local recurrence in patients with BCR after RP.67 All patients treated with RP as primary therapy at an academic medical center between 1995 and 2010 for whom samples were available and who had a BCR and either developed metastatic disease or received salvage EBRT with at least 2 years of follow-up were eligible for
Gene Expression Profiling and Protein Biomarkers for Prostate Cancer Management

Policy #       00403
Original Effective Date: 02/19/2014
Current Effective Date: 06/21/2017

retrospective analysis (N=60). Data from 5 patients were excluded for failing to meeting clinical eligibility requirements (no clarification provided) or because data were incomplete; sample blocks from 3 patients contained insufficient tumor for assay and data from 6 patients were excluded due to lack of "passing" CCP scores. Forty-seven patients were included in analysis. The outcome was classified into 3 categories: (1) metastatic disease (n=22), (2) nonresponse to salvage EBRT (n=14), and (3) durable response to salvage EBRT (n=11). Analyses were performed with a binary outcome (categories 1 and 2 combined). For each 1-unit change in the CCP score, the univariate odds ratio (OR) for metastatic disease or nonresponse was 3.72 (95% CI, 1.29 to 10.7) (see Table 13). The CCP score correlated with each of the clinical characteristics. Multivariate analysis was performed; however, due to the very small number of participants in the durable response group, confidence intervals were very wide and estimates are likely unstable.

Table 13. Univariate and Multivariate Association Between Prolaris CCP and Outcomes in Post-RP Clinical Validation Studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Outcome</th>
<th>N</th>
<th>Unadjusted Ratio (95% CI)</th>
<th>Multivariate Ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cuzick et al (2011)</td>
<td>BCR</td>
<td>366</td>
<td>HR=1.89 (1.54 to 2.31)</td>
<td>1.77 (1.40 to 2.22)</td>
</tr>
<tr>
<td></td>
<td>Prostate cancer death</td>
<td>337</td>
<td>HR=2.92 (2.38 to 3.57)</td>
<td>2.56 (1.85 to 3.53)</td>
</tr>
<tr>
<td>Cooperberg et al (2013)</td>
<td>BCR</td>
<td>413</td>
<td>HR=2.1 (1.6 to 2.9)</td>
<td>1.7 (1.3 to 2.4)</td>
</tr>
<tr>
<td>Bishoff et al (2014)</td>
<td>BCR</td>
<td>582</td>
<td>HR=1.60 (1.35 to 1.90)</td>
<td>1.47 (1.23 to 1.76)</td>
</tr>
<tr>
<td>Koch et al (2016)</td>
<td>Metastatic disease</td>
<td>47</td>
<td>OR=3.72 (1.29 to 10.7)</td>
<td>10.4 (2.05 to 90.1)</td>
</tr>
<tr>
<td></td>
<td>or nonresponse</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Per 1-unit increase in CCP. Adjusted for PSA, Gleason score, pathological T stage and grade, positive surgical margins, extracapsular extension, bladder involvement, seminal vesicle involvement, positive lymph node, and age.

** Per 1-unit increase in CCP. Adjusted for Gleason score, PSA, Ki67, and cancer extent.

*** Per 1-unit increase in CCP. Adjusted for Cancer of the Prostate Risk Assessment–Surgical.

† Per 1-unit increase in CCP. Adjusted for PSA, Gleason score, and adjuvant treatment.

‡ Per 1-unit increase in CCP. Adjusted for Gleason score, time from surgery to BCR, and PSA.

Although not a study of management, post-RP, Freedland et al (2013) described the prognostic ability of the CCP score for predicting BCR in men who received primary EBRT.68 The retrospective data included 141 men diagnosed with prostate cancer who had biopsy samples and follow-up of at least 3 years who were treated with EBRT from 1991 to 2006. Nineteen (13%) of men experienced BCR by 5 years. The univariate hazard ratio for BCR for each 1-unit increase in CCP was 2.55 (95% CI, 1.43 to 4.55). The multivariable hazard ratio for BCR associated with 1-unit increase in CCP, including adjustment for pretreatment PSA, Gleason, percent positive cores, and concurrent androgen deprivation therapy, was 2.11 (95% CI, 1.05 to 4.25).

Systematic Reviews
As described in a previous section, results of a systematic review and meta-analysis supported by the manufacturer were reported. Seven published studies were identified; all been reviewed in the previous paragraphs (needle biopsy conservative management cohorts, postprostatectomy cohorts and EBRT cohort). Including 4 validity studies that reported outcomes of BCR in post-RP cohorts, the pooled estimate of the hazard ratio, calculated with random-effects meta-analytic methods, for BCR for a 1-unit increase in CCP score was 1.9 (95% CI, 1.6 to 2.3). Two studies reported outcomes for disease-specific mortality. Only
Gene Expression Profiling and Protein Biomarkers for Prostate Cancer Management

Policy # 00403
Original Effective Date: 02/19/2014
Current Effective Date: 06/21/2017

one of those was a post-RP study so the pooled hazard ratios are not relevant here. There was evidence of heterogeneity in both models; reviewers did not report any variables associated with heterogeneity.

Clinical Utility
We identified no studies to directly support the clinical utility of Prolaris.

In a decision-curve analysis, Cooperberg et al (2013) found the CAPRA-S score superior to CCP alone (as well as treat-none or treat-all strategies) in men after prostatectomy. A combined CCP/CAPRA-S predictor appeared only slightly better than CAPRA-S alone for thresholds of approximately 30% or more. For example, at a threshold of 30% (ie, meaning a man would value the harm-to-benefit of treatment such as RT as 3:7), the combined CCP/CAPRA-S score would detect about 2 more men per 100 likely to experience BCR if the false-positive rate was fixed. However, the lack of confidence intervals for the decision-curve analysis, together with the small difference, is consistent with an uncertain net benefit obtained by adding CCP to the CAPRA-S score.

Section Summary: Prolaris
Analytic validity of gene expression analysis for prostate cancer management using Prolaris was reported by Warf et al (2015) and supported by results from the MAQC project.

Four identified studies examined the clinical validity of Prolaris in men after RP using a BCR or systemic disease end point. Cuzick et al (2011) found the CCP score offered little improvement in the AUC (2%) over clinicopathologic predictors and did not examine reclassification. Cooperberg et al (2013) found the AUC for BCR improved from 0.73 (CAPRA-S alone) to 0.77 by adding CCP score. Bishoff et al (2014) and Koch et al (2016) did not report any classification or discrimination measures. Koch et al (2016) was performed in patients who had a BCR following RP. All validation studies were Simon category C or D.

No direct evidence is available to support the clinical utility of Prolaris for improving net outcomes of patients with localized prostate cancer or following RP. The indirect chain of evidence is also incomplete. Decision-curve analysis did not provide convincing evidence of meaningful improvement in net benefit in BCR by incorporating the CCP score. Prolaris CCP score may have an association with BCR but disease-specific survival outcomes were not reported. A larger number of disease-specific survival events and precision estimates for discrimination measures are needed. All validation studies were Simon category C or D. Simon category C studies are not sufficient to determine with confidence the prognosis of CCP score.

Decipher Prostate Cancer Classifier
The Decipher test classifies patients as low risk, who can delay or defer RT after prostatectomy, or high risk, as those who would potentially benefit from early radiation. The gene expression classifier is a continuous risk score between 0 and 1, with higher risk scores indicating a greater probability of developing metastasis.

Analytic Validity
Published data on the analytic validity of the Decipher prostate cancer classifier consists of 1 study, which was performed on surgical resection specimens from patients with prostate cancer identified to be in a...
postsurgery high-risk population. The Decipher test platform was performed in FFPE tissue to assess the differential expression in the discovery, validation, and clinical application. Matched FFPE and unfixed fresh-frozen specimens from paired tumor and normal samples from kidney, lung, and colon were compared and the microarray signals derived from the degraded RNA extracted from FFPE specimens was found to be highly analogous to the signals from the RNA in the fresh frozen specimens.

According to GenomeDx’s website, additional analytic performance studies were conducted, and the test was subjected to reagent and analytical verification studies in the laboratory according to CLIA guidelines, reproducibility was demonstrated by evaluation of day-to-day and operator-operator precision, and the assay showed concordant results between the clinical laboratory, R&D laboratories, and pathology.

Clinical Validity
The clinical validity of the Decipher test (genomic classifier [GC]) has been reported in 10 studies to predict metastasis, mortality, or BCR after RP in patients with postoperative high-risk features like pathologic stage T2 with positive margins, pathologic stage T3 disease, or a rising PSA (see Tables 14 and 15).

Table 14. Studies Evaluating the Decipher Genomic Classifier

<table>
<thead>
<tr>
<th>Author</th>
<th>Design</th>
<th>Outcome</th>
<th>Sites</th>
<th>Dates</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Observation and RT samples</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Erho (2013) (validate)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>186</td>
</tr>
<tr>
<td>Karnes (2013)</td>
<td>Case cohort from registry (Simon C)</td>
<td>Mets (5 y)</td>
<td>Mayo Clinic</td>
<td>2000-2006</td>
<td>219</td>
</tr>
<tr>
<td>Ross (2014)</td>
<td>Case cohort from registry (Simon C)</td>
<td>Mets (5 y)</td>
<td>Mayo Clinic</td>
<td>2000-2006</td>
<td>85</td>
</tr>
<tr>
<td>Ross (2014)</td>
<td>Case cohort from registry (Simon C)</td>
<td>Mets (5 y)</td>
<td>Mayo Clinic</td>
<td>2000-2006</td>
<td>85</td>
</tr>
<tr>
<td>Cooperberg (2015)</td>
<td>Case cohort from registry (Simon C)</td>
<td>PC mortality</td>
<td>CapSURE Registry</td>
<td>2000-2006</td>
<td>185</td>
</tr>
<tr>
<td></td>
<td><strong>Observation-only samples</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Klein (2015); Klein (2016)</td>
<td>Retrospective cohort from registry (Simon C)</td>
<td>Mets (5 y); Mets (10 y)</td>
<td>Cleveland Clinic</td>
<td>1993-2001</td>
<td>169</td>
</tr>
<tr>
<td>Ross (2016)</td>
<td>Case cohort from registry (Simon C)</td>
<td>Mets (10 y)</td>
<td>Johns Hopkins</td>
<td>1992-2010</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td><strong>RT-only samples</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Den (2014)</td>
<td>Retrospective cohort from registry (Simon C)</td>
<td>BCR</td>
<td>Thomas Jefferson</td>
<td>1999-2009</td>
<td>139</td>
</tr>
</tbody>
</table>

BCR: biochemical recurrence; CapSURE: Cancer of the Prostate Strategic Urologic Research Endeavor; Mets: metastases; PC: prostate cancer; RT: radiotherapy.

* Appears to be subgroup with BCR from Karnes et al (2013).
Table 15. Reported Prognostic Accuracies (Clinical Validity) of Decipher and Comparators

<table>
<thead>
<tr>
<th>Author</th>
<th>Outcome</th>
<th>GC</th>
<th>Comparator</th>
<th>GC + Comparator</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Observation and RT samples</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Erho (2013) (training)</td>
<td>Metastasis</td>
<td>0.90 (0.87 to 0.94)</td>
<td>0.76 (0.67 to 0.83)</td>
<td>0.91 (0.87 to 0.94)</td>
</tr>
<tr>
<td>Erho (2013) (validate)</td>
<td></td>
<td>0.75 (0.70 to 0.81)</td>
<td>0.69 (0.60 to 0.77)</td>
<td>0.74 (0.65 to 0.82)</td>
</tr>
<tr>
<td>Karnes (2013)</td>
<td>Metastasis</td>
<td>0.79 (0.68 to 0.87)</td>
<td>0.64 (0.55 to 0.72)</td>
<td></td>
</tr>
<tr>
<td>Ross (2014)</td>
<td>Metastasis</td>
<td>0.82 (0.76 to 0.86)</td>
<td>0.70 (0.66 to 0.75)</td>
<td>0.75 (0.69 to 0.80)</td>
</tr>
<tr>
<td>Cooperberg (2015)</td>
<td>PC mortality</td>
<td>0.78 (0.68 to 0.87)</td>
<td>0.75 (0.55 to 0.84)</td>
<td></td>
</tr>
<tr>
<td><strong>Observation-only samples</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Klein (2015); Klein (2016)</td>
<td>Metastasis</td>
<td>5 y: 0.77 (0.66 to 0.87)</td>
<td>5 y: 0.75 (0.65 to 0.84)</td>
<td>5 y: 0.79 (0.65 to 0.85)</td>
</tr>
<tr>
<td>Ross (2016)</td>
<td>Metastasis</td>
<td>0.76 (0.65 to 0.84)</td>
<td>0.77 (0.69 to 0.85)</td>
<td>0.87 (0.77 to 0.94)</td>
</tr>
<tr>
<td>Glass (2016)</td>
<td>Clinical recurrence</td>
<td>0.80 (NR)</td>
<td>0.73 (NR)</td>
<td>0.84 (NR)</td>
</tr>
<tr>
<td><strong>RT-only samples</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Den (2014)</td>
<td>BCR post RT</td>
<td>0.75 (0.67 to 0.84)</td>
<td>0.70 (0.61 to 0.79)</td>
<td>0.78 (0.69 to 0.86)</td>
</tr>
<tr>
<td>Den (2015)</td>
<td>Metastasis post RT</td>
<td>0.83 (0.72 to 0.89)</td>
<td>0.66 (0.56 to 0.78)</td>
<td>0.85 (0.79 to 0.93)</td>
</tr>
<tr>
<td>Freedland (2016)</td>
<td>Metastasis post RT</td>
<td>0.85 (0.73 to 0.88)</td>
<td>0.65 (0.54 to 0.81)</td>
<td>NR</td>
</tr>
</tbody>
</table>

- Clinical classifier includes Gleason score, extracapsular extension, positive surgical margins, seminal vesicle invasion, or lymph node involvement.
- Cancer of the Prostate Risk Assessment‒Surgical.
- Stephenson nomogram.
- Only reported compared with single clinical predictors.
- AUC CI obtained by digitizing figure.
- Gleason score.
- Briganti score.
- National Comprehensive Cancer Network risk categories.

All studies were conducted from registry data. The development study had a nested case-control design. The 5- and 10-year results of 1 study were published separately. Four were case-cohort studies, and 5 used retrospective cohorts. Owing to apparent overlap in samples, the number of unique patients in the studies is difficult to ascertain. Six studies were supported by GenomeDx, which offers the Decipher test; all studies identified multiple authors as company employees. Studies were considered according to whether post RP included men were observed or treated with RT (adjuvant or salvage), resulting in the following groupings: (1) observation or RT, (2) observation only, and (3) RT only.

Five studies, including the test (validation) sample from the development study, examined men observed following RP and undergoing adjuvant or salvage RT. Median follow-up periods ranged from 6.4 to 16.9 years. The distributions of Gleason scores in the studies varied—from 17.8% to 49.3% with Gleason scores of 8 or higher and 0.4% to 15.1% with 6 or lower. Extracapsular extension of the tumor ranged from 42.7% and 72.3% of men across the studies.

**Validation Studies: Observation and RT Samples**

Cooperberg et al (2015) evaluated the prognostic accuracy of the test for prostate cancer mortality; the others, for the development of metastasis (see Table 15). Karnes et al (2013) reported a 2.4% five-year
cumulative incidence of metastasis in 338 men with GC scores of less than 0.4, but 22.5% in the 77 men with scores 0.6 or more. In men who had developed BCR, Ross et al (2014) found the GC score associated with 5-year cumulative incidence of metastases—10.0% in men with scores of 0.4 or lower versus 54.0% in those with higher scores. The GC AUCs for predicting metastases ranged from 0.75 to 0.82. Two studies found a clinical classifier achieved AUCs of 0.69 and 0.70, respectively. Karnes et al (2013) examined only single clinicopathologic predictors with an AUC for the Gleason score of 0.64. Erho et al (2013) reported favorable reclassification compared with the Gleason score, but applied cutoffs not currently used. Karnes et al found improved reclassification to lower risk—among the 150 men not developing metastases, the GC reclassified 23 men with Gleason scores of 8 or more into the lowest risk group and 11 men with Gleason scores of 7 or less to higher risk. However, reclassification of men experiencing metastases from lower to higher risk is likely the most important role for the test, given that RT, although effective, is generally avoided. Yet, among the 69 men developing metastases in Karnes et al, of the 29 with Gleason scores of 7 or lower, 10 were correctly reclassified to the highest GC risk (score >0.6), but of the 40 men with Gleason scores of 8 or higher, 10 were incorrectly reclassified to the lowest GC risk group (score <0.4). For prostate cancer mortality, compared with CAPRA-S, Cooperberg et al found that the GC improved reclassification somewhat—of the 19 men with CAPRA-S scores of 5 or lower, 12 were correctly reclassified to the highest GC risk and 1 was incorrectly reclassified with a CAPRA-S greater than 6 to low risk; all men had CAPRA-S scores of 3 or more.

Validation Studies: Observation-Only Samples
Three validation studies reported in 4 publications, Klein et al (2015), Klein et al (2016), Ross et al (2016), and Glass et al (2016) excluded men receiving any adjuvant therapy following RP over median follow-up periods ranging from 7.8 to 9 years. The Klein and the Ross studies assessed the prognostic accuracy for metastasis through 5 years or 10 years. Glass et al (2016) assessed the prognostic accuracy for clinical recurrence, defined as local, regional, or distant recurrence or metastasis confirmed by clinical or radiologic evidence. Ross et al (2016) reported a 6.5% 5-year cumulative incidence of metastases in men with GC scores of 0.45 or lower, compared with 30.3% in those with scores higher than 0.60. The AUCs for development of metastases were 0.77 and 0.76 for the GC in the 2 studies, and essentially the same as the best comparator (see Table 15). Glass et al (2016) reported a 2.6% clinical recurrence rate by 10 years for patients with GC scores less than 0.45 and 13.5% recurrence rate for those with score 0.45 or greater. The AUCs for clinical recurrence at 10 years were 0.80 for GC and 0.73 for CAPRA-S. In the samples examined by Ross et al and Glass et al, combining the GC with the best clinicopathologic tool improved the AUC. Although neither study included a "standard" reclassification table, Ross et al reported 10-year cumulative incidence of metastases stratified by GC and CAPRA-S. The GC appeared to discriminate within CAPRA-S categories, but appeared to add little to a score greater than 5. Of note, the reclassification employed a different lower cutoff score (0.45 instead of 0.40), “…based on a recently refined optimization algorithm.”

Validation Studies: RT-Only Samples
Three analyses of overlapping retrospectively assembled cohorts of men undergoing either adjuvant or salvage RT. One study examined the prognostic ability of the GC for BCR, while the other 2 examined its prognostic ability for metastases. The median follow-up in Den et al (2014) and Den et al (2015) exceeded 10 years; the median follow-up in Freedland et al (2016) was 7.4 years. Just over three-quarters of the men in the studies had positive surgical margins or a larger proportion than in the other validation studies. Den et
Gene Expression Profiling and Protein Biomarkers for Prostate Cancer Management

Policy # 00403
Original Effective Date: 02/19/2014
Current Effective Date: 06/21/2017

al (2014) found that the GC’s AUC for BCR was 0.75 compared with 0.70 for the Stephenson nomogram. In Den et al (2015), the AUC for metastases was 0.83 versus 0.66 for CAPRA-S; 7 (21.2%) of men with high GC scores (>0.6) developed metastases compared with 12 (15.2%) men with CAPRA-S scores exceeding 5.62. However, overall only 19 (10.1%) men had developed metastases. Among the 160 men who did not, the GC reclassified 27 of 67 men with high CAPRA-S scores into a low-risk group, but given the small number of men developing metastases, the reclassifications were somewhat uncertain. Finally, the authors explored whether the classifier might identify men likely to benefit from adjuvant RT over salvage, suggesting that adjuvant therapy might be preferred in men with a GC score greater than 0.4. However, that result was based on only 14 men with GC scores of 0.4 and 3 men with values that were lower. In Freedland et al (2016), the C index for metastases was 0.85 for the GC compared to 0.63 for the CAPRA-S and 0.65 for the Briganti nomogram. Twenty men developed metastases. In a reclassification analysis, 31 (39%) patients in the upper 2 tertiles of risk by Briganti were classified as low risk by the GC and 1 of them developed metastases during follow-up. Seventy-three (49%) patients who were categorized as intermediate or high risk with CAPRA-S were classified as GC low-risk; 3 developed metastases during follow-up.

Clinical Utility

Direct Evidence
No studies reporting direct evidence were identified.

Indirect Evidence: Decision Curves
Seven studies included decision curves comparing net benefit of different strategies using metastases or survival as the outcome (see Table 16). In observation and RT samples from Karnes et al (2013) and Ross et al (2014), over a 15% to 25% range of thresholds for decision making (ie, suspected probability of developing metastases), relying on the GC result for adjuvant RT decisions, compared with treating based on the best comparator test, would be expected to correctly identify as few as no men or as many as 4 per 100 likely to experience metastases, assuming no increase in false positives. No confidence intervals were provided for the net benefit estimates and uncertainty cannot be evaluated. In the 2 observation-only samples, although the GC improved the net benefit over a “treat none” strategy over 15% to 25% thresholds, it appeared to offer little over the comparator test (eg, about 1 additional patient likely to experience metastases without an increase in false positives). In Ross et al, the net benefit for CAPRA-S score exceeded that of the GC, with the net benefit of the GC plus CAPRA-S score slightly better than the CAPRA-S score alone. Finally, among men undergoing RT, decision curves suggested that the test would identify 3 or 4 men developing metastases per 100 tested at a fixed false-positive rate. Lobo et al (2015) reported an individualized decision analysis comparing the GC to “usual care” using data from the cohorts in Karnes et al (2013) and Den et al (2014). The usual care probabilities of receiving each treatment were derived from the published literature. A 6% threshold for the GC score was used for GC-based treatment. Using the cohort from Karnes et al (2013), the estimated 10-year probability of metastasis or death was 0.32 (95% CI, 0.32 to 0.33) for usual care compared to 0.31 (95% CI, 0.30 to 0.32) for GC-based treatment. In the cohort from Den et al (2014), the estimated 10-year probability of metastasis or death was 0.28 (95% CI, 0.27 to 0.29) for usual care compared to 0.26 (95% CI, 0.25 to 0.27) for GC-based treatment.
Table 16. Reported Net Benefit of the Decipher Classifier vs Comparators

<table>
<thead>
<tr>
<th>Author</th>
<th>Outcome</th>
<th>Range of Net Benefit vs Treat None</th>
<th>Best Comparator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observation and RT samples</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Karnes et al (2013)</td>
<td>Metastasis</td>
<td>0.009 to 0.020</td>
<td>-0.004 to 0.003</td>
</tr>
<tr>
<td>Ross et al (2014)</td>
<td>Metastasis</td>
<td>0.09 to 0.13</td>
<td>0.036 to 0.040</td>
</tr>
<tr>
<td>Cooperberg et al (2015)</td>
<td>PC mortality</td>
<td>0.003&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.003&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Lobo et al (2016) with Karnes et al (2013)</td>
<td>Metastasis or death</td>
<td>NR</td>
<td>0.017</td>
</tr>
<tr>
<td>Observation-only samples</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Klein et al (2015)</td>
<td>Metastasis</td>
<td>0.008 to 0.025</td>
<td>0.000 to 0.012</td>
</tr>
<tr>
<td>Ross et al (2016)</td>
<td>Metastasis</td>
<td>0.09 to 0.12</td>
<td>0.003 to 0.004</td>
</tr>
<tr>
<td>RT-only samples</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Den et al (2015)</td>
<td>Metastasis post RT</td>
<td>0.09 to 0.11</td>
<td>0.03 to 0.04</td>
</tr>
</tbody>
</table>

<sup>a</sup> For 25% threshold.

Indirect Evidence: Changes in Management
Several studies have compared physician’s treatment recommendations before and after receiving GC results. Because the studies did not include information on outcomes and clinical validity has not been established, it is not known whether these treatment decisions represent a clinical improvement in management.

Indirect Evidence: Analysis of Association Between the GC and Treatment Effects
Ross et al (2016) reported results of a retrospective, comparative study of RT after RP for 422 men with pT3 disease or positive margins. The men were from 4 cohorts previously described (Karnes 2013; Den 2014; Ross 2016; Freedland 2016). The 4 treatment groups were adjuvant RT (n=111), minimal residual disease salvage RT (n=70), salvage RT (n=83), and no RT (n=157). The primary end point was metastasis. Thirty-seven men developed metastasis and the median follow-up was 8 years. Both CAPRA-S (HR=1.39; 95% CI, 1.18 to 1.62) and Decipher (HR=1.28; 95% CI, 1.08 to 1.52) were independently associated with metastasis in multivariable analysis. There was no evidence that treatment effect was dependent on genomic risk (interaction p=0.16 for CAPRA-S, p=0.39 for Decipher). Men with low CAPRA-S or low Decipher scores had a low risk of metastatic events regardless of treatment selection and men with high CAPRA-S or Decipher scores benefitted from adjuvant RT compared to the other treatments.

Section Summary: Decipher Prostate Cancer Classifier
The analytic validity of the Decipher test has been reported in 1 published study. Clinical validity has been evaluated in overlapping validation samples (including the development test set). The validation studies consisted of observational data obtained from registries with archived samples. All validation studies are Simon category C. Simon category C studies are not sufficient to determine with confidence the prognosis of high risk individuals. Ten-year disease-specific survival outcomes from a Simon category A or multiple, independent Simon category B studies are needed. Although each study evaluated different outcomes (ie, metastasis, prostate cancer–specific mortality, BCR) in samples with different populations, all studies reported some incremental improvement in discrimination. Confidence intervals of AUC frequently overlapped between Decipher and comparators. Results did not consistently demonstrate meaningful
improvement in reclassification—possibly most importantly to higher risk categories. The performance over clinicopathologic predictors did not appear consistently and meaningfully improved.

**Ongoing and Unpublished Clinical Trials**
A currently unpublished trial (ie, a category B study) that might influence this review is listed in Table 18.

### Table 18. Summary of Key Trials

<table>
<thead>
<tr>
<th>NCT No.</th>
<th>Trial Name</th>
<th>Planned Enrollment</th>
<th>Completion Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCT00002874</td>
<td>Radiation Therapy With or Without Bicalutamide in Treating Patients With Stage II, Stage III, or Recurrent Prostate Cancer</td>
<td>840</td>
<td>Dec 2016</td>
</tr>
<tr>
<td>NCT02454595</td>
<td>Registry to Measure the Impact of Adding Genomic Testing</td>
<td>200</td>
<td>Nov 2016</td>
</tr>
<tr>
<td>NCT026668276</td>
<td>The Impact of a Gene Expression Profile on Treatment Choice and Outcome Among Minority Men Newly Diagnosed With Prostate Cancer: A Randomized Trial</td>
<td>300</td>
<td>Aug 2022</td>
</tr>
<tr>
<td>NCT02609269</td>
<td>Decipher Genomics Resource Information Database (GIRD)</td>
<td>10,000</td>
<td>Dec 2020</td>
</tr>
<tr>
<td>NCT02723734</td>
<td>Validation Study on the Impact of Decipher Testing – VANDAAM Study</td>
<td>50</td>
<td>May 2021</td>
</tr>
<tr>
<td>NCT02080689</td>
<td>Prospective Clinical Utility Study to Assess the Impact of Decipher on Treatment Decisions after Surgery (PRO-IMPACT)</td>
<td>286</td>
<td>Mar 2017</td>
</tr>
<tr>
<td>NCT01350180</td>
<td>Assessing DNA Changes in High Risk Prostate Cancer to Determine Prognosis</td>
<td>132</td>
<td>Dec 2016</td>
</tr>
</tbody>
</table>

NCT: national clinical trial.

### Summary of Evidence

**Initial Management Decision: Active Surveillance vs Therapeutic Intervention**
For individuals who have clinically localized prostate cancer who receive ProLaris, the evidence includes 1 study of analytic validity and retrospective cohort studies of clinical validity using archived samples and a decision curve analysis providing indirect evidence of clinical utility. Relevant outcomes include overall survival, disease-specific survival, test accuracy and validity, quality of life, and treatment-related morbidity. Evidence of improved clinical validity or prognostic accuracy for prostate cancer death using ProLaris Cell Cycle Progression score in patients managed conservatively after needle biopsy has shown some improvement in areas under the receiver operating characteristic curve over clinicopathologic risk stratification tools. All validation studies are Simon category C. There is limited indirect evidence for potential clinical utility. The evidence is insufficient to determine the effects of the technology on health outcomes.

For individuals who have clinically localized prostate cancer who receive Oncotype DX Prostate, the evidence includes 2 studies of analytic validity, case-cohort and retrospective cohort studies of clinical validity using archived samples, and a decision curve analysis examining indirect evidence of clinical utility. Relevant outcomes include overall survival, disease-specific survival, test accuracy and validity, quality of life, and...
Gene Expression Profiling and Protein Biomarkers for Prostate Cancer Management

Policy # 00403  
Original Effective Date: 02/19/2014  
Current Effective Date: 06/21/2017

treatment-related morbidity. Evidence for clinical validity and potential clinical utility of Oncotype DX Prostate in patients with clinically localized prostate cancer derives from a study predicting adverse pathology after radical prostatectomy. The validity of using tumor pathology as a surrogate for risk of progression and cancer-specific death is unclear. It is also unclear whether results from an radical prostatectomy population can be generalized to an active surveillance population. All validation studies are Simon category C or D. The evidence is insufficient to determine the effects of the technology on health outcomes.

For individuals who have clinically localized prostate cancer who receive the ProMark protein biomarker test, the evidence includes 1 study of analytic validity, 1 retrospective cohort study of clinical validity using archived samples, and no studies of clinical utility. Relevant outcomes include overall survival, disease-specific survival, test accuracy and validity, quality of life, and treatment-related morbidity. There is insufficient evidence to support improved outcomes with ProMark given that only a single clinical validity study was available. The evidence is insufficient to determine the effects of the technology on health outcomes.

ManagementDecision After Radical Prostatectomy

For individuals who have intermediate- or low-risk prostate cancer after radical prostatectomy who receive Prolaris, the evidence includes 1 study of analytic validity and retrospective cohort studies of clinical validity using archived samples. Relevant outcomes include overall survival, disease-specific survival, test accuracy and validity, quality of life, and treatment-related morbidity. Evidence of improved clinical validity or prognostic accuracy for prostate cancer death using the Prolaris Cell Cycle Progression score in patients postprostatectomy has shown some improvement in areas under the receiver operating characteristic curve over clinicopathologic risk stratification tools. All validation studies are Simon category C or D. The evidence is insufficient to determine the effects of the technology on health outcomes.

For individuals who have high-risk prostate cancer after radical prostatectomy who receive the Decipher prostate cancer classifier, the evidence includes 1 study of analytic validity, prospective and retrospective studies of clinical validity using overlapping archived samples, decision curve analyses examining indirect evidence of clinical utility, and prospective decision impact studies without pathology or clinical outcomes. Relevant outcomes include overall survival, disease-specific survival, test accuracy, test validity, quality of life, and treatment-related morbidity. The clinical validity of the Decipher genomic classifier has been evaluated in samples of patients with high-risk prostate cancer undergoing different interventions following radical prostatectomy. Studies reported some incremental improvement in discrimination. However, it is unclear whether there is consistent improved reclassification—particularly to higher risk categories—or whether the test could be used to predict which men will benefit from radiotherapy. All validation studies are Simon category C. The evidence is insufficient to determine the effects of the technology on health outcomes.

References


Gene Expression Profiling and Protein Biomarkers for Prostate Cancer Management

Policy #: 00043
Original Effective Date: 02/19/2014
Current Effective Date: 06/21/2017

42. MAQC Consortium. The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. Nat Biotechnol. Sep 2006;24(9):1151-1161. PMID 16964229


Policy #       00403  
Original Effective Date: 02/19/2014  
Current Effective Date: 06/21/2017  


Policy History  
Original Effective Date: 02/19/2014  
Current Effective Date: 06/21/2017  
02/06/2014 Medical Policy Committee review  
02/19/2014 Medical Policy Implementation Committee approval. New policy.  
06/04/2015 Medical Policy Committee review  
06/17/2015 Medical Policy Implementation Committee approval. Policy statement unchanged.  
08/03/2015 Coding update: ICD10 Diagnosis code section added; ICD9 Procedure code section removed.  
06/02/2016 Medical Policy Committee review  
06/20/2016 Medical Policy Implementation Committee approval. Promark and Decipher tests added to the policy. Policy statement updated by adding “protein biomarkers”. Title change.  
01/01/2017 Coding update: Removing ICD-9 Diagnosis Codes and CPT coding update  
06/01/2017 Medical Policy Committee review  

©2017 Blue Cross and Blue Shield of Louisiana  
An independent licensee of the Blue Cross and Blue Shield Association  
No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means,  
electronic, mechanical, photocopying, or otherwise, without permission from Blue Cross and Blue Shield of Louisiana.
Gene Expression Profiling and Protein Biomarkers for Prostate Cancer Management

Policy # 00403
Original Effective Date: 02/19/2014
Current Effective Date: 06/21/2017

06/21/2017 Medical Policy Implementation Committee approval. Coverage eligibility unchanged. Extensive updates to rationale and references.
08/01/2017 Coding update
Next Scheduled Review Date: 06/2018

Coding
The five character codes included in the Blue Cross Blue Shield of Louisiana Medical Policy Coverage Guidelines are obtained from Current Procedural Terminology (CPT)®, copyright 2016 by the American Medical Association (AMA). CPT is developed by the AMA as a listing of descriptive terms and five character identifying codes and modifiers for reporting medical services and procedures performed by physician.

The responsibility for the content of Blue Cross Blue Shield of Louisiana Medical Policy Coverage Guidelines is with Blue Cross and Blue Shield of Louisiana and no endorsement by the AMA is intended or should be implied. The AMA disclaims responsibility for any consequences or liability attributable or related to any use, nonuse or interpretation of information contained in Blue Cross Blue Shield of Louisiana Medical Policy Coverage Guidelines. Fee schedules, relative value units, conversion factors and/or related components are not assigned by the AMA, are not part of CPT, and the AMA is not recommending their use. The AMA does not directly or indirectly practice medicine or dispense medical services. The AMA assumes no liability for data contained or not contained herein. Any use of CPT outside of Blue Cross Blue Shield of Louisiana Medical Policy Coverage Guidelines should refer to the most current Current Procedural Terminology which contains the complete and most current listing of CPT codes and descriptive terms. Applicable FARS/DFARS apply.

CPT is a registered trademark of the American Medical Association. Codes used to identify services associated with this policy may include (but may not be limited to) the following:

<table>
<thead>
<tr>
<th>Code Type</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPT</td>
<td>81479, 81599, 84999</td>
</tr>
<tr>
<td></td>
<td>New code eff 1/1/17: 81539</td>
</tr>
<tr>
<td></td>
<td>New code eff 8/1/17: 0005U</td>
</tr>
<tr>
<td>HCPCS</td>
<td>No codes</td>
</tr>
<tr>
<td>ICD-10 Diagnosis</td>
<td>C61</td>
</tr>
</tbody>
</table>

*Investigational – A medical treatment, procedure, drug, device, or biological product is Investigational if the effectiveness has not been clearly tested and it has not been incorporated into standard medical practice. Any determination we make that a medical treatment, procedure, drug, device, or biological product is Investigational will be based on a consideration of the following:

A. Whether the medical treatment, procedure, drug, device, or biological product can be lawfully marketed without approval of the U.S. FDA and whether such approval has been granted at the time the medical treatment, procedure, drug, device, or biological product is sought to be furnished; or

B. Whether the medical treatment, procedure, drug, device, or biological product requires further studies or clinical trials to determine its maximum tolerated dose, toxicity, safety, effectiveness, or effectiveness as compared with the standard means of treatment or diagnosis, must improve health outcomes, according to the consensus of opinion among experts as shown by reliable evidence, including:

1. Consultation with the Blue Cross and Blue Shield Association TEC or other nonaffiliated technology evaluation center(s);
2. Credible scientific evidence published in peer-reviewed medical literature generally recognized by the relevant medical community; or
3. Reference to federal regulations.

† Indicated trademarks are the registered trademarks of their respective owners.
Gene Expression Profiling and Protein Biomarkers for Prostate Cancer Management

Policy #  00403
Original Effective Date: 02/19/2014
Current Effective Date: 06/21/2017

NOTICE: Medical Policies are scientific based opinions, provided solely for coverage and informational purposes. Medical Policies should not be construed to suggest that the Company recommends, advocates, requires, encourages, or discourages any particular treatment, procedure, or service, or any particular course of treatment, procedure, or service.