Use of genomic markers to risk stratify men with prostate cancer

JOHN W. DAVIS

New genomic biomarkers are being developed to refine the risk for men with prostate cancer, but for healthcare professionals to believe in them, they must be both valid and of clinical utility. John Davis reviews three of these biomarkers, providing advice on how to interpret results and make decisions.

© KTSDesign/Science Photo Library

John W. Davis, *MD, Associate Professor of Urology, University of Texas MD Anderson Cancer Center, Houston, Texas, USA*

Healthcare professionals look to high-impact journals and meetings to improve their knowledge of novel trends in their specialty and to evaluate whether new information will build upon their current management strategies. As new information is presented, there is certainly the validity question (is it true, accurate, free of bias?), followed by the clinical utility question that eventually boils down to the question: 'do you believe in it?'

The management of localised prostate cancer has recently been challenged with several genomic biomarkers now commercially available in the USA (limited availability in Europe), with plans for additional global expansion. The 'do you believe?' question is an interesting area of study, as it challenges us to re-look at our current clinical management, expose unmet needs, and look for clinical decision points where risk refinement can be useful. It also challenges us to re-think biomarkers as stages of development and with proper nomenclature.

BIOMARKER DEVELOPMENT

Discovery

We all understand that biomarkers have to be discovered to exist, and there are numerous discovery-level publications on various biomarkers and using various techniques. The key questions at this phase are:

- what is the exact technique being used and the discovered marker(s)?
- why do we think this marker will compete well with other concepts in the field, *ie* other markers, imaging, *etc*?

Many markers are discovered on basic immunohistochemistry staining platforms, and may or may not ever be pursued into a commercially available product.

Validation

Next, biomarkers must be validated in several appropriate populations. This step is more of a statistical exercise, and the markers must look valuable in univariate and multivariate models, and must uniquely improve upon existing predictive models.

Clinical utility

A biomarker must also show clinical utility: do healthcare professionals make different and improved decisions using the marker? Do the markers reinforce decisions that were previously made with uncertain data? How often does a biomarker show a significantly more favourable or unfavourable result, compared with a 'consistent with clinical findings' result? This is where the answer to the 'do you believe?' question can be yes or no – even with statistical validity established. In part, this may be related to the amount of 'needle motion' a biomarker produces (what is the typical range of results obtained?) and how often such needle motion occurs outside of clinical prediction.

The needle motion of a biomarker can also be thought of in terms of the negative or positive predictive features, *ie* how favourable it can read, and how unfavourable it can read. The ideal biomarker would perform well on both ends of the scale, but it may be clinically useful if a biomarker performs really well on one end of the scale, and weakly on the other.

Hypothetical biomarker

Let me illustrate with a hypothetical biomarker and performance numbers. If we discovered and validated a biomarker that predicted whether or not a patient with low-risk prostate cancer could remain on active surveillance for 10 years free from upgrading and delayed treatment, we would first begin answering that question with clinical features: approximately 25%.[1] This is essentially a prevalence of disease starting assumption. If our hypothetical biomarker had dual positive and negative predictive power, the low end of the result would re-stratify patients as having <5% upgrading/treatment, and the high end would have >95%. Based upon the baseline prevalence of upgrading at 25%, we would expect only the 25% of upgraded patients to score in that high range.

Such a powerful biomarker would be easy to believe in, as in post-validation use one could easily order tests on 10 patients and show dramatic risk refinement. Even if the hypothetical biomarker had one-way value, it might be easy to believe in. If the marker were strong on negative predictive power, it might show the <5% upgrading/treatment risk versus the high end results showing 50% upgrading (but not >95%). Therefore, the 'do you believe?' question relates to whether or not you expect novel biomarkers to provide statistical superiority for large populations (one extreme) or to provide prediction in absolute terms that can be easily applied to one person (the other extreme).

Nomenclature

Finally, a nomenclature point for the clinical utility question relates to whether or not the endpoint of a biomarker provides pure risk refinement versus a link to therapy.[2] If a biomarker gives a risk of biochemical recurrence after a radical prostatectomy, it is often considered *prognostic*, and analogous to a weather forecast that you can put a percentage on but cannot change. If a biomarker predicts biochemical recurrence after radical prostatectomy plus success for adjuvant radiation therapy, it can be considered *predictive*. Given the implications of a powerful *predictive* biomarker, clinical utility is a built-in feature, whereas with *prognostic* biomarkers, clinical utility must be separately assessed.

COMMERCIALLY AVAILABLE BIOMARKERS

With this background, we can now transition to a concise review of three commercialised genomic biomarkers in prostate cancer and illustrate these practical points. I will present in alphabetical order of the commercial product name. Table 1 provides key comparisons for these first three products on the market.

Decipher

The Decipher test from GenomeDx (San Diego, CA, USA) is a 22-gene panel corresponding to RNAs from coding and non-protein-coding regions of the genome.[3] It can be tested on radical prostatectomy tissue to address the question as to the likelihood of lymph node or bone metastases developing in the first 5 years after surgery. A low score might indicate observation is the better plan, while a high score might indicate that early (adjuvant) radiation therapy is better.

A retrospective analysis from Thomas Jefferson Medical College tested Decipher scores among men after radical prostatectomy who underwent adjuvant or salvage radiation therapy.[4] They found that men with lower Decipher scores equally benefited from early or delayed radiation, while men with higher Decipher scores benefited greater from early radiation.

In a case simulation study, the knowledge of Decipher scores increased the recommendation for observation by 20% for low scores, and increased treatment recommendations by 16% for higher scores.[5]

In summary, this test is custom designed for the highest-risk cases where early metastatic progression is a concern, rather than biochemical recurrence only, and the test result is a straightforward prognosis of the event.

Oncotype Dx

The Oncotype Dx test from Genomic Health (Redwood City, CA, USA) focuses on the

	Decipher	Oncotype Dx	Prolaris
Tissues tested	RP for high risk: pT3, positive margin, PSA rise	Biopsy: for NCCN very-low to intermediate risk	Biopsy or RP
Clinical endpoints	Early regional nodes or bone metastasis	Risk of unfavourable pathology: pT3 and/or Gleason ≥4+3	Biopsy: 10-year mortality with conservative management RP: biochemical recurrence risk
Clinical utility	Adjuvant/salvage therapy	Active surveillance or immediate therapy	Biopsy: active surveillance or immediate therapy RP: adjuvant/ salvage therapy
Cost (USD)	$4250	$3825	$3400
RP, radical prostatectomy; pT3, pathologic stage with extraprostatic extension and/or seminal vesicle invasion; NCCN, National Comprehensive Cancer Network; USD, United States dollar equivalent.			

Table 1. Comparison of key features of three commercialised genomic tests for prostate cancer (adapted from Davis[3])

opposite side of the spectrum – the biopsy result showing favourable disease. In current times, the long-term mortality benefit from treating low-risk and low-volume Gleason 3+4 is questionable. Yet, anywhere from 20 to 30% of such cases can be upgraded at radical prostatectomy.

The Oncotype Dx test is a 12 cancer-related gene panel that reflects several pathways – stromal response, cellular organisation, androgen signalling and proliferation.[6] The test gives a direct prognosis for the finding of adverse pathology at radical prostatectomy, such as pathologic T3 disease and/or upgrading the Gleason 4+3 or higher. The test will give a unique genomic prostate score (GPS) on a scale of 0–100. That score is then translated to a percentage risk of having the unfavourable pathology, and it is given a categorical name such as 'very low', 'low' or 'intermediate' risk for unfavourable pathology.

This test reflects the above discussion on negative versus positive predictive power. A low GPS score of 0–10, for example, can rate unfavourable pathology as <5%; however, a very high GPS score in the 50 range only makes the prognosis 50–60% range. Therefore, one might have a stronger confidence ('belief') in a very low score in terms of recommending active surveillance – especially in an older patient where the mortality risk from unfavourable pathology may be less.

Prolaris

The Prolaris test from Myriad Genetics (Salt lake City, UT, USA) is a 46-gene panel of cell cycle progression genes that measures proliferation as cells go into their division cycles. The general concept was developed in breast cancer and then validated retrospectively in two large prostatectomy data sets and a large cohort of men placed on observation.[7-9]

The test can be run on biopsy or radical prostatectomy tissue, and due to the validation datasets, the endpoints reported are different. The test is also different from Decipher, in that the prognosis reported takes into account the clinical parameters. For the common need to predict outcomes on active surveillance, a Prolaris score for a low-risk patient will report a 10-year probability of cancer-related mortality with only conservative management.

The range of results can be as low as 1% or as high as 11%, with 4.5% being the average. These ranges of results have been criticised as being too limited. However, the low scores can certainly reinforce the surveillance decision – especially in older or comorbid patients with shorter life expectancies. It may seem minor that one test generates a 3% mortality and another a 6% mortality, but in the context of low-risk disease, we should not be expecting anyone to have 95% mortality, or we should be reclassifying the disease.

For intermediate- and high-risk disease, the ranges are much higher (3–40% type ranges), and there may be increased clinical utility here. In fact, the high range of Prolaris in intermediate risk might have greater mortality than the lower range of Prolaris in high risk, indicating that the test can have independent prognostic power apart from clinical features. The Prolaris score can also be run on radical prostatectomy tissue and the output will be a 10-year biochemical recurrence risk, which might be useful to make postoperative radiotherapy decisions.

INTERPRETING RESULTS AND MAKING DECISIONS

As indicated with each marker, a test result can score very high or very low, and be correlated with a clear clinical decision. The odds of this occurring for your individual patient may vary based upon the prevalence of such disease in the population actually tested. A key point highlighted in clinical utilities studies is that some tests are 'consistent' with clinical information, just as some imaging studies we order are normal or rule out a condition of interest.

Approximately 50% of Prolaris studies ordered in post-commercial release studies are consistent with clinical prognostic information, while the remaining are more or less aggressive than clinical prognostic information (company data, unpublished). When translated to clinical decisions, Crawford *et al.*[10] found that 40% reduced therapy recommendations, 35% were no change and 25% increased therapy recommendations. GenomeDx reports a similar metric that in their validation studies, the high-risk test score is present in 19% of the population tested, while 60% are in the low risk (unpublished).

It is unhelpful to consider 'consistent' results as 'unnecessary' or 'useless' as there are ranges of interpretation possible. A very young patient may have treatment as a strong default position and only a very less aggressive Prolaris or Oncotype Dx score would change that decision. A much older patient might have surveillance as a very strong default, and only a very aggressive Prolaris or Oncotype Dx score would change that decision. In other cases, the test may just provide reinforcement of a decision.

We have encountered two good recent examples. A 50-year-old male with low-risk disease and morbid obesity (body mass index >55) tested with Prolaris in the very less aggressive range. This helped confirm our preference to work on weight loss first and defer therapy to another time. A 75-year-old with Gleason 3+4 requested radical prostatectomy, but had significant comorbidities, including recent deep-vein thrombosis/pulmonary embolism. A considerably less aggressive Prolaris score helped convince him to go on active surveillance. In this latter case, it is important to point out that the surgeon voted for active surveillance before and after Prolaris testing, while the patient was the decider that changed his mind. This is a good example of shared decision-making, and emphasises that clinical utility can be measured at the level of the physician, patient or combined.

In the post-radical prostatectomy testing (Prolaris or Decipher), the unmeasured variable will be the patient's acceptance for further therapy (*ie* radiation), which may relate to the quality of his recovery from surgery. Therefore, if a patient has perfect continence and potency, he may be more willing to be tested for risk of recurrence and proceed with adjuvant therapy. If recovery is still evolving, the patient may accept only observation and possible salvage radiation no matter what, and testing would be more 'academic' at this point. That said, I have observed patients with both types of testing who were unwilling to undergo adjuvant therapy, but due to a high-risk Prolaris or Decipher test, they planned very frequent PSA testing with salvage radiation at the very earliest of indications, *ie* PSA detectable at 0.1ng/ml. This has occurred for one of the patients.

WHAT CAN WE EXPECT IN THE FUTURE?

At present, the 'do you believe?' question is answered very differently by physicians. I have not observed anyone adopt a practice pattern of routine use for all, but rather selected use for specific situations (myself, as the anecdotes above illustrate) and certainly some do not find utility in the current products, or are unaware of them. This result may seem curious as all three products can show statistical validity and superior metrics to clinical variables. Yet they do come at a cost (see Table 1) and there may be varying expectations as to how powerful a biomarker should be at changing default clinical decisions.

With time, we hope that genomic markers will be more specifically linked to a therapy, such that the decisions are measured at that level rather than in only prognostic percentages. Genomic markers might also migrate into prospective clinical trials such that they can be inclusion/exclusion factors that enhance comparative cohorts for the desired endpoints. Thus, a high-risk surgery trial with neoadjuvant therapy would require a genomic classification of the highest-risk disease.

Declaration of interests

JWD has speaking agreements with Myriad Genetics and Intuitive Surgical, and is involved in a scientific trial for GenomeDx.

REFERENCES

1. Cooperberg MR, Carroll PR, Klotz L. Active surveillance for prostate cancer: progress and promise. *J Clin Oncol* 2011;29:3669–76.
2. Oldenhuis CNAM, Oosting SF, Gietema JA, *et al.* Prognostic versus predictive value of biomarkers in oncology. *Eur J Cancer* 2008;44:946–53.
3. Davis JW. Novel commercially available genomic tests for prostate cancer: a roadmap to understanding their clinical impact. *BJU Int* 2014;114:320–2.
4. Den RB, Feng FY, Showalter TN, *et al.* Genomic prostate cancer classifier predicts biochemical failure and metastases in patients after postoperative radiation therapy. *Int J Radiat Oncol Biol Phys* 2014;89:1038–46.
5. Badani KK, Thompson DJ, Brown G, *et al.* Effect of a genomic classifier test on clinical practice decisions for patients with high-risk prostate cancer after surgery. *BJU Int* 2015;115:419–29.
6. Klein EA, Cooperberg MR, Magi-Galluzzi C, *et al.* A 17-gene assay to predict prostate cancer aggressiveness in the context of Gleason grade heterogeneity, tumor multifocality, and biopsy undersampling. *Eur Urol* 2014;66:550–60.
7. Cuzick J, Berney DM, Fisher G, *et al.* Prognostic value of cell cycle progression signature for prostate cancer death on conservatively managed needle biopsy cohort. *Br J Cancer* 2012;106:1095–9.
8. Cuzick J, Berney DM, Fisher G, *et al.* Prognostic value of an RNA expression signature derived from cell cycle proliferation genes in patients with prostate cancer: a retrospective study. *Lancet Oncol* 2011;12:245–55.
9. Cooperberg MR, Simko JP, Cowan JE, *et al.* Validation of a cell-cycle progression gene panel to improve risk stratification in a contemporary prostatectomy cohort. *J Clin Oncol* 2013;31:1428–34.
10. Crawford ED, Scholz MC, Kar AJ, *et al.* Cell cycle progression score and treatment decisions in prostate cancer: results from an ongoing registry. *Curr Med Res Opin* 2014;30:1025–31.