A whole-blood RNA transcript-based prognostic model in men with castration-resistant prostate cancer: a prospective study

Robert W Ross, Matthew D Galsky*, Howard I Scher, Jay Magidson, Karl Wassmann, Gwo-Shu Mary Lee, Leah Katz, Sumit K Subudhi, Aseem Anand, Martin Fleisher, Philip W Kantoff, William K Oh*

Summary

Background Survival for patients with castration-resistant prostate cancer is highly variable. We assessed the effectiveness of a whole-blood RNA transcript-based model as a prognostic biomarker in castration-resistant prostate cancer.

Methods Peripheral blood was prospectively collected from 62 men with castration-resistant prostate cancer on various treatment regimens who were enrolled in a training set at the Dana-Farber Cancer Institute (Boston, MA, USA) from August, 2006, to June, 2008, and from 140 patients with castration-resistant prostate cancer in a validation set from Memorial Sloan-Kettering Cancer Center (New York, NY, USA) from August, 2006, to February, 2009. A panel of 168 inflammation-related and prostate cancer-related genes was assessed with optimised quantitative PCR to assess biomarkers predictive of survival.

Findings A six-gene model (consisting of *ABL2*, *SEMA4D*, *ITGAL*, and *C1QA*, *TIMP1*, *CDKN1A*) separated patients with castration-resistant prostate cancer into two risk groups: a low-risk group with a median survival of more than 34·9 months (median survival was not reached) and a high-risk group with a median survival of 7·8 months (95% CI 1·8–13·9; p<0·0001). The prognostic utility of the six-gene model was validated in an independent cohort. This model was associated with a significantly higher area under the curve compared with a clinicopathological model (0·90 [95% CI 0·78–0·96] *vs* 0·65 [0·52–0·78]; p=0·0067).

Interpretation Transcriptional profiling of whole blood yields crucial prognostic information about men with castration-resistant prostate cancer. The six-gene model suggests possible dysregulation of the immune system, a finding that warrants further study.

Funding Source MDX.

Lancet Oncol 2012; 13: 1105–13

Published Online October 9, 2012 http://dx.doi.org/10.1016/S1470-2045(12)70263-2

See Comment page 1067

*RWR and MG contributed equally to this report

Division of Solid Tumor Oncology, Department of Medicine, Dana-Farber Cancer Institute, Boston, MA, USA (R W Ross MD, G-S M Lee PhD, L Katz MD, Prof P W Kantoff MD); **Harvard Medical School, Boston, MA, USA** (R W Ross, G-S M Lee, L Katz, P W Kantoff); **Division of Hematology/Oncology, Department of Medicine, Tisch Cancer Institute, Mount Sinai School of Medicine, New York, NY, USA** (M D Galsky MD, Prof W K Oh MD); **Genitourinary Oncology Service, Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, NY, USA** (Prof H I Scher MD, S K Subudhi MD, A Anand MD); **Weill Cornell Medical College, New York, NY, USA** (H I Scher, S K Subudhi, A Anand); **Statistical Innovations, Belmont, MA, USA** (J Magidson PhD); **GeneNews, Boston, MA, USA** (K Wassmann MBA); **and Department of Clinical Laboratories, Memorial Sloan-Kettering Cancer Center, New York, NY, USA** (M Fleisher PhD)

Correspondence to: Prof William K Oh, Division of Hematology and Medical Oncology, Mount Sinai School of Medicine, Tisch Cancer Institute, One Gustave L Levy Place, Box 1079, New York, NY 10029, USA **william.oh@mssm.edu**

Introduction

Castration-resistant prostate cancer is a strikingly heterogeneous disease state that affects patients with varying metastatic burden and symptoms.[1] As a result of this heterogeneity, the overall survival of patients with castration-resistant prostate cancer can be extremely variable, ranging from several months to several years. The ability to accurately predict prognosis in men with castration-resistant prostate cancer is crucial to assist with patient counselling and to optimise clinical-trial design and patient stratification.

Several studies have correlated clinical and laboratory variables, including age, functional status, extent of bone and other metastases, prostate-specific antigen (PSA), alkaline phosphatase, and lactate dehydrogenase, with survival in patients with castration-resistant prostate cancer.[2–4] Additionally, points-based nomograms have been developed combining these variables.[5,6] While such nomograms have improved the ability to individualise prognosis, they offer only moderate predictive discrimination, highlighting the need for improved models.

Interactions between blood cells and the peripheral tissue through which blood circulates, including neoplastic tissue, might alter the gene expression of blood cells. Indeed, recent studies have shown that gene-expression profiling of peripheral blood cells could yield diagnostic and prognostic information regarding various disease states.[7–10] Expression profiling of blood offers several practical advantages compared with expression profiling of tumour tissue, including the minimally invasive nature of sample acquisition, relative ease of standardisation of sampling protocols, and the ability to obtain repeated samples over time. In this study, we tested the hypothesis that transcriptional profiling of whole blood could yield prognostic information in men with astration-resistant prostate cancer.

Methods

Patient population

The training set comprised 62 patients with castration-resistant prostate cancer, with or without the presence of radiographic metastases, and on various treatment regimens, enrolled at the Dana-Farber Cancer Institute from August, 2006, to June, 2008, on a genitourinary oncology clinical database and biorepository protocol. Whole-blood samples were prospectively collected in PAXgene Blood RNA tubes (PreAnalytiX, Hombrechtikon,

Switzerland) from patients in this cohort. The validation set consisted of 140 patients with castration-resistant prostate cancer, with or without the presence of radiographic metastases, and on various treatment regimens, enrolled on a genitourinary oncology clinical database and biorepository protocol at Memorial Sloan-Kettering Cancer Center (New York, NY, USA). Banked PAXgene Blood RNA tubes were available from this latter cohort of patients, collected between August, 2006, and February, 2009. Two patients had poor quality RNA, leaving 138 patient samples for validation. Clinical data, including survival outcomes, were gathered as a standard component of each protocol, consistent with methods as previously described.[11] The protocols were approved by the institutional review boards at each participating institution and written informed consent was obtained from all participants.

Procedures

From 1998 to 2008 Source MDx did a systematic search for genes that express mRNA in whole blood that can be measured with commercial quantitative PCR. Candidate genes were selected on the basis of a review of published works and consultation with medical experts. During this period about 2000 such candidate genes were classed as having measurable gene expression on the basis of analysis of whole-blood samples collected in PAXgene Blood RNA tubes from normal volunteers.

See Online for appendix

Subsequently, pilot studies were done from 2005 to 2007 for seven major solid tumours (prostate, colon, lung, breast, ovarian, cervical, and melanoma) versus normal case controls establishing genes unique to specific cancers, common genes of significance across cancers, and the potential importance of the early-growth-response gene family.[12,13] The prostate-cancer pilot studies done during this period assessed 382 unique genes on the basis of the selection process described above (figure 1). These gene assays included cytokines, chemokines, cell signalling, and other genes that track inflammatory disorders and immune system response associated with cancer and genes that play a part in angiogenesis, apoptosis, tumour suppression, cell cycle, DNA repair, and tumour progression.

The preliminary exploratory phase included two independent cohorts (figure 1). A pilot study in Russia compared the gene expression differences for newly diagnosed, non-treated patients with early stage prostate cancer (N=50) and blood samples with healthy normal controls (N=50); whole-blood samples were also collected in PAXgene tubes at the same site for both cases and controls.

A second pilot study done at the Dana-Farber Cancer Institute (Boston, MA, USA) similarly collected and compared PAXgene blood samples from newly diagnosed, untreated patients with prostate cancer (N=40) with blood sample collected from healthy normal controls in Colorado, USA. Collectively, these studies yielded a 168-gene panel that distinguished gene expression of patients with prostate cancer from healthy normal controls in whole blood (appendix p 3). The 168-gene panel was used in this current study, as well as for further assessment of detection of early stage prostate cancer (appendix p 2).[14]

Whole-blood samples were collected in PAXgene Blood RNA tubes and were processed to total RNA within 30 days of phlebotomy with the Qiagen PAXgene Blood RNA Kit (Qiagen, Valencia, CA, USA). Representative RNA samples met metrics for quality and integrity (RNA integrity number [RIN]≥6·3) according to the Bioanalyzer 2100 (Agilent Technologies, Palo Alto, CA, USA) in combination with the RNA 6000 Nano or Pico Series II LabChip (Agilent Technologies, Palo Alto, CA, USA).

First-strand complementary DNA was synthesised from random hexamer-primed RNA templates using

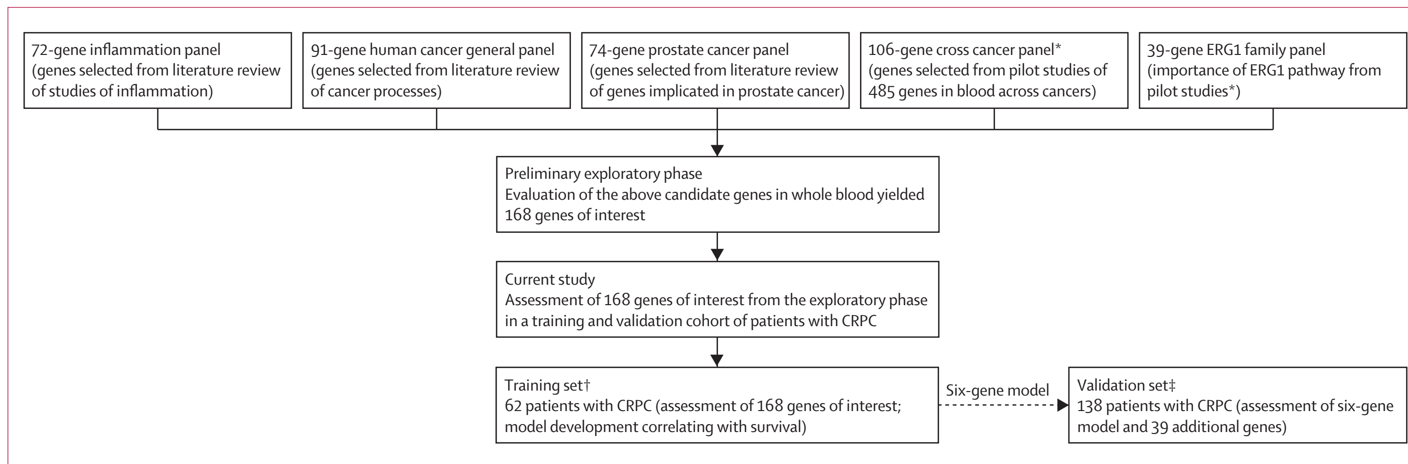

Figure 1: **Study profile**

CRPC=castration-resistant prostate cancer. *Pilot studies assessing expression of 485 genes in whole blood from patients with seven major solid tumours. †Dana-Farber Cancer Institute. ‡Memorial Sloan-Kettering Cancer Center.

TaqMan (Applied Biosystem, Division of Life Technologies Corporation, Carlsbad, CA, USA) reverse transcription reagents. We undertook quantitative PCR analysis of the 18S rRNA content of newly synthesised complementary DNA, with the 7900HT fast real-time PCR System (Applied Biosystems, Foster City, CA, USA) as a quality check of the first-strand synthesis reaction.

Target-gene amplification was done in a quantitative PCR reaction with TaqMan Universal PCR Master Mix (Applied Biosystem, Division of Life Technologies Corporation, Carlsbad, CA, USA) and Precision Profiles (Source MDx, Boulder, CO, USA). Individual target-gene amplification was multiplexed with the eukaryotic 18S rRNA endogenous control and run in triplicate in a 384-well format on the 7900HT fast real-time PCR system. For quality control, all replicate cycle threshold (CT) values (both target gene and endogenous control) were independently checked and automatically filtered by rule. Normalised CT values (ΔCT) for each amplified target gene replicate were calculated. Resulting triplicate ΔCT values for individual target genes were averaged yielding a final ΔCT value. Additional details regarding specimen collection, processing, and analysis are detailed in the appendix (p 4).

Statistical analysis

We tested the hypothesis that whole-blood RNA transcript-based prognostic biomarkers can classify patients with castration-resistant prostate cancer into low-risk and high-risk groups on the basis of a risk score. The relation between gene expression and the time to death was examined with Cox proportional hazard models. Models were generated (one-gene and two-gene) yielding the best prediction of the survival of patients with castration-resistant prostate cancer. A gene might not be predictive on its own but could contribute to a model by enhancing the effect of another gene in the model.[14,15] To allow inclusion of such synergistic gene pairs in a model, since such genes cannot be identified by examining one-gene models alone, we examined all possible one-gene and two-gene models.

We developed separate models according to survival time measured from the date the patient was determined to be castration-resistant and the date of blood draw. Since patient age was not significantly related to survival time based on either definition, models were developed on the basis of gene expression data alone. Final gene models summarised and interpreted were those for which all genes in the model were incrementally significant at the 0·05 level. Various comparisons were made between different models and examined for consistency. Additionally, the p values for the significant gene effects were compared and examined for consistent patterns when the survival time was measured from the time the patient was determined to be castration resistant, as opposed to the time of the blood draw.

Since there were more genes (168) than patients (62), to reduce the likelihood of overfitting, our final model included the best six genes obtained from the best two-gene models, for which best was defined by taking all models that contained significant effects, and rank ordering them according to their entropy R^2.[16] The best two-gene models containing unique genes were used to select the genes included in a final six-gene Cox hazard model.

Since heterogeneity is common in hazard models, and the omission of such can create biased results,[17] we applied latent class modelling to examine whether distinct subpopulation segments had different expected survival times. Latent class (also known as finite mixture) models have been shown to be useful in determining both the size of these different segments (risk groups) and classifying patients into the most appropriate risk group.[18,19] In medicine, such models have frequently

	Training set	Validation set
Number of patients	62	138
Age, years*	68 (49–86)	72 (48–93)
Gleason score at diagnosis	7 (5–10)	NA
ECOG performance status	0 (0–3)	NA
PSA (ng/mL)	21 (0·04–3905)	65·8 (<0·05–3096)
Haemoglobin (g/dL)	12·6 (8·5–18·4)	NA
Alkaline phosphatase (IU/L)	93 (27–1565)	99 (29–1818)
Metastatic disease	54 (87%)	124 (90%)
Previous chemotherapy	21 (34%)	62 (45%)
Visceral metastases†	16 (29%)	NA
Duration of CRPC at time of blood draw (months)	16 (0–93)	NA
Median follow-up (months)‡	11 (7–15)	30 (22–40)

Data are number, number (range), median (range), or number (%). NA=not available. ECOG=Eastern Cooperative Oncology Group. PSA=prostate-specific antigen. CRPC=castration-resistant prostate cancer. *Age at time of blood draw. †Data not available for six patients. ‡Median follow-up calculated from time of blood draw, according to Schemper and Smith.[24]

Table 1: **Baseline patient characteristics**

Regression coefficient (gene 1)	Gene 1	Regression coefficient (gene 2)	Gene 2	Entropy R^2
2·3	*ABL2*	–1·3	*C1QA*	0·21
3·4	*SEMA4D*	–2·4	*TIMP1*	0·19
6·8	*SEMA4D*	–4·7	*MYD88*	0·18
7·4	*SEMA4D*	–4·2	*SVIL*	0·17
2·3	*ITGAL*	–2·0	*CDKN1A*	0·17
1·6	*ABL2*	–1·0	*C1QB*	0·17
3·1	*ABL2*	–3·0	*PYCARD*	0·17
3·4	*ABL2*	–2·7	*MNDA*	0·17

Cox p values for individual genes: *ABL2*, p=0·0001; *CDKN1A*, p=0·0007; *C1QA*, p=0·0010; *SEMA4D*, p=0·0025; *TIMP1*, p=0·0312; *ITGAL*, p=0·0516.

Table 2: **Top eight two-gene models**

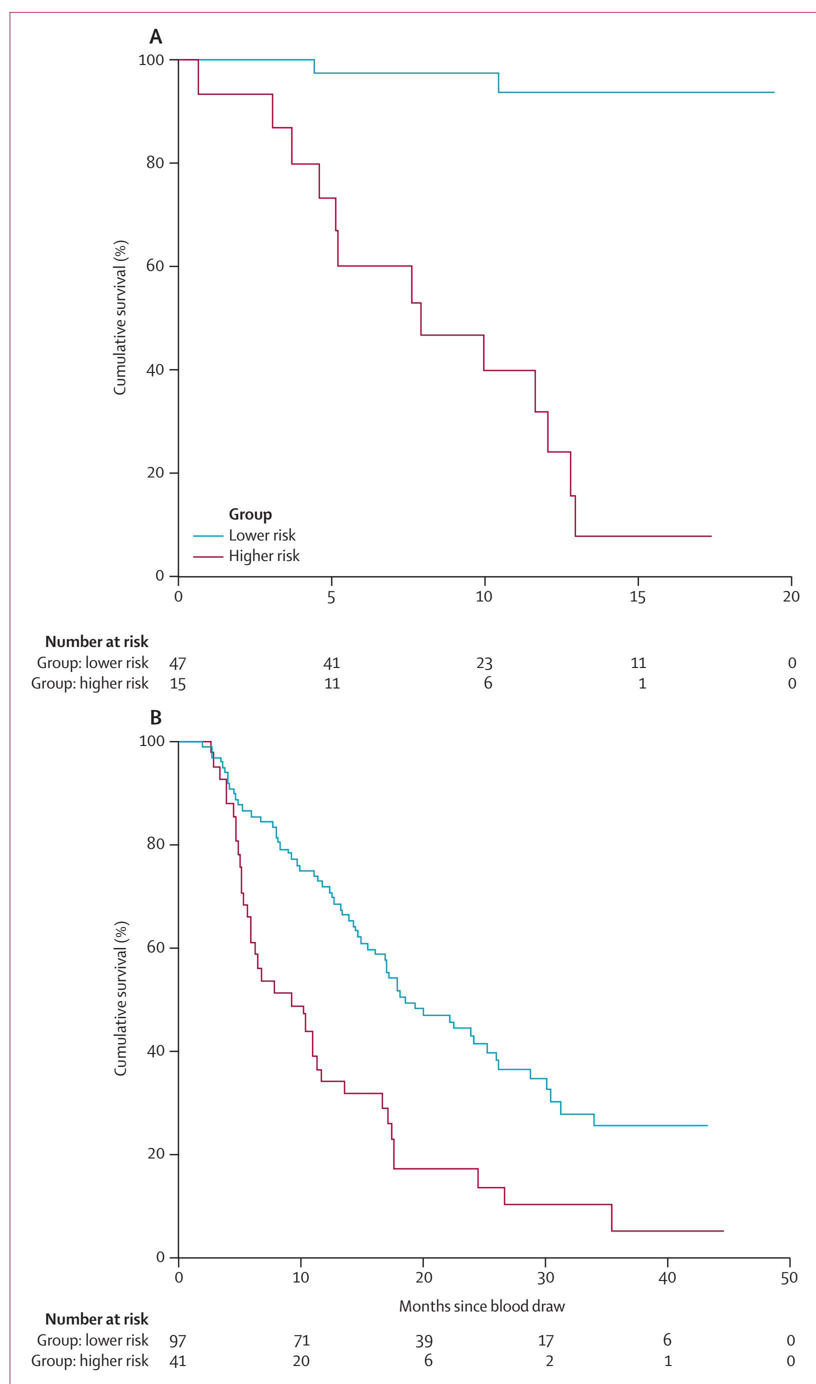

Figure 2: **Kaplan-Meier assessment of prediction of the six-gene model in the training set (A) and validation set (B)**
In training and validation sets, survival for lower-risk group is significantly longer than higher-risk group, p<0·0001.

been used to identify the characteristics of long-term survivors.[20,21] Software incorporating general maximum likelihood algorithms for estimating such random-effects models includes GLLAMM (version 8)[22] and Latent GOLD (version 4.5).[23] Here we used the latter programme because of a regularisation option, which we used to restrict the baseline hazard function for each risk group to be a non-decreasing function of time.

Stepwise Cox regression analyses were used to compare the prognostic ability of the six-gene model versus clinicopathological variables. To assess the incremental improvement of the clinicopathological variables versus the six-gene model, the six-gene-model score was entered first and each clinicopathological variable was allowed to enter, one at a time, if its incremental p value was less than 0·05. To assess the incremental improvement of the six-gene model compared with a clinicopathological model, the clinical variables were allowed to enter first until no longer significant and then the six-gene model score was allowed to enter if significant.

On the basis of the analysis of 168 target genes in the training dataset, the resulting six-gene model was tested on an independent validation dataset. Validation was done on a masked basis. Individual model risk scores were computed for each individual patient according to the prespecified coefficients established in the training set for the six-gene model. Additionally, a cutpoint developed in the training model was applied to split the samples into two groups (high-risk and low-risk groups) and Kaplan-Meier survival curves were generated for the two groups on the basis of the prespecified cutpoint. After appending these scores to the validation data, survival status information was unmasked and a log-rank p value was obtained from a Cox model with the dichotomous variable representing the two risk groups being used as a single covariate, with a p value of less than 0·05 resulting in a successful validation.

Role of the funding source

Source MDx contributed to the study design and analysed the study specimens. WKO, RWR, MDG, JM, KW, and PWK had access to the raw data. The corresponding author had full access to all of the data and had final responsibility for the decision to submit for publication.

Results

The baseline characteristics of patients in the training set and validation set were similar, though a slightly higher proportion of patients in the validation set had received previous chemotherapy (table 1).

Blood samples were available from all 62 patients in the training set for gene-expression profiling. As of June 20, 2008, a total of 47 patients in the training set were alive and 15 (24%) had died. All 168 target genes and all pairs of these genes were entered directly as predictors in Cox proportional hazards models. All resulting one-gene and two-gene models, after eliminating those with non-significant p values, were ranked from most to least significant (by their entropy R^2 value), on the basis of their association with survival status. The two-gene prognostic models did not depend on whether survival was measured from the time of blood draw or the time of development of castration resistance (table 2, appendix p 5).

A six-gene model (*ABL2, C1QA; SEMA4D, TIMP1*; and *ITGAL, CDKN1A*) was then constructed based on genes

Low risk: 2×*ABL2*+*SEMA4D*+*ITGAL* –*C1QA*–*TIMP1*–*CDKN1A*<21.21

High risk: 2×*ABL2*+*SEMA4D*+*ITGAL* –*C1QA*–*TIMP1*–*CDKN1A*>21.21

in the top two two-gene models (*ABL2, C1QA*; *SEMA4D, TIMP*) plus the unique genes in the fifth two-gene model (*ITGAL, CDKN1A*; table 2). The third two-gene model (*SEMA4D, MYD88*) and the fourth two-gene model (*SEMA4D, SVIL*) were not included because of the redundancy of *SEMA4D*. Latent class analysis, identifying long-term and short-term survival groups, was used to obtain a cutpoint for two risk groups,

Since the coefficient estimates were all close to 1, –1, or 2, they were therefore rounded to these numbers. The six-gene model was developed using June, 2008, census data and assessed on the basis of an updated census on October, 2009. This model separated patients with castration-resistant prostate cancer into two groups (figure 2 and table 3): high-risk men with a median survival of 7·8 months (95% CI 1·8–13·9) and low-risk (median survival of more than 34·9 months; median survival time not reached). The Kaplan-Meier in figure 2A shows the large difference in survival between the lower and high-risk men in the training set (hazard ratio [HR] 9·7, 95% CI 2·9–33·1; $p_{log\ rank}$<0·0001). The independent validation set also resulted in a highly significant difference (p<0·0001), as depicted in figure 2B.

The prognostic ability of the six-gene model was compared with a model including clinicopathological variables that have been shown to be of prognostic significance in patients with castration-resistant prostate cancer.[5] Of the seven variables included in the Halabi nomogram,[5] data were available for six variables (alkaline phosphatase, Eastern Cooperative Oncology Group [ECOG] performance status, haemoglobin, and presence of visceral metastases, PSA, and Gleason score) from 56 of 62 patients included in the training set (lactate dehydrogenase values not available). Of these 56 patients, 13 had died, and 43 were alive at the time of the analysis. Several Cox models were developed with survival measured since castration-resistant prostate cancer diagnosis and since blood draw, results from both of which were largely consistent. Only three of these variables were significant (ECOG performance status, haemoglobin, and PSA), and PSA was no longer significant when either of the other two variables were also included in the model (table 4).

A stepwise Cox regression of survival since blood draw based on the clinical variables resulted in a model with only two variables: ECOG performance status and haemoglobin. A comparison was made between the two risk groups obtained from this model estimated with the clinicopathological variables, with the two risk groups obtained from the six-gene model (appendix pp 6–9). For comparability with the six-gene model, a cutoff value for the model based on these two clinical variables was chosen to obtain two risk groups of the same size as the risk groups obtained with the six-gene model—43 low-risk patients and 13 high-risk subjects. The Kaplan-Meier analysis based on these two clinical groups yielded a p value of 0·0006 (HR 5·4, 95% CI 1·3–2·2). With the two risk groups obtained from the six-gene model, the Kaplan-Meier analysis yielded a p value less than 0·0001 (HR 21·9, 5·6–85·2; figure 3).

	Mean estimate	Median estimate
Training set		
Lower risk	30·1 (1·3; 27·6–32·6)	>34·9*
Higher risk	9·7 (1·8; 6·2–13·2)	7·8 (3·1; 1·8–13·9)
Overall	25·2 (1·5; 22·2–28·3)	32·3
Validation set		
Low risk	22·8 (1·6; 19·7–25·9)	18·5 (2·6; 13·5–23·6)
High risk	13·0 (1·8; 9·5–16·6)	9·2 (2·4; 4·4–14·0)
Overall	20·1 (1·3; 17·5–22·7)	16·6 (1·5; 13·8–19·5)

Data are mean estimate (SE; 95% CI) or median estimate (SE; 95% CI). Training set, p<0·0001; validation set, p<0·0001 (p values are comparing survival in high-risk *vs* low-risk groups). *Low-risk training set patients have a median survival of more than 34·9 months (a range estimate): a point estimate for the median survival was not available for the low-risk group because the percentage of low-risk patients predicted to survive longer than the maximum observed survival time during a follow-up census in October, 2009, was higher than 50%.

Table 3: **Training set and validation set mean and median survival (months) from blood draw (six-gene Cox model based on gene expression measurement)**

	Score	Degrees of freedom	p value
Time since CRPC diagnosis (months)			
ECOG performance status	4·53	1	0·03
Haemoglobin	6·49	1	0·01
Gleason	0·04	1	0·85
Visceral disease	2·71	1	0·10
Log (alkaline phosphatase)	0·62	1	0·43
Log (PSA)	5·39	1	0·02
Six-gene model	31·45	6	<0·0001
Time since blood draw (months)			
ECOG performance status	14·17	1	0·0002
Haemoglobin	7·24	1	0·007
Gleason	0·10	1	0·75
Visceral disease	1·91	1	0·17
Log (alkaline phosphatase)	0·15	1	0·70
Log (PSA)	6·01	1	0·01
Six-gene model	29·27	6	<0·0001

CRPC=castration-resistant prostate cancer. ECOG=Eastern Cooperative Oncology Group. PSA=prostate-specific antigen.

Table 4: **Cox model for survival measured from CRPC diagnosis or from time of blood draw**

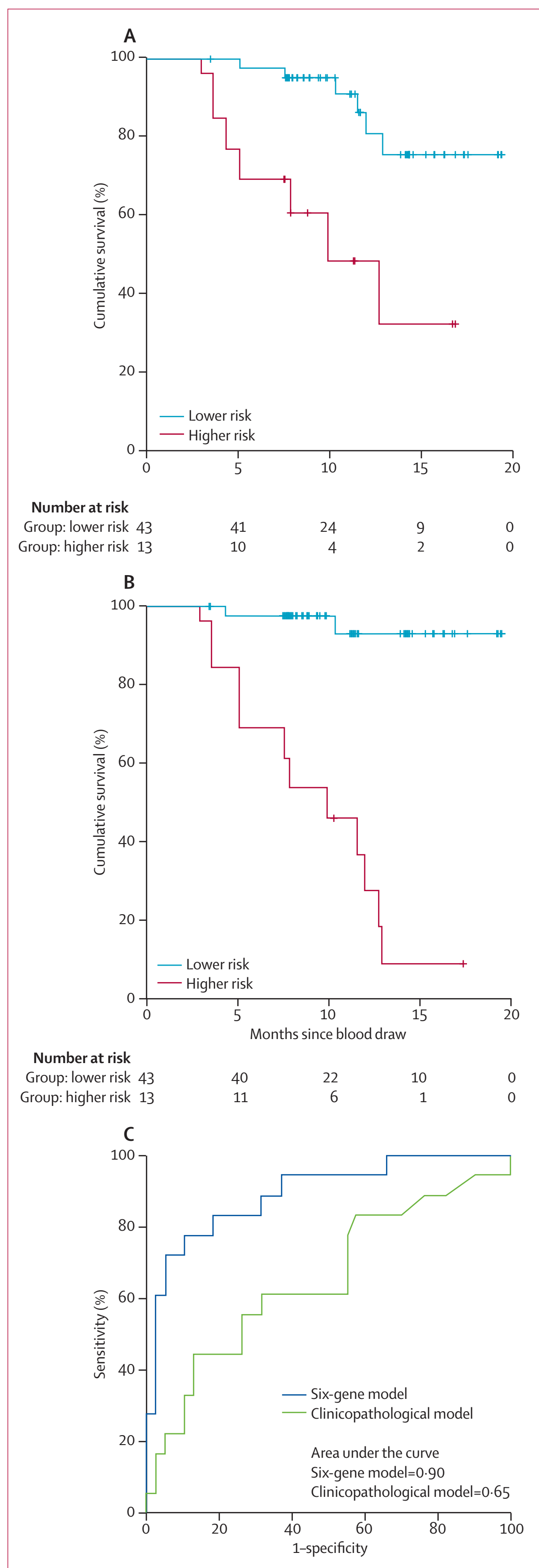

Figure 3: **Results from comparison of six-gene and clinicopathological models**
The following analyses are shown: clincopathological variables (A), six-gene signature (B), and the comparison of the receiver operating characteristic curve of the six-gene model versus the clinicopathological model based on 1·5-year survival of patients in the training set (C).

For increased power of detection of differences between the six-gene and clinicopathological models, we used an additional 16 months of follow-up data obtained from the October, 2009, census. Between June, 2008, and October, 2009, an additional ten patients had died, nine of whom had clinical data available, increasing the total deaths to 22 of 56 patients with clinical data. Receiver operating characteristic curves were generated for the clinicopathological model and six-gene model, based on the 1·5-year survival since blood draw for the patients in the training set (figure 3). The six-gene model was associated with a significantly higher (p=0·0067) area under the curve (0·90, 95% CI 0·78–0·96) than in the clinicopathologic model (0·65, 0·52–0·78). To establish whether the significant improvement of the six-gene model over the clinicopathological model was maintained if additional clinicopathological variables were included in the latter model, this difference analysis of the area under the curve was repeated with the best three-variable and four-variable models. In both cases, the results showed that the six-gene model was significantly better than the clinicopathological model (p=0·0072 for the three variable model; p=0·0034 for the four variable model; appendix pp 6–9).

An additional analysis was undertaken to determine whether the clinicopathological variables (visceral disease, alkaline phosphatase, ECOG performance status, haemoglobin) could improve on the gene model. The six-gene model score was entered first into a Cox model and the clinicopathological variables allowed to enter in a stepwise manner (table 5). When survival was measured from blood draw, none of these variables were incrementally significant. When survival was measured from date of castration-resistant prostate cancer diagnosis, only the presence of visceral metastatic disease entered into the model (p=0·054), increasing the overall model score χ^2 statistic from 35·1 to 38·1 (on the basis of 56 patients), an increase which can be explained by chance alone.

To assess the incremental improvement of the six-gene model over the clinicopathological variables, a model based solely on the clinicopathological variables yielded significant results according to whether survival was measured from castration-resistant prostate cancer diagnosis (p=0·008) or blood draw (p=0·002). The six-gene model (based on the same set of 56 patients) also showed significant results according to whether survival time was measured from castration-resistant prostate cancer diagnosis (adjusted p<0·0001) or blood draw (adjusted p=0·0003), where the associated p values reported in table 5 were adjusted to take into account six degrees of freedom for the six distinct genes in the model.

Gene-expression analysis of 138 patients with castration-resistant prostate cancer was done in the validation set. At the time of the analysis, 97 (70%) of 138 patients had died. The results validated that the

six-gene model developed in the training set successfully predicted survival status in this external cohort of patients with castration-resistant prostate cancer and separated patients into a high-risk and low-risk group ($p<0\cdot0001$) on the basis of the risk score and cutpoint developed in the training set (table 3, figure 2). The median survival for the low-risk group was 18·5 (SD 2·6; 95% CI 13·5–23·6) months *vs* 9·2 (2·4; 4·4–14·0) months for the high-risk group (HR 2·3, 95% CI 1·39–3·70). Notably, this successful validation was achieved in a setting in which investigators were masked to clinical outcomes, with an independent cohort treated at an external institution with a wide range of therapeutic regimens, and with samples obtained at various timepoints during the course of treatment (table 1).

The primary variables used in the clinical model for the training cohort were not available for the validation cohort and as a result, the comparison between the six-gene model and the clinicopathological variables was not repeated for the validation cohort.

A cell fractionation study was subsequently designed to investigate the cellular origins of the genes comprising the signature. In this study whole-blood samples from 15 patients with castration-resistant prostate cancer at the Dana Farber Cancer Institute were collected for purification of peripheral blood mononuclear cells (PBMCs). Subsequently, four different cell types were enriched from the purified PBMC fraction and levels of RNA transcripts from both enriched and depleted B cells, monocytes, natural killer cells, T cells as well as the original PBMC fraction, were quantitatively analysed. Normalised target-gene expression values from PBMC samples were compared with those from enriched (and depleted) cell fractions to determine whether an increase in expression was seen in a specific cellular fraction(s). Expression levels of cell-specific markers were also analysed in parallel for each cellular fraction generated in the enrichment process, to determine the fold-enrichment of specific cell types. A comparison of the averaged relative expression values in enriched cell fractions was done to investigate potential differences in the levels of expression across cell types, which might correlate with differences seen in whole blood.

This analysis revealed that the genes included in the six-gene model were preferentially expressed in three different enriched cell fractions. *C1QA*, *CDKN1A,* and *TIMP1* showed an increased expression in enriched monocytes, *ITGAL* showed an increased expression in enriched natural killer cells, and *ABL2* and *SEMA4D* showed a slight increase in expression in enriched T cells (appendix p 10).

	Regression coefficient	SE	Wald test	Degrees of freedom	p value
Six-gene signature alone*					
Survival time from CPRC diagnosis	0·66	0·136	23·3	1	<0·0001
Survival time from blood draw	0·57	0·120	22·2	1	<0·0001
Survival from CRPC diagnosis (gene signature entered first)†					
Six-gene signature	0·73	0·162	20·5	1	<0·0001
Visceral metastases	–1·28	0·642	4·0	1	0·054
Survival from CRPC diagnosis (clinicopathological variables entered first)†					
Haemoglobin	–0·39	0·213	3·3	1	0·070
Six-gene signature	0·63	0·139	20·4	1	<0·0001
Survival from blood draw (clinicopathological variables entered first)†					
ECOG	0·60	0·318	3·5	1	0·061
Haemoglobin	–0·48	0·228	4·4	1	0·036
Six-gene signature	0·50	0·126	15·9	1	<0·0001

CRPC=castration-resistant prostate cancer. ECOG=Eastern Cooperative Oncology Group. *Results of separate Cox models. †Results of stepwise Cox model.

Table 5: **Effect of combining the six-gene signature and clinicopathological variables**

Discussion

The prognosis of patients with castration-resistant prostate cancer is extremely heterogeneous and the ability to predict prognosis has been limited by models with only moderate predictive discrimination. We postulated that gene-expression profiling of whole blood, which might be influenced by neoplastic tissue through which blood circulates, could provide insights into the behaviour of malignant disease. Indeed, the peripheral-blood six-gene model was capable of predicting survival in patients with castration-resistant prostate cancer.

An understanding of the function of the protein products of the genes comprising the signature could provide mechanistic insight into the powerful prognostic effect of this gene-expression model. The protein products of *ABL2*, *ITGAL*, and *SEMA4D* are components of an integrated system required for T-cell motility, antigen surveillance, and T-helper-cell activity with respect to B-cell activation.[25–29] The down-regulation of these genes in patients with increased mortality is indicative of a down-regulation of cellular and humoral immunity. Similarly, the upregulation of *C1QA*, *CDKN1A*, and *TIMP1* in patients with increased mortality is indicative of a process driving monocyte differentiation towards the production of tissue macrophages and by default, away from dendritic cells.[30–35] Notably, our group has previously shown that elevated plasma levels of TIMP1 are associated with poor survival in patients with castration-resistant prostate cancer.[36] Although we are limited in our understanding of the causal relation between this complex whole-blood transcriptional signature and prognosis in patients with castration-resistant prostate cancer, and the interpretation of the role of individual genes is probably oversimplified, the genes comprising our model suggest a biasing of the immune system towards macrophage differentiation, accompanied by a decrease in both cell-mediated and humoral immunity, in patients with higher mortality.

Importantly, other groups have also shown the potential prognostic importance of circulating immune factors.

Panel: Research in context

Systematic review

In preparation for the current report, articles published between Jan 1, 1980, and Jan 1, 2010, were reviewed for clinicopathological and laboratory parameters associated with survival in patients with castration-resistant prostate cancer by searching PubMed with the terms "prognosis", "castration-resistant prostate cancer", "hormone-refractory prostate cancer", "androgen-independent prostate cancer" and "survival". This synthesis of published works provided a framework for the development of novel approaches to prognostication in patients with castration-resistant prostate cancer.

Interpretation

Our study shows that a whole-blood gene signature predicts survival in patients with castration-resistant prostate cancer better than several routinely measured clinical variables alone. Additional validation, incorporating other known prognostic indicators, including circulating tumour cells, is required for integration of the prognostic signature into clinical care.

The prognostic significance of peripheral-blood immune signatures in patients with castration-resistant prostate cancer was explored by Olmos and colleagues.[37] In their study, a nine-gene signature in a poor-prognosis group of patients with castration-resistant prostate cancer was associated with increased CD71+ early erythroid cells and decreased B-cell and T-cell immune response. Furthermore, increased serum macrophage inhibitory cytokine 1 at baseline and after one cycle of docetaxel has been shown to predict response and survival in patients with castration-resistant prostate cancer.[38]

The six-gene signature is of clinical relevance. First, such a test might assist with patient counselling. Second, the signature could be used to stratify patients enrolled in phase 3 trials. Finally, as the treatment armamentarium for castration-resistant prostate cancer has expanded to include immune modulation with sipuleucel-T, a natural extension of this work is to identify whether such a whole-blood signature could serve as a predictive biomarker.

Our study is limited by the lack of data regarding clinicopathological variables in the training and validation sets that might be of prognostic importance. For instance, data regarding Gleason grade, ECOG performance status, and specific therapeutic interventions were not available for the validation cohort. Additional limitations include the lack of uniform treatment and follow-up among the patients with castration-resistant prostate cancer. However, patients in both cohorts were enrolled in a clinical biorepository protocol that included prospective collection of outcome data. The heterogeneity in treatments received by individual patients mimics that of real world clinical practice, adding to the generalisability of the results. Other clinicopathological variables have been correlated with prognosis in patients with castration-resistant prostate cancer and enumeration of circulating tumour cells (CTCs) using the CellSearch platform (Veridex, Raritan, New Jersey) has been cleared by the US Food and Drug Administration for this purpose.[39] The effect of adding pretreatment CTCs to our model cannot be established. However, our analysis suggests that the six-gene model predicts survival better than several routinely measured clinical variables alone. Moreover, adding the clinical variables to the six-gene model provides little, if any, improvement in prognostic ability.

In summary, our study describes the development and validation of a whole-blood gene signature that correlates with survival of patients with castration-resistant prostate cancer. Because of the lack of data regarding several variables of potential prognostic importance, additional prospective validation of the signature is warranted (panel). Such a validation could be integrated into a large randomised therapeutic trial in castration-resistant prostate cancer, ensuring capture of parameters such as CTCs and concomitant drugs (eg, corticosteroids, bisphosphonates) as well as mandating uniform timing of sample collection (eg, baseline pretreatment) and follow-up. Additional studies are exploring the prognostic significance of the six-gene signature in other prostate cancer clinical states, the stability of the signature during the course of a patient's illness, and the predictive ability of this signature in patients with prostate cancer treated with immune-based therapies.

Contributors

RWR did the literature search and contributed to the study design, data collection, data analysis, data interpretation, and writing of the report. MDG was involved in data collection, data analysis, data interpretation, and writing of the report. HIS was involved with study design, data collection, data analysis, data interpretation, and writing of the report. JM was involved with study design, data analysis, data interpretation, and providing graphs and figures. KW was involved with study design, data analysis, data interpretation, and writing of the report. G-SML was involved with data collection and data analysis. LK was involved with data collection and data analysis. SKS was involved with data interpretation and writing. AA was involved with data collection, data analysis, and data interpretation. MF was involved in data analysis and interpretation. PWK was involved in study design, study accrual, data interpretation, and writing of the report. WKO was involved in study design, data collection, data analysis, data interpretation, writing of the report, and funding.

Conflicts of interest

RWR served as a consultant to Source MDx. JM served as a consultant to Source MDx. KW was an employee of Source MDx. All other authors declare no conflicts of interest.

Acknowledgments

This study was funded by Source MDx. Source MDx is no longer in business.

References

1 Scher HI, Morris MJ, Basch E, Heller G. End points and outcomes in castration-resistant prostate cancer: from clinical trials to clinical practice. *J Clin Oncol* 2011; **29:** 3695–704.

2 Berry WR, Laszlo J, Cox E, Walker A, Paulson D. Prognostic factors in metastatic and hormonally unresponsive carcinoma of the prostate. *Cancer* 1979; **44:** 763–75.

3 Emrich LJ, Priore RL, Murphy GP, Brady MF. Prognostic factors in patients with advanced stage prostate cancer. *Cancer Res* 1985; **45:** 5173–79.

4 Kelly WK, Scher HI, Mazumdar M, Vlamis V, Schwartz M, Fossa SD. Prostate-specific antigen as a measure of disease outcome in metastatic hormone-refractory prostate cancer. *J Clin Oncol* 1993; **11:** 607–15.

5 Halabi S, Small EJ, Kantoff PW, et al. Prognostic model for predicting survival in men with hormone-refractory metastatic prostate cancer. *J Clin Oncol* 2003; **21:** 1232–37.

6 Smaletz O, Scher HI, Small EJ, et al. Nomogram for overall survival of patients with progressive metastatic prostate cancer after castration. *J Clin Oncol* 2002; **20:** 3972–82.

7 Alcorta D, Preston G, Munger W, et al. Microarray studies of gene expression in circulating leukocytes in kidney diseases. *Exp Nephrol* 2002; **10:** 139–49.

8 Bull TM, Coldren CD, Moore M, et al. Gene microarray analysis of peripheral blood cells in pulmonary arterial hypertension. *Am J Respir Crit Care Med* 2004; **170:** 911–19.

9 Ma J, Liew CC. Gene profiling identifies secreted protein transcripts from peripheral blood cells in coronary artery disease. *J Mol Cell Cardiol* 2003; **35:** 993–98.

10 Osman I, Bajorin DF, Sun TT, et al. Novel blood biomarkers of human urinary bladder cancer. *Clin Cancer Res* 2006; **12:** 3374–80.

11 Oh WK, Hayes J, Evan C, et al. Development of an integrated prostate cancer research information system. *Clin Genitourin Cancer* 2006; **5:** 61–66.

12 Rosenberg JE, Ross RW, Wassmann K, et al. Gene expression similarities in whole-blood RNA transcript analysis in geographically disparate prostate cancer populations. ASCO Genitourinary Cancers Symposium 2010; San Francisco, CA, USA; March 5–7, 2010: abstr 181.

13 Ross RW, Bankaitis-Davis D, Siconolfi L, et al. Sensitivity and specificity of a whole-blood RNA transcript-based diagnostic test for the diagnosis of prostate cancer (CaP) compared with prostate-specific antigen (PSA) alone. 44th annual meeting of the American Society of Clinical Oncology; Orlando, FL, USA; May 29–June 2, 2009: abstr 5052.

14 Magidson J, Wassmann K. The role of proxy genes in predictive models: an application to early detection of prostate cancer. http://www.statisticalinnovations.com/technicalsupport/Suppressor.AMSTAT.pdf (accessed June 6, 2012).

15 Hanczar B, Zucker JD, Henegar C, Saitta L. Feature construction from synergic pairs to improve microarray-based classification. *Bioinformatics* 2007; **23:** 2866–72.

16 Magidson J. Qualitative variance, entropy, and correlation ratios for nominal dependent variables. *Soc Sci Res* 1981; **10:** 177–94.

17 Heckman JJ, Singer B. A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica* 1984; **52:** 271–320.

18 Land K, Nagin D, McCall P. Discrete-time hazard regression models with hidden heterogeneity: The semiparametric mixed poisson regression approach. *Sociol Methods Res* 2001; **29:** 342–73.

19 Vermunt JK, ed. A general latent class approach to unobserved heterogeneity analysis of event history data. In: Hagenaars J, McCutcheon A, eds. Applied latent class analysis. Cambridge: Cambridge University Press, 2002: 383–407.

20 McDonald JD, Alessandro R. Mixture modeling of recurrent event times with long-term survivors: analysis of Hutterite birth intervals. *Stat Methods Appl* 2001; **10:** 257–72.

21 Steele F. A discrete-time multilevel mixture model for event history data with long-term survivors, with an application to an analysis of contraceptive sterilization in Bangladesh. *Lifetime Data Anal* 2003; **9:** 155–74.

22 Skrondal A, Rabe-Hesketh S. Generalized latent variable modeling: Multilevel, longitudinal and structural equation models. Boca Raton: Chapman & Hall/CRC Press, 2004.

23 Vermunt JK, Magidson J. Technical Guide for Latent GOLD 4.0: Basic and Advanced. Belmont Massachusetts: Statistical Innovations Inc, 2005.

24 Schemper M, Smith TL. A note on quantifying follow-up in studies of failure time. *Control Clin Trials* 1996; **17:** 343–46.

25 Giblin PA, Lemieux RM. LFA-1 as a key regulator of immune function: approaches toward the development of LFA-1-based therapeutics. *Curr Pharm Des* 2006; **12:** 2771–95.

26 Hogg N, Laschinger M, Giles K, McDowall A. T-cell integrins: more than just sticking points. *J Cell Sci* 2003; **116:** 4695–705.

27 Sierra JR, Corso S, Caione L, et al. Tumor angiogenesis and progression are enhanced by Sema4D produced by tumor-associated macrophages. *J Exp Med* 2008; **205:** 1673–85.

28 Smith A, Stanley P, Jones K, Svensson L, McDowall A, Hogg N. The role of the integrin LFA-1 in T-lymphocyte migration. *Immunol Rev* 2007; **218:** 135–46.

29 Zipfel PA, Zhang W, Quiroz M, Pendergast AM. Requirement for Abl kinases in T cell receptor signaling. *Curr Biol* 2004; **14:** 1222–31.

30 Chirco R, Liu XW, Jung KK, Kim HR. Novel functions of TIMPs in cell signaling. *Cancer Metastasis Rev* 2006; **25:** 99–113.

31 Hornebeck W, Lambert E, Petitfrère E, Bernard P. Beneficial and detrimental influences of tissue inhibitor of metalloproteinase-1 (TIMP-1) in tumor progression. *Biochimie* 2005; **87:** 377–83.

32 Lu J, Wu X, Teh BK. The regulatory roles of C1q. *Immunobiology* 2007; **212:** 245–52.

33 Lu JH, Teh BK, Wang L, et al. The classical and regulatory functions of C1q in immunity and autoimmunity. *Cell Mol Immunol* 2008; **5:** 9–21.

34 Ouellet S, Vigneault F, Lessard M, Leclerc S, Drouin R, Guérin SL. Transcriptional regulation of the cyclin-dependent kinase inhibitor 1A (p21) gene by NFI in proliferating human cells. *Nucleic Acids Res* 2006; **34:** 6472–87.

35 Terheyden P, Loos M, Störkel S, Kaul M. Human macrophages simultaneously express membrane-C1q and Fc-receptors for IgG. *Immunol Lett* 2005; **101:** 202–09.

36 Oh WK, Vargas R, Jacobus S, et al. Elevated plasma tissue inhibitor of metalloproteinase-1 levels predict decreased survival in castration-resistant prostate cancer patients. *Cancer* 2011; **117:** 517–25.

37 Olmos D, Clark J, Brewer D, et al. An evaluation of blood mRNA expression array signatures derived from unsupervised analyses in the identification of prostate cancers with poor outcome. *Proc Am Soc Clin Oncol* 2011; **29** (suppl)**:** 4577.

38 Zhao L, Lee BY, Brown DA, et al. Identification of candidate biomarkers of therapeutic response to docetaxel by proteomic profiling. *Cancer Res* 2009; **69:** 7696–703.

39 de Bono JS, Scher HI, Montgomery RB, et al. Circulating tumor cells predict survival benefit from treatment in metastatic castration-resistant prostate cancer. *Clin Cancer Res* 2008; **14:** 6302–09.