Prognostic value of blood mRNA expression signatures in castration-resistant prostate cancer: a prospective, two-stage study

David Olmos, Daniel Brewer*, Jeremy Clark*, Daniel C Danila, Chris Parker, Gerhardt Attard, Martin Fleisher, Alison H M Reid, Elena Castro, Shahneen K Sandhu, Lorraine Barwell, Nikhil Babu Oommen, Suzanne Carreira, Charles G Drake, Robert Jones, Colin S Cooper, Howard I Scher, Johann S de Bono*

Lancet Oncol 2012; 13: 1114–24

Published Online October 9, 2012 http://dx.doi.org/10.1016/S1470-2045(12)70372-8

See Comment page 1067

*DO, DB, and JC contributed equally

Drug Development Unit, The Royal Marsden NHS Foundation Trust, Sutton, UK (D Olmos MD, G Attard MD, A H M Reid MD, S K Sandhu MD, N B Oommen MD, Prof J S de Bono MD)**; Section of Medicine** (D Olmos, G Attard, A H M Reid, S K Sandhu, S Carreira PhD, J S de Bono)**, Section of Molecular Carcinogenesis** (D Brewer PhD, J Clark PhD, Prof C S Cooper PhD)**, Cancer Genetics Department** (E Castro MD)**, The Institute of Cancer Research, Sutton, UK; Clinical Research Programme, Spanish National Cancer Research Centre, Madrid, Spain** (D Olmos)**; Genitourinary Oncology Service, Department of Medicine, Sidney Kimmel Center for Prostate and Urologic Cancers** (D C Danila MD, Prof H I Scher MD)**, Department of Laboratory Medicine** (M Fleisher MD)**, Memorial Sloan-Kettering Cancer Center, New York, NY, USA; Department of Medicine, Weill Cornell College of Medicine, New York, NY, USA** (D C Danila, H I Scher)**; Academic Urology Unit, The Royal Marsden NHS Foundation Trust, Sutton and London, UK** (C Parker MD, E Castro)**; The Beatson West of Scotland Cancer Centre, Glasgow, UK** (L Barwell MS)**; James Buchanan Brady Urological Institute, Sidney Kimmel Comprehensive Cancer Centre, Johns Hopkins University, Baltimore, MD, USA** (C G Drake MD)**; and**

Summary

Background Biomarkers are urgently needed to dissect the heterogeneity of prostate cancer between patients to improve treatment and accelerate drug development. We analysed blood mRNA expression arrays to identify patients with metastatic castration-resistant prostate cancer with poorer outcome.

Methods Whole blood was collected into PAXgene tubes from patients with castration-resistant prostate cancer and patients with prostate cancer selected for active surveillance. In stage I (derivation set), patients with castration-resistant prostate cancer were used as cases and patients under active surveillance were used as controls. These patients were recruited from The Royal Marsden Hospital NHS Foundation Trust (Sutton, UK) and The Beatson West of Scotland Cancer Centre (Glasgow, UK). In stage II (validation-set), patients with castration-resistant prostate cancer recruited from the Memorial Sloan-Kettering Cancer Center (New York, USA) were assessed. Whole-blood RNA was hybridised to Affymetrix U133plus2 microarrays. Expression profiles were analysed with Bayesian latent process decomposition (LPD) to identify RNA expression profiles associated with castration-resistant prostate cancer subgroups; these profiles were then confirmed by quantative reverse transcriptase (qRT) PCR studies and correlated with overall survival in both the test-set and validation-set.

Findings LPD analyses of the mRNA expression data divided the evaluable patients in stage I (n=94) into four groups. All patients in LPD1 (14 of 14) and most in LPD2 (17 of 18) had castration-resistant prostate cancer. Patients with castration-resistant prostate cancer and those under active surveillance comprised LPD3 (15 of 31 castration-resistant prostate cancer) and LDP4 (12 of 21 castration-resistant prostate cancer). Patients with castration-resistant prostate cancer in the LPD1 subgroup had features associated with worse prognosis and poorer overall survival than patients with castration-resistant prostate cancer in other LPD subgroups (LPD1 overall survival 10·7 months [95% CI 4·1–17·2] *vs* non-LPD1 25·6 months [18·0–33·4]; p<0·0001). A nine-gene signature verified by qRT-PCR classified patients into this LPD1 subgroup with a very low percentage of misclassification (1·2%). The ten patients who were initially unclassifiable by the LPD analyses were subclassified by this signature. We confirmed the prognostic utility of this nine-gene signature in the validation castration-resistant prostate cancer cohort, where LPD1 membership was also associated with worse overall survival (LPD1 9·2 months [95% CI 2·1–16·4] *vs* non-LPD1 21·6 months [7·5–35·6]; p=0·001), and remained an independent prognostic factor in multivariable analyses for both cohorts.

Interpretation Our results suggest that whole-blood gene profiling could identify gene-expression signatures that stratify patients with castration-resistant prostate cancer into distinct prognostic groups.

Funding AstraZeneca, Experimental Cancer Medicine Centre, Prostate Cancer Charity, Prostate Cancer Foundation.

Introduction

Prostate cancer is a very heterogeneous disease; many patients are diagnosed at an early stage and do not need treatment or are cured with radical treatment.[1] Other patients present with advanced disease or recurrent disease despite initial curative treatment, and eventually succumb due to metastatic castration-resistant prostate cancer.[2] The molecular heterogeneity of castration-resistant prostate cancer, as well as difficulty in acquiring tumour tissue from patients with prostate cancer, makes the identification and validation of multipurpose blood-based or urine-based biomarker assays crucially important to individualise management of prostate cancer.[3] Such tests are repeatable, less invasive, and easily implemented in clinical practice.[3–5] Serum prostate-specific antigen (PSA) has been widely studied in the context of management of prostate cancer[6] but is not a reliable intermediate endpoint of overall survival.[6,7] In recent years the development of high-throughput technologies has allowed the identification of other useful tissue-based and fluid-based biomarkers.[6,7] For example, the presence of circulating tumour cells (CTCs) in peripheral blood is a prognostic biomarker and a measure of therapeutic response in patients with prostate cancer.[8,9]

Tumour gene-expression signatures have contributed to molecular classifications of cancer but as potential

biomarkers have not been widely implemented in clinical practice, partly due to challenges with acquiring fresh tumour samples.[10,11] Blood cells express 16 000–20 000 gene transcripts, which respond to microenvironmental and macroenvironmental changes (eg, prostate-cancer bone-marrow invasion, which affects haemopoiesis by competing with primitive haemopoietic-cell niches).[12,13] Moreover, chromosomal aberrations and epigenetic changes have been identified in peripheral blood mononuclear cells from patients with solid tumours including prostate cancer, and are associated with disease burden and prognosis.[14–17]

In this study, we postulated that patients with aggressive metastatic castration-resistant prostate cancer would have a different whole-blood gene-expression array profile to patients with low-grade, low-disease-burden, prostate cancer suitable for active surveillance. We also postulated that this expression profile would have clinical use. We therefore initially compared whole-blood gene-expression patterns from two separate groups of patients with prostate cancer (for the stage I derivation test cohort): patients with progressing metastatic castration-resistant prostate cancer undergoing a change in treatment (cases); and patients with clinically indolent disease eligible and consenting to participate in an active surveillance research trial (controls). Whole-blood RNA was acquired with the PAXgene (PreAnalytiX, Hombrechtikon, Switzerland) system method, which prevents RNA degradation and restricts ex-vivo transcriptional changes in anticoagulated whole blood,[18] providing high-quality RNA for expression profiling. We initially studied the clinical relevance of these expression signatures. Finally, we confirmed the broad relevance of the identified signatures in an independent patient population with castration-resistant prostate cancer treated at Memorial Sloan-Kettering Cancer Center (MSKCC), to confirm the validity of the identified expression profile (stage II; validation cohort).

Methods

Study design and patients

The primary aim of the first stage of this study was to elucidate whole-blood expression profiles associated with aggressive castration-resistant prostate cancer; we used whole blood from patients participating in an active surveillance protocol as a control. Secondary aims included the correlation of these expression profiles with outcome. The results of these analyses, done on whole-blood RNA samples acquired from patients treated at The Royal Marsden Hospital NHS Foundation Trust (Sutton, UK) and The Beatson West of Scotland Cancer Centre (Glasgow, UK; stage I) were then validated in the second stage of this study (stage II).

Patients included in stage I attended routine clinical appointments at these centres between August, 2007, and April, 2008. Patients were enrolled in two groups: (1) patients with advanced castration-resistant prostate cancer; and (2) patients undergoing active surveillance in a prospective research trial. The analyses in stage II assessed whole-blood RNA from an independent cohort of patients with progressing metastatic castration-resistant prostate cancer treated at MSKCC (New York, USA) between February, 2007, and July, 2011. Patients at both The Royal Marsden and MSKCC provided written informed consent for CTC enumeration. This study was approved by the relevant local institutional review boards and this report was written according to the REMARK (Reporting recommendations for tumor *MARK*er) prognostic studies criteria.[19]

All patients had a histological diagnosis of prostate cancer and provided informed and written consent for these studies, before sample collection. For the first part of this study there was no formal estimation of sample size; for the second stage, on the basis of the Fisher-Irwin test,[20] we estimated that 66 patients were needed to validate our unfavourable prognostic signature (expressed in at least 25% of patients with castration-resistant prostate cancer), which is associated with a minimum five-fold increase in the risk of 18-month mortality (assuming an 18-month mortality rate in patients with castration-resistant prostate cancer of 50%, an α error of 5%, and a β error of 20%).

Procedures

For each patient, 2·5 mL of peripheral venous blood was collected in 5 mL PAXgene tubes. All samples were taken at least 1 month after cessation of any prostate-cancer therapy. In stage I, additional blood was collected 1 month after the first sampling in patients who had not yet been started on a new prostate-cancer treatment. Samples were frozen at –20°C within an hour of being drawn. Whole-blood RNA was isolated and purified with the PAXgene Blood RNA Kit according to the manufacturer's instructions. RNA quality and quantity measures were done with a 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA) and an ND-1000 spectrophotometer (Thermo Scientific, Newark, DE, USA), respectively. Only purified RNA samples with an RNA integration number of seven or more were selected for further analyses. In a proportion of patients with castration-resistant prostate cancer, CTC enumeration was also done with the CellSearch system (Veridex, Raritan, NJ, USA), as described previously.[8]

50 ng of purified RNA was amplified, fragmented, and labelled with the Ovation Whole Blood Solution (NuGen Technologies, San Carlos, CA, USA). Expression profiles were assessed with Human Genome U133 Plus 2.0 arrays (Affymetrix, Santa Clara, CA, USA). Briefly, arrays were hybridised with 4 µg of cDNA, washed, and scanned. Then, expression signal estimates were derived from raw data files with a robust multiarray analysis algorithm. Complete microarray datasets, in compliance with MIAME (minimum

Institute of Cancer Sciences, University of Glasgow, Glasgow, UK (R Jones MD)

Correspondence to:
Prof Johann S de Bono, Section of Medicine, The Institute of Cancer Research, Drug Development Unit, Royal Marsden NHS Foundation Trust, Downs Road, Sutton, Surrey SM2 5PT, UK
johann.de-bono@icr.ac.uk

For more on the **prospective trial** see http://prostate.icr.ac.uk/ActiveSurveillance.htm

information about a microarray experiment) guidelines[21] are made available on Gene Expression Omnibus (GEO; accession GSE37199). Latent process decomposition (LPD), an unsupervised Bayesian approach, was used to classify samples into subgroup processes. The optimal number of LPD processes in our dataset was derived and each sample was then said to be a member of a particular process if its probability of belonging was greater than 0·5.[22] Hierarchical clustering analyses and principal component analysis (PCA) were also undertaken (appendix).

See Online for appendix

The random forest machine learning algorithm was used to develop signatures suitable as tests for an LPD process membership.[23] Performance was determined by ten-fold cross-validation. The Ingenuity pathway analysis application (Ingenuity, Redwood City, CA, USA) was used for functional network analyses of lists of genes expressed differentially between the LPD processes (appendix).

First strand cDNA was synthesised using 500 ng of each RNA, random nonamers, and an upstream gene-specific primer for ERG (GGTAGTGGAGATGTGAGAGAAGG), as described by Clipsham and collegues.[24] Subsequently, the expression of significantly upregulated genes (appendix) in the Affymetrix microarrays was verified by quantitative reverse transcriptase (qRT) PCR using specific probes in a 384-well custom, TaqMan, low-density array with the Prism 7900HT instrument (Applied Biosystems, Foster City, CA, USA). Each probe was studied in duplicate and then normalised relative to the 18S rRNA endogenous control (ΔCt=endogenous Ct–experimental Ct). This ratio of experimental to endogenous signal was then compared with the transcript level ratio for the same transcripts from the Affymetrix results. TMPRSS2/ERG transcripts were studied by nested-PCR with the method and primers described by Clark and colleagues[25] (appendix). Amplified products were visualised by agarose-gel electrophoresis.

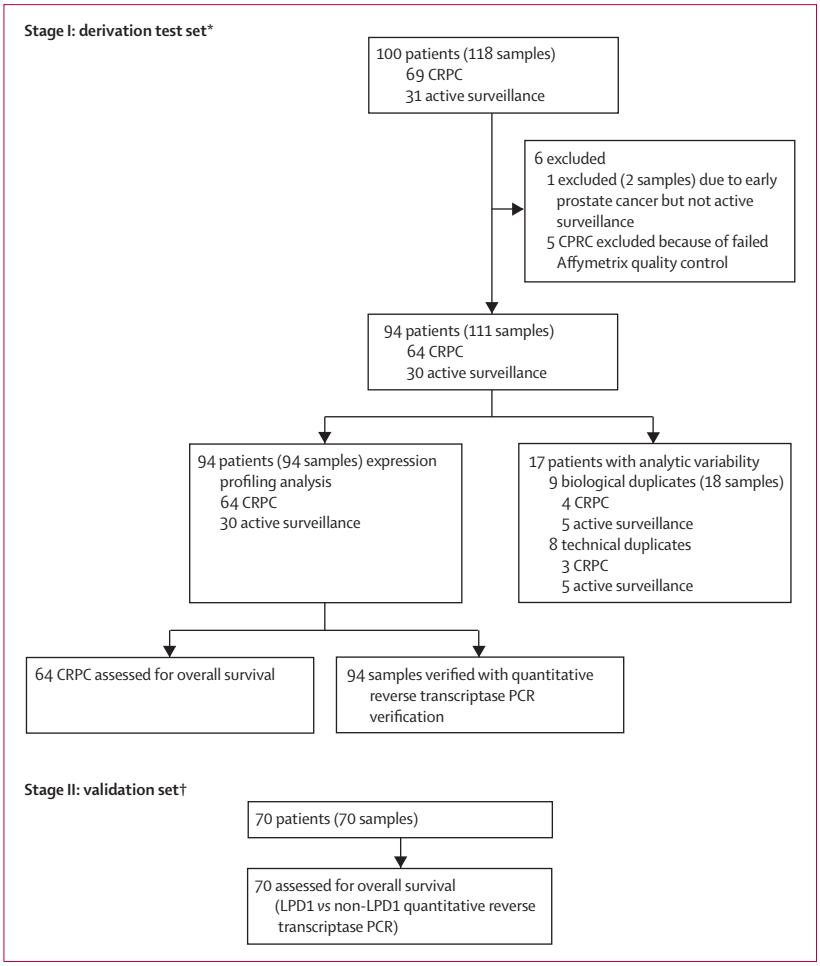


Figure 1: **Study profile**
CRPC=castration-resistant prostate cancer. LPD=latent process decomposition. *Identification of a signature associated to aggressive CRPC and poor outcome. †Validation of the prognostic value of the expression signature generated in stage I.

Statistical analyses

The associations of LPD expression-pattern groups with baseline clinical characteristics were analysed with the χ^2 test or Fisher's exact test, Spearman correlation rank test, or Mann-Whitney U test for categorical, ordinal, or continuous variables, respectively. Overall survival was calculated from time of blood draw to death (events) or until last follow-up (censored event). Date cutoffs for these analyses were Sept 30, 2010, for stage I and June 6, 2012, for stage II. Univariable and multivariable analyses were done by proportional hazard (Cox) regression analysis. Overall survival curves were derived with the Kaplan-Meier method and compared using the log-rank test. Model selection was done on multivariable Cox-regression models with Akaike-information-criteria-based bidirectional selection to drop non-contributing factors. Unless otherwise specified all analyses were done with R software (version 2.11.0) or SPSS (version 19.0).

Role of the founding source

AstraZeneca provided support for data collection and processing of samples. The sponsor of the study had no role in study design, data analysis, data interpretation, or writing of the report. DO, DB, JC, and JSdB had access to the raw data. DO and JSdB had full access to all of the data and had final responsibility for the decision to submit for publication.

Results

A total of 118 blood samples were obtained from 100 patients with prostate cancer during stage I (figure 1). After quality control, only samples from 94 patients were eligible for data expression studies. For the validation series in stage II, we obtained 70 samples from patients with castration-resistant prostate cancer treated at MSKCC. In the active surveillance group, the median age was 69 years (range 50–83), the median presenting PSA was 7·2 ng/mL (range 3–20) with 93% (29 of 30) of patients having a T1 tumour, or a Gleason score of 6 or

less, or both. The clinical characteristics of the patients with CPRC in the derivation and validation sets are shown in table 1.

Two main sources of intrapatient variability, technical and biological, were assessed. Thus, eight technical replicates (two samples taken from the same patient at the same time) and nine biological replicates (two samples taken at different timepoints from the same patients) were analysed. For the technical replicates the mean Pearson correlation between pairs was 0·99, with fold changes greater than 1·5 and 2·0 in 3·2% and 0·2% of the 54613 probesets analysed, respectively, on average across samples. For biological replicates the mean Pearson correlation between pairs was 0·98 with fold changes greater than 1·5 and 2·0 in 3·1% and 0·2%, respectively. Both sources of intrapatient variability in this study were less than the observed interpatient variability.

	Stage I *vs* stage II patients with CRPC			Stage I patients with CRPC (derivation set, n=64)			Stage II patients with CRPC (validation set, n=70)		
	Stage I (n=64)	Stage II (n=70)	p value	LPD1 (n=16)	non-LPD1 (n=48)	p value	LPD1 (n=31)	non-LPD1 (n=39)	p value
Age, years									
Median (range)	68·5 (46·0–83·1)	71·7 (51·6–86·9)	0·139	67·3 (46·0–79·8)	69·0 (53·1–83·1)	0·377	72·9 (51·9–82·3)	70·5 (51·6–86·9)	0·745
Performance status									
ECOG 0	19 (29·7%)	21 (30·0%)	0·855	5 (31·3%)	14 (29·2%)	0·214	7 (22·6%)	14 (35·9%)	0·482
ECOG 1	42 (65·6%)	47 (67·1%)	..	9 (56·3%)	33 (68·8%)	..	23 (74·6%)	24 (61·5%)	..
ECOG ≥2	3 (4·7%)	2 (2·9%)	..	2 (12·5%)	1 (2·1%)	..	1 (3·2%)	1 (2·6%)	..
Gleason score									
Gleason ≤6	10 (17·5%)	12 (17·1%)	0·647	1 (6·7%)	9 (21·4%)	0·126	7 (22·6%)	5 (12·8%)	0·25
Gleason 7	13 (22·8%)	21 (30·0%)	..	6 (40·0%)	7 (16·7%)	..	13 (41·9%)	24 (61·5%)	..
Gleason ≥7	34 (59·6%)	37 (52·9%)	..	8 (53·3%)	26 (61·9%)	..	11 (35·5%)	10 (25·6%)	..
Metastasis									
Bone	55 (88·7%)	68 (97·1%)	0·083	15 (93·7%)	40 (83·3%)	0·43	31 (100%)	37 (94·9%)	0·499
Visceral	9 (14·1%)	11 (15·5%)	0·815	3 (18·8%)	6 (12·5%)	0·679	5 (16·1%)	6 (15·4%)	0·932
Baseline PSA, ng/mL									
Median (range)	178 (1–3683)	70 (3–5219)	0·806	373 (20–3683)	89 (1–3609)	0·005	259 (11–5219)	61 (3–2413)	0·005
PSADT									
PSADT <3 months	37 (57·8%)	..*	NA	11 (68·8%)	26 (54·2%)	0·212	..*	..*	..*
PSADT ≥3 and <6 months	17 (26·6%)	..*	NA	4 (25·0%)	13 (27·1%)	..	..*	..*	..*
PSADT ≥6 months	10 (15·6%)	..*	NA	1 (6·3%)	9 (18·8%)	..	..*	..*	..*
Haemoglobin, g/L									
Median (range)	1·1 (0·9–1·4)	1·2 (0·9–1·5)	0·013	1·1 (0·9–1·4)	1·1 (0·9–1·4)	0·555	1·1 (0·9–1·5)	1·3 (0·9–1·5)	<0·0009
Alkaline phosphatase, IU/L									
Median (range)	139 (46–2684)	138 (48–2051)	0·631	390 (57–2684)	100 (46–2166)	0·002	170 (48–2051)	103 (52–1057)	0·026
Serum albumin (g/L)									
Median (range)	36 (17–45)	42 (37–48)	<0·0001	32 (27–41)	37 (17–45)	0·004	41 (37–47)	43 (36–48)	0·011
Lactate dehydrogenase, IU/L									
Median (range)	177 (122–749)	254 (132–1014)	<0·0001	278 (144–749)	168 (122–645)	0·009	286 (132–1014)	243 (144–393)	0·009
CTCs									
<5 CTCs/7·5 mL of blood	13 (41·9%)	29 (41·4%)	0·962	3 (30%)	10 (47·6%)	0·353	7 (22·6%)	22 (56·4%)	0·004
≥5 CTCs/7·5 mL of blood	18 (58·1%)	41 (58·6%)	..	7 (70%)	11 (52·4%)	..	24 (77·4%)	17 (43·6%)	..
Unknown	33	..		6	27	..	..	..	..
***TMPRSS2/ERG* translocation**									
Nested PCR positive	11 (17·2%)	..*	NA	5 (31·3%)	6 (12·5%)	0·085	..*	..*	..*
Previous chemotherapy history									
Number of lines (median [range])	0 (0–2)	1 (0–2)	<0·0001	0 (0–2)	0 (0–2)	0·077	1 (0–2)	1 (0–2)	0·592
Previous docetaxel	19 (29·7%)	47 (67·1%)	<0·0001	7 (43·8%)	12 (25·0%)	0·155	22 (71·0%)	25 (64·1%)	0·544
Concomitant steroids									
Yes	22 (34·4%)	36 (51·4%)	0·047	5 (31·3%)	17 (35·4%)	0·761	18 (58·1%)	18 (46·2%)	0·322

CRPC=castration-resistant prostate cancer. LPD=latent process decomposition. ECOG=Eastern Cooperative Oncology Group. PSA=prostate-specific antigen. PSADT=PSA doubling time. NA=not applicable. CTC=circulating tumour cells. *No information available.

Table 1: **Baseline characteristics of patient with CRPC and associations with the LPD1 signature according to qRT-PCR results**

Initial gene expression analyses using PCA showed that plate and centre made a contribution to the overall gene expression profile patients. Therefore, in all differential analyses these effects were taken into account. The expression data showed that 3009 probesets were differentially expressed between castration-resistant prostate cancer and active surveillance patient samples (appendix). The top ten genes most significantly upregulated in patients with castration-resistant prostate cancer were *HEMGN*, *CA1*, *IFI27*, *RAP1GAP*, *PBX1*, *ERAF*, *MMP8*, *CRISP3*, *ABCC13*, and *BPGM*. Multiple probesets for an individual gene were

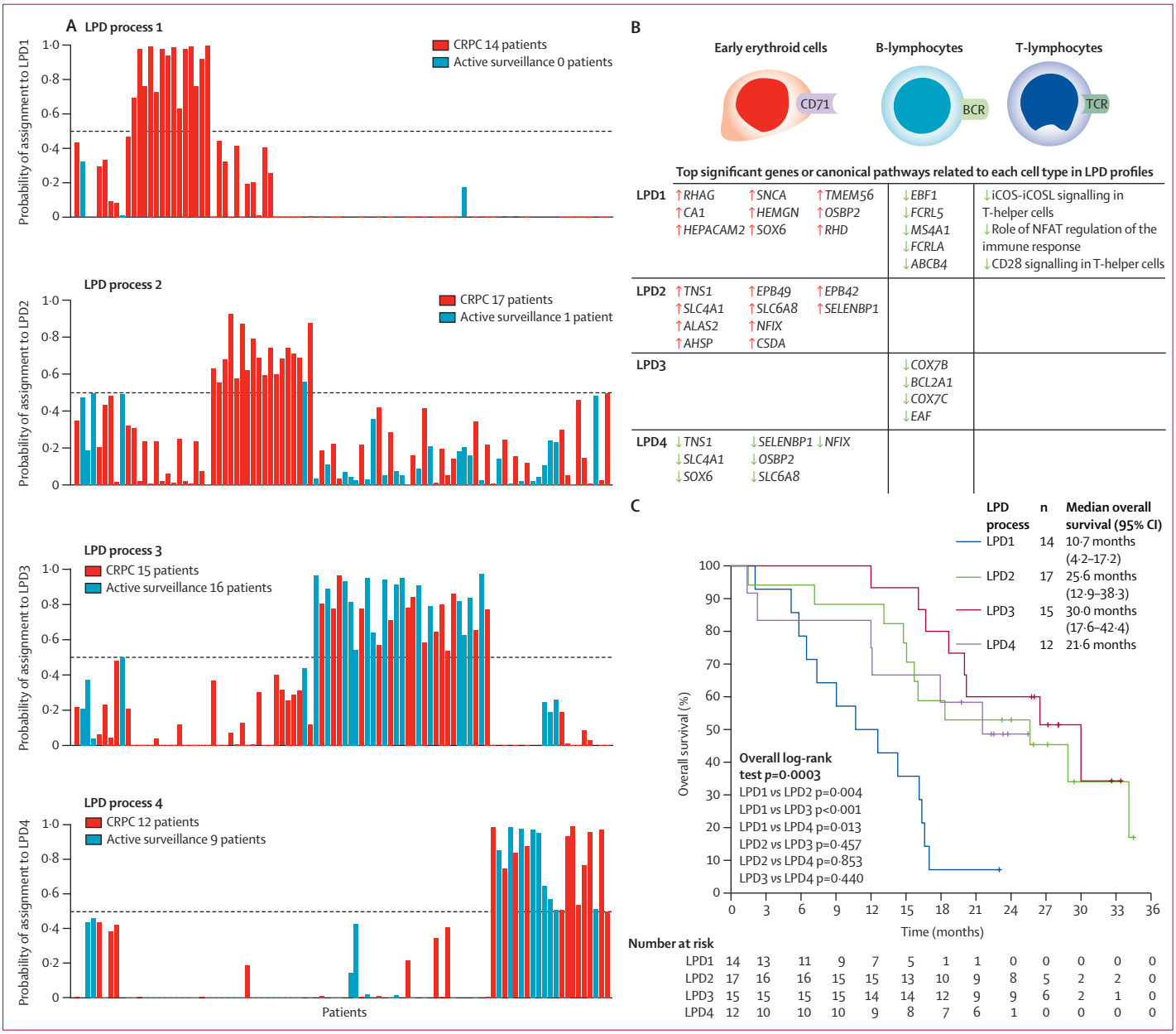


Figure 2: **Latent process decomposition analyses (patients grouping, associated genes, and survival)**
(A) LPD division of patients samples (n=94). LPD analysis of Affymetrix blood expression data divided the samples into four processes (LPD1–4), each represented here by a bar chart, red for CRPC and blue for active surveillance. Each sample is represented in all four processes and the height of each bar corresponds to how well a sample matches to a LPD process. A sample is assigned to an LPD group if its probability of assignment score is greater than 0·5 (discontinuous horizontal lines). (B) Many of the top ten significant genes and pathways differentially expressed when each LPD was compared with each of the other LPD process were associated with CD71+ early erythroid cells, B-cell, and T-cell lineages. The most significantly upregulated or downregulated genes in relation to cell type and LPD process are shown. Red arrows represent upregulated genes and pathways, green arrows downregulated genes and pathways. (C) Kaplan-Meier overall survival curves in months according to LPD group for 58 patients with CRPC only. No significant differences in overall survival were seen between the LPD2, LPD3, and LPD4 subgroups. The six patients not included were not subclassified by LPD analysis and had a median overall survival of 15·8 (95% CI 8·8–22·7) months. LPD=latent process decomposition. CPRC=castration-resistant prostate cancer. BCR=B-cell receptor. TCR=T-cell receptor. iCOS-iCOSL=inducible costimulator and its ligand. NFAT=nuclear factor of activated T cells.

frequently ranked in close proximities, confirming the quality of this dataset.

Unsupervised LPD analyses optimally divided the expression dataset into four processes (LPD1–4; figure 2), independent of treatment centre (p=0·74). 14 patients were in LPD1, 18 in LPD2, 31 in LPD3, and 21 in LPD4. Ten patients did not reach the threshold for inclusion into any LPD process. Differential expression analyses were done between all processes; for this purpose expression of each probe in an LPD group was compared with the expression in the remaining groups. Thus, 541 probesets were differentially expressed in LPD2, 2179 in LPD3, and 10 063 in LPD4 (appendix). In LPD1, 2740 probesets were differentially expressed (1584 overexpressed, 1156 underexpressed). The significant probesets that showed the greatest difference in LPD1 are shown in figure 3. Of the 25 top upregulated transcripts in LPD1, nine were associated with genes highly expressed in CD71+ early erythroid cells but not in whole blood (*RHAG*, *CA1*, *HEPACAM2*, *SNCA*, *HEMGN*, *SOX6*, *TMEM56*, *OSBP2*, *RHD*), and nine of the 20 most downregulated transcripts were associated with B-cell related genes (*EBF1*, *FCRL5*, *MS4A1*, *FCRLA*, *ABCB4*).

A

Ten most upregulated genes		Ten most down-regulated genes	
Gene symbol	Log_2 change	Gene symbol	Log_2 change
GYPA	3·88	*EBF1*	−1·87
RHAG	3·85	*MACROD2*	−1·72
IFI27	3·75	*OSBPL10*	−1·72
MMP8	3·70	*PAX5*	−1·67
CRISP2	3·55	*LOC283663*	−1·65
MAOA	3·52	*FCRL5*	−1·61
CRISP3	3·46	*MS4A1*	−1·61
RAP1GAP	3·35	*IGHD*	−1·60
RHD	3·35	*FCRLA*	−1·58
MMP8	3·33	*PRSS33*	−1·57

B

Ingenuity canonical pathways	p value
iCOS-iCOSL signalling in T-helper cells	0·00003
Calcium-induced T-lymphocyte apoptosis	0·00003
Role of NFAT in regulation of the immune response	0·00012
CD28 signalling in T-helper cells	0·00012
Methane metabolism	0·00014

C

Probeset	Symbol	Gini importance metric
201174_s_at	TERF2IP	5·3984552
213096_at	TMCC2	5·1047048
204467_s_at	SNCA	3·5190276
209046_s_at	GABARAPL2	2·5084932
207827_x_at	SNCA	1·8047234
202129_s_at	RIOK3	1·3102857
212330_at	TFDP1	1·0905238
203040_s_at	HMBS	0·9611742
1552713_a_at	SLC4A1	0·8626886
201060_x_at	STOM	0·4762092

Figure 3: **Top differentially expressed genes, canonical pathways, and probesets in LPD1 signature**

(A) Significant genes (adjusted p<0·05) associated with LPD group 1 ordered by mean magnitude change. (B) Five most canonical pathways associated with LPD group 1. (C) Ten-probeset robust signature of LPD group 1 membership. The Gini importance metric is a general indicator of feature relevance. iCOS-iCOSL=inducible costimulator and its ligand. NFAT=nuclear factor of activated T cells.

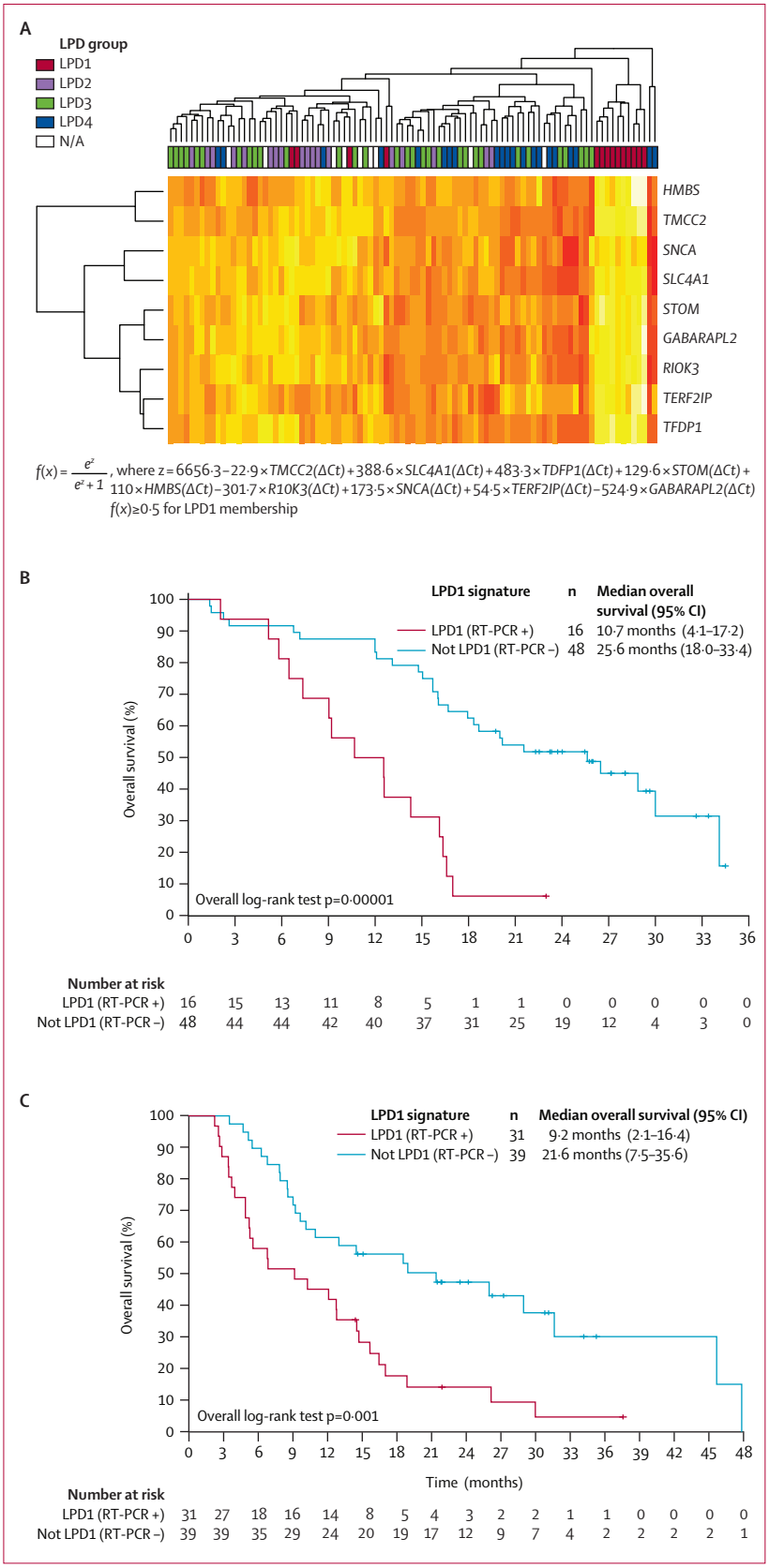
Four of the five top canonical pathways identified by Ingenuity software to be most significantly associated with genes differentially expressed in LPD1 were associated with T-cell function (figure 3; appendix). We also examined the expression of three genes associated with prostate cancer: *ERG*, *AR*, and *KLK3* (appendix); these genes were significantly upregulated in LPD1 (p=0·001, p<0·0001, and p=0·029, respectively), but their expression was not correlated with the number of CTCs (p=0·7, p=0·4, and p=0·1, respectively), despite CTC counts being higher in LPD1 patients (median 133, range 0–633). *ERG* expression was not associated with the expression of *TMPRSS2-ERG* transcripts (15 of 94 patients), although fusion-gene transcripts were more frequently detected in blood samples of LPD1 patients (p<0·0001).

Two other classic clustering algorithms, PCA and hierarchical clustering, were used to examine the sample groupings in the expression data. PCA divided the

	LPD1 (n=14)	Non-LPD1 (n=44)	p value
Performance status			
ECOG 0	4 (28·6%)	13 (29·5%)	0·944
ECOG 1	10 (71·4%)	31 (70·5%)	..
Age, years			
Median (range)	67·3 (46·0–79·8)	68·6 (53·1–83·1)	0·501
Gleason score			
Score ≤7	5 (38·5%)	14 (36·8%)	0·917
Score >7	8 (61·5%)	24 (63·2%)	..
Unknown	1	6	..
PSA, ng/mL			
Median (range)	373 (20–3683)	85 (1–3609)	0·002
PSADT, months			
Median (range)	2·3 (0·8–14·4)	2·9 (0·5–16·0)	0·148
Albumin, g/L			
Median (range)	3·2 (2·7–4·1)	3·7 (2·9–4·5)	0·002
Alkaline phosphatase, IU/L			
Median (range)	519 (75–2684)	88 (46–2166)	<0·0001
Lactate dehydrogenase value over UNL			
Median (range)	1·2 (0·8–3·9)	0·9 (0·7–3·4)	0·031
CTCs per 7·5 mL			
Median (range)	133 (0–643)	4 (0–168)	0·033
Treatments			
LHRHa	14 (100%)	44 (100%)	1·000
Antiandrogens	14 (100%)	30 (100%)	1·000
Steroids	4 (28·6%)	14 (31·8%)	0·818
Docetaxel	6 (42·9%)	10 (22·7%)	0·177
Number of systemic treatment lines			
Median (range)	3 (1–8)	2 (1–7)	0·076

Six patients with CRPC were not classified by LPD analyses. ECOG=Eastern Cooperative Oncology Group. PSA=prostate-specific antigen. PSADT=PSA doubling time. UNL=upper normal limit. CTC=circulating tumour cell. LHRHa=luteinising hormone releasing hormone analogue.

Table 2: **Baseline characteristics in the CRPC cohort according to results of whole genome arrays**



samples into three groups. Two of the PCA groups largely corresponded with LPD3 and LPD4, while LPD1 and LPD2 were intermingled (appendix). Hierarchical clustering analysis also divided the samples into three main groups (appendix). Group A (n=17) contained 11 samples from LPD1, verifying the distinctive nature of LPD1. Most samples in group B and group C were from LPD3 and LPD4, respectively. LPD2 was not well classified by hierarchical clustering, with most of these samples allocated to group B.

The 64 patients with castration-resistant prostate cancer were evenly distributed across all LPD processes (LPD1, 14; LPD2, 17; LPD3, 15; LPD4, 12; six unclassified; figure 2). This even distribution contrasted with that for patients under active surveillance who were almost entirely in LPD3 and LPD4. When only patients with castration-resistant prostate cancer were analysed, LPD1 patients had higher PSA, alkaline phosphatase, lactate dehydrogenase, and CTC counts, but lower levels of albumin (table 2). No significant differences were reported with regard to previous treatments. After a median follow-up of 29·7 months (range 1·4–36·4), patients with castration-resistant prostate cancer in LPD1 had a significantly poorer overall survival than did patients with castration-resistant prostate cancer in LPD2, LPD3, and LPD4 (p=0·0003, figure 2). As a whole, patients with castration-resistant prostate cancer in LPD1 had a hazard ratio for overall survival of 4·4 (95% CI 2·1–9·2) compared with patients not in LPD1. Survival analysis was repeated with hierarchical clustering groups: group A had a much poorer prognosis than groups B or C (hazard ratio 3·25, 95% CI 1·63–6·49; p=0·0008; appendix). The ability of hierarchical clustering to separate out the poor prognosis patients was, however, inferior to LPD.

Patients in LPD1 had a significantly poorer prognosis than did other patients with castration-resistant prostate cancer. Thus, the random forest algorithm was applied to produce an accurate robust classifier to identify LPD1 membership. This method produced a ranked list of ten probesets representing nine different genes that optimally define LPD1 (figure 3). This nine-gene classifier

Figure 4: **qRT-PCR verification, derivation, and validation of the prognostic value of the LPD1 signature in patients with CRPC**
(A) Heatmap of the nine-signature genes (annotated on the right) in the 94-patient samples assessed with TaqMan probes. The dendrogram on the top shows the arrangement of patient clusters with Euclidean distance as the distance metric. The vertical dendrogram on the left shows the strength of gene clusters within the signature. A formula based on the expression levels of the nine genes in the signature was derived using logistic regression to predict LPD1 membership. (B) Kaplan-Meier overall survival curve (months) according to LPD1 membership in 64 patients with CRPC included in the derivation test set. LPD1 membership was predicted by logistic regression with the nine-signature gene expression determined by qRT-PCR (see formula); patients with a probability of 0·5 or more were assigned to LPD1. (C) Kaplan-Meier overall survival curve (months) according to LPD1 membership in 70 patients with CRPC included in the validation set. CRPC=castration-resistant prostate cancer.

had a cross-validation sensitivity of 93·0%, specificity of 100%, and misclassification rate of 1·2% for LPD1 membership. We then verified the expression by qRT-PCR of the five most upregulated genes in LPD1 as well as the separate nine-signature genes (figure 4; appendix). Expression levels for the selected genes measured by the two methods were highly correlated (r^2 0·7–0·9) giving further validity to these data. We used logistic regression analyses to build a predictive model of LPD1 membership, on the basis of qRT-PCR $\log_2$ normalised expression of the nine-signature genes (figure 4). This model classified patients with a specificity and sensitivity of 100%, and had a cross-validated success of 88%. Using this nine-gene qRT-PCR model we were able to reclassify the ten previously unclassified patients (six castration-resistant prostate cancer and four active surveillance) by whole-gene expression LPD analysis. Overall, on the basis of these qRT-PCR results, 16 of 64 patients with castration-resistant prostate cancer from this derivation test cohort were classified into the LPD1 cohort. We then analysed the expression of these nine-genes by qRT-PCR in the validation (MSKCC) cohort; 31 of 70 (44%) patients with castration-resistant prostate cancer from this cohort were classified into the LPD1 group.

In both derivation and validation cohorts, the LPD1 signature was significantly associated with poor prognostic features including higher baseline PSA; higher baseline alkaline phosphatase; lower albumin; and higher lactate dehydrogenase levels (table 1). The LPD1 classifier nine-gene signature was also associated in the validation cohort, but not in the initial derivation cohort, with lower haemoglobin levels (p=0·0009) and an unfavourable CTC count of ≥5 per 7·5 mL of blood (p=0·004). These differences are related to the fact that patients in this validation cohort had more advanced disease.

Furthermore, in the derivation test cohort overall survival was significantly worse for patients in the LPD1 subgroup than for non-LPD1 patients (figure 4). Multivariable Cox regression analyses confirmed the independent prognostic role of LPD1 membership in castration-resistant prostate cancer for the derivation test cohort (hazard ratio 3·1, 95% CI 1·2–7·6; table 3). The LPD1 subgroup of patients had an 18-month mortality rate of 50% (eight of 16) with an odds ratio of 5·0 (95% CI 1·5–15·5) when compared with the non-LPD1 subgroup. We also studied the improvement in c-index of the addition of the nine-gene signature to the baseline model. In the derivation cohort the addition of LPD1 membership based on the nine-gene classifier improved the c-index from 0·729 to 0·749 (p=0·016).

Crucially, in the validation (MSKCC) cohort, LPD1 membership was also significantly associated with poorer overall survival (figure 4). The 18-month mortality rate for LPD1 patients was 84% (26 of 31), with an odds ratio of 5·6 (95% CI 1·7–17·9). Multivariable Cox regression analyses for this cohort confirmed the independent prognostic value of LPD1 membership (table 3). In the validation model the addition of LPD1 membership to the baseline model showed an improvement of the c-index from 0·775 to 0·784 (p=0·130).

	Stage I	p value	Stage II	p value
Univariable analysis				
LPD1 membership	4·51 (2·12–9·20)	<0·0001	2·53 (1·45–4·43)	0·001
Age (as continuous)	0·99 (0·94–1·03)	0·513	1·02 (0·98–1·06)	0·423
Gleason				
7 *vs* <7	2·96 (0·81–10·83)	0·101	0·37 (0·17–0·82)	0·014
>7 *vs* <7	2·74 (0·82–9·15)	0·102	0·35 (0·17–0·72)	0·004
PSA (times × median value)	1·07 (1·02–1·13)	0·011	1·02 (1·00–1·03)	0·020
ECOG ≥1	2·03 (0·97–4·24)	0·060	3·62 (1·73–7·68)	0·001
Albumin (as continuous)	0·88 (0·82–0·94)	<0·0001	0·81 (0·73–0·90)	<0·0001
Haemoglobin (as continuous)	1·12 (0·84–1·50)	0·432	0·74 (0·62–0·88)	<0·0001
Alkaline phosphatase (times × UNL)	1·10 (1·04–1·16)	0·001	1·07 (1·01–1·14)	0·048
Docetaxel	2·43 (1·30–4·57)	0·006	2·10 (1·05–4·22)	0·037
Number chemotherapy lines	1·59 (1·09–2·33)	0·017	1·61 (1·13–2·28)	0·008
PSADT (as continuous)	0·95 (0·87–1·05)	0·327	NA	NA
Lactate dehydrogenase ratio (times × UNL)	1·39 (0·89–2·14)	0·138	2·84 (1·99–4·04)	<0·0001
Multivariable analysis (baseline model)				
LPD1 membership	4·92 (1·36–17·74)	0·015	1·84 (0·99–3·42)	0·054
Age (as continuous)	0·98 (0·93–1·05)	0·613	0·95 (0·91–1·00)	0·042
Gleason				
7 *vs* <7	2·20 (0·47–10·35)	0·319	0·37 (0·14–0·96)	0·041
>7 *vs* <7	1·09 (0·27–4·41)	0·899	0·31 (0·13–0·73)	0·008
PSA (times × median value)	1·01 (0·93–1·09)	0·789	1·00 (0·97–1·03)	0·996
ECOG ≥1	2·52 (0·93–6·82)	0·069	3·65 (1·53–8·74)	0·004
Albumin (as continuous)	0·92 (0·81–1·05)	0·201	0·81 (0·72–0·91)	0·001
Haemoglobin (as continuous)	1·12 (0·84–1·50)	0·432	0·98 (0·77–1·25)	0·870
Alkaline phosphatase (times × UNL)	0·99 (0·86–1·14)	0·938	1·02 (0·91–1·15)	0·690
Docetaxel	9·34 (1·04–83·87)	0·046	0·18 (0·03–1·02)	0·053
Number chemotherapy lines	0·31 (0·07–1·42)	0·132	3·56 (1·43–8·84)	0·006
PSADT (as continuous)	0·97 (0·85–1·10)	0·623	NA	NA
Lactate dehydrogenase ratio (times × UNL)	3·41 (0·27–44·08)	0·346	2·86 (1·54–5·33)	0·001
Optimal multivariable analysis model with significant variables in the univariable analysis*				
LPD1 membership	3·05 (1·22–7·64)	0·017	1·84 (1·01–3·35)	0·047
Albumin (as continuous)	0·91 (0·83–0·99)	0·037	0·85 (0·76–0·95)	0·003
Alkaline phosphatase (times × UNL)	1·06 (0·98–1·14)	0·146	..†	..†
Docetaxel	12·07 (2·22–65·47)	0·010	2·44 (1·14–5·23)	0·020
ECOG ≥1	2·01 (0·89–4·51)	0·093	2·60 (1·19–5·65)	0·016
Lactate dehydrogenase ratio (times × UNL)	..†	..†	2·28 (1·58–3·30)	<0·0001

Data are hazard ratio (95% CI) unless stated otherwise. p values were calculated with a Wald test. LPD=latent process decomposition. PSA=prostate-specific antigen. ECOG=Eastern Cooperative Oncology Group. PSADT=PSA doubling time. NA=not applicable. UNL=upper normal limit. *Owing to high collinearity between previous docetaxel and the number of chemotherapy treatment lines, the later variable was removed from the optimal model. †Not included in the optimal model.

Table 3: **Cox regression modelling results for overall survival outcome**

Discussion

In this study we showed the feasibility of identifying robust gene expression profiles from peripheral blood in patients with prostate cancer. We derived a nine-gene signature that has been internally verified by qRT-PCR. Additionally, this signature had independent prognostic clinical utility, with about 75% of patients with castration-resistant prostate cancer (non-LPD1) having a significantly better prognosis than the 25% classified as belonging to the LPD1 subgroup. The prognostic value of the LPD1 signature was also validated in a second independent set of patients with castration-resistant prostate cancer treated at a different institution. The derivation of our expression profile was also supported by hierarchical clustering analyses, where cluster A identified largely the same set of patients with poor prognosis. Nonetheless, LPD analyses provided better discrimination between poor and good prognosis patients with castration-resistant prostate cancer when compared with hierarchical clustering or PCA.

This LPD1 expression signature was also associated with known prognostic biomarkers of castration-resistant prostate cancer, including high alkaline phosphatase, PSA, lactate dehydrogenase, CTC counts, and low haemoglobin and albumin levels. LPD1 subclassification was confirmed as independently prognostic for overall survival by multivariable analyses done for both derivation test and validation cohorts of patients with castration-resistant prostate cancer. The addition of LPD1 signature significantly improved the predictive performance of the model as measured by the c-statistic of both baseline models. While this c-statistic improvement was small, this does not truly reflect the potential clinical relevance of LDP1 classification. Previous studies of multiple prognostic biomarkers in various diseases have commented on the limited association between clinical relevance and improved c-statistic in such analyses.[26,27]

Seven genes in this LPD1 signature (*TFDP1, TMCC2, RIOK3, SLC4A1, SNCA1, STOM, HMBS*) were highly expressed in CD71+ early erythroid cells, but not in normal blood cells. The transferrin receptor (or CD71) is usually expressed in immature erythroid precursors,[28] and such cells are usually very rare in peripheral blood. Preclinical in-vivo studies show that primitive prostate-cancer clonogenic cells compete for haemopoietic stem cell niches in bone marrow and displace haemopoietic stem cells into blood.[12] As erythroid differentiation is the first lineage to develop from haemopoietic stem cells, this signature could represent a sensitive early marker of bone-marrow infiltration or failure.

Furthermore, eight of 15 most highly downregulated genes in LPD1, and the top four pathways identified by Ingenuity analyses, were related to B-cell and T-cell function. These data suggest that LPD1 patients with this signature could have a poorer prognosis due to altered immune function. This altered function might be due to previous anticancer treatment, a concomitant disorder, or concurrent treatment. Previous studies have reported increased levels of regulatory T cells in patients with prostate cancer,[29] with cells expressing *Fox-P3* linked to acquired treatment resistance.[30] In our study, no association was seen with concomitant diseases or treatments; specifically steroid use at the time of blood sampling for expression analyses was not associated with the LPD1 profile and the different LPD groups. Ross and colleagues[31] have also described a prognostic signature for patients with castration-resistant prostate cancer using whole-blood RNA and qRT-PCR analyses of 168 preselected genes that associate with altered cellular and humoral immunity focusing on the monocyte and macrophage lineage. Overall, these results support other emerging data implicating the effect of the immune system on outcome from castration-resistant prostate cancer; these results include recent findings describing the association of increased circulating levels of the macrophage inhibitory cytokine-1 with response to docetaxel and castration-resistant prostate cancer survival.[32] All of these findings provide further support that whole-blood expression changes related to bone-marrow lineages are prognostic in castration-resistant prostate cancer.

Panel: Research in context

Systematic review

Serum prostate-specific antigen (PSA) has been widely studied in management of prostate cancer[6] but it is not a reliable prognostic factor for overall survival.[6,7] The identification and validation of other multipurpose blood-based biomarker assays is crucial to individualise management of prostate cancer.[3] Blood cells express gene transcripts in response to microenvironmental and macroenvironmental changes such as prostate-cancer bone-marrow invasion, which affects haemopoiesis by competing with primitive haemopoietic cell niches,[12,13] and chromosomal aberrations and epigenetic changes have been identified in peripheral blood mononuclear cells from patients with solid tumours including prostate cancer.[14-17] We searched PubMed for reports published between Jan 1, 1990, and Dec 1, 2011, with the search terms "prostate cancer", "prognostic markers", and/or "blood expression profiling", without language restrictions. No publications on the relation between these three terms were identified. Therefore, we did a study to assess the prognostic value of whole-blood expression profiling for patients with castration-resistant prostate cancer.

Interpretation

Our study assessed whole-blood gene expression profiles in an initial series, which included patients with early prostate cancer with low-burden, low-risk disease and patients with castration-resistant prostate cancer. We identified by unsupervised latent process decomposition analyses an expression profile associated exclusively with castration-resistant prostate cancer, high-burden disease, and poor outcome in patients with castration-resistant prostate cancer. A reduced nine-gene signature was derived to identify patients with castration-resistant prostate cancer expressing this unfavourable gene profile. The prognostic value of this signature was then validated in a second independent series of patients with castration-resistant prostate cancer. This signature could add prognostic value to other previously described clinicopathological risk factors for castration-resistant prostate cancer. Our results suggest that whole-blood gene profiling could identify gene expression signatures that stratify patients with CRPC into very distinct prognosis groups.

These changes in gene expression are unlikely to be influenced by mRNA extracted from CTCs.[33] In our series, the most upregulated genes in LPD processes were not correlated with CTC counts and were not prostate-cancer specific genes. We extracted a median of 2·5 µg of RNA per mL of blood, which is equivalent to the RNA from about 2·5 million cells;[34] thus even in patients with very high CTC counts, their CTC RNA would be diluted by more than 100 000-fold. Although we cannot entirely exclude the fact that CTCs or circulating cell-free tumour nucleic acids can affect these profiles, we postulate that these expression profiles are most likely to be due to changes in blood cells.

Patients with early stage disease and good prognosis who were selected for management by active surveillance and were used as controls for the derivation of our castration-resistant prostate cancer signature were almost exclusively represented by the LPD3 and LPD4 subgroups. Previous studies have shown that changes in the expression patterns of total RNA extracted from peripheral blood mononuclear cells can distinguish cancer patients from non-cancer patients.[33,35] In our study, 27 (44%) patients with castration-resistant prostate cancer in our derivation set were segregated with active surveillance patients in LPD3 and LPD4. The clustering of active surveillance and castration-resistant prostate cancer cases with similar expression profiles suggests either that none of the LPD profiles correspond with a normal non-cancer expression pattern, or that these profiles in castration-resistant prostate cancer reflect a less aggressive disease state. We did not have data for a compatible age-matched non-prostate cancer control series; however, we did a series of in-silico analyses (appendix) with other PAXgene blood microarray data from healthy controls[36] and could not find similar expression profiles.

Overall, our results to date suggest that these expression profiles are clinically relevant (panel). Further studies using age-matched non-cancer patients as well as studies of multiple sequential samples from the same patient as they progress from an early castration-resistant prostate cancer state to late-stage disease are now warranted. Future studies should also assess the clinical use of whole-blood expression signatures in patients undergoing active surveillance; these studies will require blood samples from age-matched healthy volunteers as controls. Further studies could also explore microRNA (miRNA) expression profiles, which could add value to these analyses since miRNAs have been reported to distinguish between different tumour types with more accuracy than mRNA signatures.[37,38] In prostate cancer, miRNA profiling can differentiate between malignant and non-malignant prostate disease.[38] Such studies can also be pursued in other body fluids such as urine, where miRNAs might be more stable than mRNAs.

In conclusion, blood from patients with prostate cancer contains mRNA expression information that is prognostic in patients with castration-resistant prostate cancer. Expression array analyses of blood merits further assessment in predictive and pharmacodynamic biomarker studies for new anticancer drugs including immune therapies. This nine-gene qRT-PCR signature will aid further prospective studies using more widely available and less costly technologies such as microfluidic TaqMan-based qRT-PCR assays in large prospective trials.

Contributors

DO and JSdB designed the study. DO, DCD, CP, GA, MF, AHMR, LB, NBO, RJ, HIS, and JSdB recruited patients. DO, DCD, GA, AHMR, LB, RJ, HIS, and JSdB collected data and samples. DO, JC, and EC undertook the experimental procedures. DO, DB, JC, and JSdB analysed the data and DB did the statistical analyses. All authors wrote and approved the final report.

Conflicts of interest

We declare that we have no conflicts of interest.

Acknowledgments

The Drug Development Unit of The Royal Marsden NHS Foundation Trust and the Institute of Cancer Research is supported in part by a programme grant from Cancer Research UK. Funding was also provided by the Prostate Cancer Charity; an Experimental Cancer Medicine Centre grant (to the Institute of Cancer Research); The Prostate Cancer Foundation (to The Institute of Cancer Research and Memorial Sloan-Kettering Cancer Center); The NCI SPORE (Specialized Programs of Research Excellence) in Prostate Cancer, The Experimental Therapeutics Center, Geoffrey Beene Cancer Research Center, Mr William H Goodwin and Mrs Alice Goodwin, and the Commonwealth Foundation for Cancer Research (to the Memorial Sloan Kettering Cancer Center). DO was supported by a SEOM (Spanish Society of Medical Oncology) Translational Research Fellowship. DB was also supported by the Bob Champion Cancer Trust. DCD received support from the Department of Defense Prostate Cancer Research Program Physician Research Award PC081051.

References

1 Albertsen PC. Treatment of localized prostate cancer: when is active surveillance appropriate? *Nat Rev Clin Oncol* 2010; **7:** 394–400.
2 Ang JE, Olmos D, de Bono JS. CYP17 blockade by abiraterone: further evidence for frequent continued hormone-dependence in castration-resistant prostate cancer. *Br J Cancer* 2009; **100:** 671–75.
3 Fiorentino M, Capizzi E, Loda M. Blood and tissue biomarkers in prostate cancer: state of the art. *Urol Clin North Am* 2010; **37:** 131–41.
4 Basu B, Olmos D, de Bono JS. Targeting IGF-1R: throwing out the baby with the bathwater? *Br J Cancer* 2011; **104:** 1–3.
5 Schwarzenbach H, Hoon DS, Pantel K. Cell-free nucleic acids as biomarkers in cancer patients. *Nat Rev Cancer* 2011; **11:** 426–37.
6 Fradet Y. Biomarkers in prostate cancer diagnosis and prognosis: beyond prostate-specific antigen. *Curr Opin Urol* 2009; **19:** 243–46.
7 Bickers B, Aukim-Hastie C. New molecular biomarkers for the prognosis and management of prostate cancer—the post PSA era. *Anticancer Res* 2009; **29:** 3289–98.
8 Olmos D, Arkenau HT, Ang JE, et al. Circulating tumour cell (CTC) counts as intermediate end points in castration-resistant prostate cancer (CRPC): a single-centre experience. *Ann Oncol* 2009; **20:** 27–33.
9 Attard G, Swennenhuis JF, Olmos D, et al. Characterization of ERG, AR and PTEN gene status in circulating tumor cells from patients with castration-resistant prostate cancer. *Cancer Res* 2009; **69:** 2912–18.
10 Sotiriou C, Piccart MJ. Taking gene-expression profiling to the clinic: when will molecular signatures become relevant to patient care? *Nat Rev Cancer* 2007; **7:** 545–53.
11 van't Veer LJ, Bernards R. Enabling personalized cancer medicine through analysis of gene-expression patterns. *Nature* 2008; **452:** 564–70.
12 Shiozawa Y, Pedersen EA, Havens AM, et al. Human prostate cancer metastases target the hematopoietic stem cell niche to establish footholds in mouse bone marrow. *J Clin Invest* 2011; **121:** 1298–312.

13 Shiozawa Y, Pienta KJ, Taichman RS. Hematopoietic stem cell niche is a potential therapeutic target for bone metastatic tumors. *Clin Cancer Res* 2011; **17:** 5553–58.

14 Ozen M, Hopwood VL, Johnston DA, et al. Aneuploidy index in blood: a potential marker for early onset, androgen response, and metastasis in human prostate cancer. *Urology* 1999; **53:** 381–85.

15 Toncheva D, Nacheva M. Chromosomal instability in cancer patients. *Hum Genet* 1998; **103:** 367–71.

16 Dotan ZA, Dotan A, Ramon J, Avivi L. Altered mode of allelic replication accompanied by aneuploidy in peripheral blood lymphocytes of prostate cancer patients. *Int J Cancer* 2004; **111:** 60–66.

17 Dotan ZA, Dotan A, Ramon J, Avivi L. Aberrant allele-specific replication, independent of parental origin, in blood cells of cancer patients. *BMC Cancer* 2008; **8:** 390.

18 Rainen L, Oelmueller U, Jurgensen S, et al. Stabilization of mRNA expression in whole blood samples. *Clin Chem* 2002; **48:** 1883–90.

19 McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM, and the Statistics Subcommittee of the NCI-EORTC Working Group on Cancer Diagnostics. Reporting recommendations for tumor marker prognostic studies. *J Clin Oncol* 2005; **23:** 9067–72.

20 Gail M, Gart JJ. The determination of sample sizes for use with the exact conditional test in 2 x 2 comparative trials. *Biometrics* 1973; **29:** 441–48.

21 Brazma A, Hingamp P, Quackenbush J, et al. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet* 2001; **29:** 365–71.

22 Rogers S, Girolami M, Campbell C, Breitling R. The latent process decomposition of cDNA microarray data sets. *IEEE/ACM Trans Comput Biol Bioinformatics* 2005; **2:** 143–56.

23 Breiman L. Random forests. *Mach Learn* 2001; **45:** 5–32.

24 Clipsham RC, McCabe ER. Single-tube gene-specific expression analysis by high primer density multiplex reverse transcription. *Mol Genet Metab* 2001; **74:** 435–48.

25 Clark J, Merson S, Jhavar S, et al. Diversity of TMPRSS2-ERG fusion transcripts in the human prostate. *Oncogene* 2007; **26:** 2667–73.

26 Janes H, Pepe MS, Gu W. Assessing the value of risk predictions by using risk stratification tables. *Ann Intern Med* 2008; **149:** 751–60.

27 Ware JH. The limitations of risk factors as prognostic tools. *N Engl J Med* 2006; **355:** 2615–17.

28 Kono M, Kondo T, Takagi Y, Wada A, Fujimoto K. Morphological definition of CD71 positive reticulocytes by various staining techniques and electron microscopy compared to reticulocytes detected by an automated hematology analyzer. *Clin Chim Acta* 2009; **404:** 105–10.

29 Miller AM, Lundberg K, Ozenci V, et al. CD4+CD25high T cells are enriched in the tumor and peripheral blood of prostate cancer patients. *J Immunol* 2006; **177:** 7398–405.

30 Valdman A, Jaraj SJ, Compérat E, et al. Distribution of Foxp3-, CD4- and CD8-positive lymphocytic cells in benign and malignant prostate tissue. *APMIS* 2010; **118:** 360–65.

31 Ross RW, Galsky MD, Scher HI, et al. A whole-blood RNA transcript-based prognostic model in men with castration-resistant prostate cancer: a prospective study. *Lancet Oncol* 2012; published online Oct 8. http://dx.doi.org/10.1016/S1470-2045(12)70263-2.

32 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.

33 Showe MK, Vachani A, Kossenkov AV, et al. Gene expression profiles in peripheral blood mononuclear cells can distinguish patients with non-small cell lung cancer from patients with nonmalignant lung disease. *Cancer Res* 2009; **69:** 9202–10.

34 Hausen P, Stein H. On the synthesis of RNA in lymphocytes stimulated by phytohemagglutinin. 1. Induction of uridine-kinase and the conversion of uridine to UTP. *Eur J Biochem* 1968; **4:** 401–06.

35 Aarøe J, Lindahl T, Dumeaux V, et al. Gene expression profiling of peripheral blood cells for early detection of breast cancer. *Breast Cancer Res* 2010; **12:** R7.

36 Kennedy L, Vass JK, Haggart DR, et al. Hematopoietic lineage transcriptome stability and representation in PAXgene collected peripheral blood utilising SPIA single-stranded cDNA probes for microarray. *Biomark Insights* 2008; **3:** 403–17.

37 Calin GA, Croce CM. MicroRNA signatures in human cancers. *Nat Rev Cancer* 2006; **6:** 857–66.

38 Chiam K, Ricciardelli C, Bianco-Miotto T. Epigenetic biomarkers in prostate cancer: current and future uses. *Cancer Lett* 2012; published online March 3. http://dx.doi.org/10.1016/j.canlet.2012.02.011.