[CANCER RESEARCH 63, 4196–4203, July 15, 2003]

Survival Analysis of Genome-Wide Gene Expression Profiles of Prostate Cancers Identifies New Prognostic Targets of Disease Relapse[1,2]

Susan M. Henshall, Daniel E. H. Afar,[3] Jordan Hiller,[3] Lisa G. Horvath, David I. Quinn, Krishan K. Rasiah, Kurt Gish,[3] Dorian Willhite, James G. Kench, Margaret Gardiner-Garden, Phillip D. Stricker, Howard I. Scher, John J. Grygiel, David B. Agus, David H. Mack, and Robert L. Sutherland[4]

Cancer Research Program, Garvan Institute of Medical Research, St. Vincent's Hospital, Darlinghurst, Sydney, New South Wales 2010, Australia [S. M. H., L. G. H., D. I. Q., K. K. R., J. G. K., M. G-G., R. L. S.]; Genomics Research, Eos Biotechnology, South San Francisco, California [D. E. H. A., J. H., K. G., D. W., D. H. M.]; Department of Medical Oncology (J. J. G.), and Department of Urology, St. Vincent's Hospital, Darlinghurst, Sydney, New South Wales 2010, Australia [P. D. S.]; Cedars-Sinai Prostate Cancer Center, Los Angeles, California [D. B. A.]; and Genitourinary Oncology Service, Memorial Sloan-Kettering Cancer Center, New York, New York [H. S.]

ABSTRACT

Current models of prostate cancer classification are poor at distinguishing between tumors that have similar histopathological features but vary in clinical course and outcome. Here, we applied classical survival analysis to genome-wide gene expression profiles of prostate cancers and preoperative prostate-specific antigen (PSA) levels from each patient, to identify prognostic markers of disease relapse that provide additional predictive value relative to PSA concentration. Three of ~200 probesets showing strongest correlation with relapse were identified as the gene for the putative calcium channel protein, trp-p8, with loss of *trp-p8* mRNA expression associated with a significantly shorter time to PSA relapse-free survival. We observed subsequently that *trp-p8* is lost in the transition to androgen independence in a prostate cancer xenograft model and in prostate cancer tissue from patients treated preoperatively with antiandrogen therapy, suggesting that *trp-p8* is androgen regulated, and its loss may be associated with more advanced disease. The identification of trp-p8 and other proteins implicated in the phosphatidylinositol signal transduction pathway that are associated with prostate cancer outcome, both here and in other published work, suggests an integral role for this pathway in prostate carcinogenesis. Thus, our findings demonstrate that multivariable survival analysis can be applied to gene expression profiles of prostate cancers with censored follow-up data and used to identify molecular markers of prostate cancer relapse with strong predictive power and relevance to the etiology of this disease.

INTRODUCTION

Prostate cancer will account for an estimated 30% (189,000) of new cancer cases in men in the United States in 2002 (1). Many of these newly diagnosed cases are a result of the extensive use of PSA[5] screening and the subsequent diagnosis of prostate cancer at an early stage and age. However, despite the introduction of PSA screening, the mortality from prostate cancer has remained relatively constant. The implications of this are that: (*a*) there are a large group of men diagnosed with prostate cancer for whom radical treatment is probably unnecessary and who will die with their prostate cancer rather than from it; and (*b*) there are a group of men for whom early detection offers the possibility of cure that may be denied by delay. Consequently, identifying these groups of men at the time of diagnosis is critical to the optimal management of prostate cancer.

Although the benefits of PSA screening are widely debated, this serum marker remains one of only a few preoperative parameters of prognostic utility. To enhance the predictive value of individual parameters with outcomes, nomograms have been developed that incorporate parameters that are measured routinely in clinical practice to predict the probability of PSA relapse-free survival of individual patients both before and after therapy (2–6). Models such as these currently form the basis of routine clinical decision-making, but such classification systems cannot explore differences in outcomes observed between cancers with similar histopathological features. Hence, there remains a critical need for increased accuracy in the subcategorization of prostate cancers to identify those with an aggressive phenotype. One approach is to define patterns of gene expression that correlate with disease phenotype and patient outcome. Here, we undertook a systematic search for novel biomarkers of prostate cancer prognosis by outcome-based analyses of transcript profiles.

MATERIALS AND METHODS

Tissue Collection and Preparation of RNA. A cohort of 72 fresh-frozen prostate cancers was collected from patients with localized prostate cancer treated by RP at St. Vincent's Hospital. The primary outcome, disease-specific relapse, was measured from the date of RP and was defined as a rise in serum PSA >0.3 ng/ml with subsequent additional rises. After inking of the external limits of the prostate immediately after removal and before formalin-fixation, up to six 5-mm core biopsies were taken and stored at −80°C for later RNA extraction. The proportion of invasive cancer in the biopsy sample was then estimated retrospectively by either frozen sectioning of the biopsy and H&E staining, or by examination of archival formalin-fixed, paraffin-embedded tissue surrounding the biopsy site. Only those biopsies that contained ≥75% invasive cancer were used for subsequent transcript profiling. Only one biopsy per patient was analyzed.

Xenograft Model. The androgen-dependent LuCaP-35 (7) prostate cancer xenograft (generously provided by Robert L. Vessella, University of Washington, Seattle, WA) was grown as s.c. tumors in nude male mice. To study the androgen-withdrawal process, tumor-bearing mice were castrated and monitored for tumor regression and PSA levels. Tumors were harvested from mice before castration and at various time points (1–100 days) postcastration, and were processed for microarray analysis. For data analysis and identification of androgen-regulated genes, *i.e.,* genes that behaved similar to PSA, the LuCaP-35 xenografts were binned into two groups (days 0–2 postcastration *versus* 5–100 days postcastration), because PSA levels were high at days 0–2 and dropped precipitously at day 5 postcastration. Genes that showed a significant ($P < 0.01$) difference in the means of each group were identified by a standard Student's t test.

RNA Extraction and Microarray Protocols. Preparation of total RNA from fresh-frozen prostate and xenograft tissue was performed by extraction with Trizol reagent (Life Technologies, Inc., Gaithersburg, MD) and was reverse transcribed using a primer containing oligodeoxythymidylic acid and a T7 promoter sequence. The resulting cDNAs were then *in vitro* transcribed in the presence of biotinylated nucleotides (Bio-11-CTP and Bio-16-UTP) using the T7 MEGAscript kit (Ambion, Austin, TX).

Received 10/25/02; accepted 5/15/03.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] Supported by grants from the National Health and Medical Research Council of Australia, The Cancer Council New South Wales, the R. T. Hall Trust, Freedman Foundation, Royal Australasian College of Surgeons, Australasian Urological Foundation, Prostate Cancer Foundation of Australia, and David Wilson Trust.

[3] Present address: Protein Design Labs Inc., Fremont, CA 94555.

[2] Supplementary data are submitted for review as part of this manuscript.

[4] To whom correspondence should be addressed, at Cancer Research Program, Garvan Institute of Medical Research, 384 Victoria Street, Darlinghurst, New South Wales, 2010 Australia. Phone: 612-9295-8322; Fax: 612-9295-8321; E-mail: r.sutherland@garvan.org.au.

[5] The abbreviations: PSA, prostate-specific antigen; Ca^{2+}, calcium; HR, hazard ratio; IQR, interquartile range; IP3R, inositol triphosphate receptor; ISH, *in situ* hybridization; NHT, neoadjuvant hormone therapy; pFDR, positive false discovery rate; RP, radical prostatectomy; EST, expressed sequence tag; Gn-RH, gonadotrophin-releasing hormone; DIG, digoxigenin.

The biotinylated targets were hybridized to the Eos Hu03, a customized Affymetrix GeneChip (Affymetrix, Santa Clara, CA) oligonucleotide array comprising 59,619 probesets representing 46,000 unique sequences including both known and FGENESH predicted exons that were based on the first draft of the human genome. The Hu03 probesets consist of perfect match probes only, most probesets having 6 or 7 probes. Hybridization signals were visualized using phycoerythrin-conjugated streptavidin (Molecular Probes, Eugene, OR).

Normalization of the gene expression data was performed as follows. The probe-level intensity data from each array were fitted to a fixed γ distribution, using an inverse γ function to map the empirical cumulative distribution of intensities to the desired γ distribution. This procedure is akin to other per-chip normalization procedures, such as fixing the mean and SD of each chip to a standard value, except it is more stringent in that it fixes the entire distribution rather than one or two parameters. The purpose of per-chip normalization is to remove between-chip variation, on the assumption that it is attributable to nonbiological factors, *i.e.,* technical noise. The scale parameter for the γ distribution was chosen to yield a distribution with an arbitrary mean value of 300, and the shape parameter of 0.81 was chosen to reproduce the typical shape of the empirical distribution seen in good samples.

A single measure of average intensity was calculated for each probeset using Tukey's trimean of the intensity of the constituent probes (8). The trimean is a measure of central tendency that is resistant to the effects of outliers. Finally, a background subtraction was applied to each average intensity measure to correct for nonspecific hybridization. The average intensity measure of a "null" probeset consisting of 491 probes with scrambled sequence was subtracted from all of the other probesets on the chip.

Statistical Methods. Before survival analysis, a screen was applied to the expression data to eliminate probesets without meaningful variation. For each probeset, the ratio of the 90th percentile to the 15th percentile intensity measure was required to be at least 2, and the minimum expression level was required to be at least 150 average intensity units. This screen reduced the initial set of 59,619 probesets to a subset of 8,521 probesets for additional examination. Cox proportional hazards analyses with pretreatment PSA concentration dichotomized at 20 ng/ml and gene expression modeled as a continuous variable were computed for each probeset, to identify gene expression that predicts PSA recurrence (9). To assist interpretation, we next calculated the IQR HR for each probeset. Because the expression data are treated here as continuous covariates, HRs expressed in their natural scale illustrate only the change in risk of relapse associated with a change of 1 unit on the expression scale, a change too small to be comprehended easily. To put the HRs and associated confidence limits on a more interpretable scale, we present here the HR associated with a change in expression values equivalent to 1 IQR of the sample data for each probeset. The IQR is simply the 75th percentile minus the 25th percentile, and thus contains the middle 50 percent of observations. The IQR HR was computed by multiplying the regression coefficient for each probeset by its own IQR before exponentiation.

The multiple hypothesis testing problem has been recognized as an important issue to address in microarray research. The many tests that are performed simultaneously on thousands of probesets greatly increases the chances of making type 1 errors (or false-positive findings). To assess the effect of multiple hypothesis testing, we adapted a method developed by Storey and Tibshirani (10) for calculating the pFDR, an estimate of the proportion of false positives present in a set of findings. This technique was developed explicitly for use with microarray data, for which the usual assumption of independence among tests is untenable. The procedure involves simulation of null data by randomly permutating the relapse status of subjects and reperforming the survival analyses. In each simulation, the number of relapsers and nonrelapsers (17 and 55, respectively) remains constant, but these designations are shuffled and assigned to patients at random. By performing the permutation many times and noting the number of findings at a given level of significance each time, an estimate of the expected number of false positives under null conditions is obtained. This figure is then divided by the number of actual findings to obtain an estimate of the proportion of false-positive findings.

Variables of clinical relevance were also modeled in univariate analyses for their ability to predict disease-free survival in the 72 prostate cancers using the Cox proportional hazards model. *Trp-p8* mRNA expression assessed by ISH, was reported as proportions within histological groups and compared between groups using a Fisher's Exact test.

Cluster analysis was used to explore and visualize the findings from the survival analyses. The distance metric used was the square root of $(1 - r)$, where r is the standard Pearson product-moment correlation, and the clustering algorithm used was Ward's minimum variance method (11).

All of the statistical analyses were performed using SAS (SAS Institute Inc., Cary, NC).

Tissue Microarray and ISH. Tissue microarrays were constructed as described previously (12), and were comprised of prostate cancer samples from 95 patients that are part of a previously published cohort of patients treated for localized prostate cancer by RP alone at St. Vincent's Hospital (13). Forty-seven of the 95 prostate cancer samples were independent of the cohort used for microarray analysis. In addition, 13 prostate cancer specimens were collected from patients treated for localized prostate cancer by RP who had received at least 3 months (range, 3–10 months) of preoperative neoadjuvant hormonal treatment (5 with antiandrogens alone, 6 with a combination of a Gn-RH analogue and antiandrogens and 2 with a Gn-RH analogue alone). *Trp-p8* expression in these 13 samples was assessed on conventional tissue sections.

For ISH, a 424-bp probe for *trp-p8* was derived from the 3′ end of the *trp-p8* gene and transcribed to produce a DIG-labeled riboprobe using an RNA DIG-labeling kit (Roche, Mannheim, Germany). ISH was performed on the Ventana Discovery instrument (Ventana Medical Systems, Tucson, AZ) using the RiboMap kit with protease P2 for 2 min (Ventana Medical Systems) and hybridization for 8 h at 65°C. Chromogenic detection was achieved with the BlueMap detection system as described by the manufacturer (Ventana Medical Systems). Independent readings of the pathology and the ISH measurements were performed.

RESULTS

Expression Profiling of Prostate Cancers. In this study, we sought to discover novel biomarkers that might predict for PSA relapse after RP using outcome-based statistical tools to analyze gene expression profiles of 72 prostate cancers. A criterion for selection was the ability to predict recurrence better than preoperative serum PSA concentration alone, because PSA is one of only a handful of markers that provide preoperative prognostic information. The 72 prostate tissues were collected at the time of RP from patients undergoing treatment for localized prostate cancer at St. Vincent's Hospital Campus. At last follow-up (median = 28.25 months; range, 4.9–90.3 months), 17 of the 72 (23.6%) patients had relapsed, of which 14 demonstrated a rise in postoperative PSA levels, whereas 3 patients were diagnosed with a rising PSA and local recurrence of disease. Consistent with published data (5, 6, 13), the significant predictors of prostate cancer relapse in this cohort on univariate analysis were Gleason score (HR = 1.88; $P = 0.027$), surgical margins (HR = 4.90; $P = 0.035$), and preoperative PSA concentration (HR = 4.43; $P = 0.006$; Table 1). The overall relapse rate of 23.6% and median time to relapse of 14 months in this group of 72 patients was similar to that observed in a cohort of 732 patients treated for localized prostate cancer by RP at the same institution between 1986 and 1999 (13).

Survival Analysis. The intensity of each probeset value was entered as a continuous explanatory variable in a Cox proportional

Table 1 *Clinicopathological characteristics of the prostate cancer cohort (n = 72) that were utilized in the survival analysis*

Variable	HR (confidence levels)	*P*
Gleason score[a]	1.88 (1.08–3.29)	0.027
Preoperative PSA concentration <20 ng/ml *vs.* ≥20 ng/ml	4.43 (1.53–12.79)	0.006
Seminal vesicle involvement positive *vs.* negative	2.33 (0.88–6.14)	0.086
Surgical margins positive *vs.* negative	4.90 (1.12–21.5)	0.035

[a] Gleason score was modeled as a continuous variable.

Table 2 *The first 50 genes, ranked by P, identified by survival analysis to be associated with prostate cancer relapse*

Rank	UniGene cluster	Genbank accession	Gene title	Risk of relapse[a]	*P*
1	Hs.189095	NM_020436	Sal-like 4	2.040	0
2	Hs.132860	AA845608	ESTs	0.341	0
3	Hs.75616	NM_014762	24-Dehydrocholesterol reductase (seladin-1)	0.293	0
4	Hs.42321	NM_173605	Hypothetical protein LOC283518	2.133	0
5	Hs.80667	NM_004726	RALBP1 associated Eps domain containing 2 (REPS2)	0.172	0
6	Hs.163543	NM_144704	Hypothetical protein FLJ30473	3.241	0
7	Hs.286	NM_000968	Ribosomal protein L4	0.215	0
8	Hs.114670	D49387	Leukotriene B4 12-hydroxydehydrogenase	2.380	0
9	Hs.366053	NM_024080	Transient receptor potential cation channel, subfamily M, member 8 (trp-p8)	0.260	0
10	Hs.366	AL389978	Immunoglobulin heavy chain variable region	2.436	0.001
11	Not available	BE250014	ESTs	0.295	0.001
12	Hs.257391	NM_032293	Hypothetical protein DKFZp761J1523	3.138	0.001
13	Hs.264330	AK024677	*N*-acylsphingosine amidohydrolase (acid ceramidase)-like	0.256	0.001
14	Hs.123468	NM_033225	CUB and Sushi multiple domains 1	0.185	0.001
15	Not available	AV656002	EST	0.251	0.001
16	Hs.129977	AW452344	ESTs	0.229	0.001
17	Hs.277728	NM_012429	SEC14-like 2	0.348	0.001
18	Hs.117950	NM_006452	Phosphoribosylaminoimidazole carboxylase	0.321	0.001
19	Hs.424973	BC018081	Clone IMAGE:4793702	0.225	0.001
20	Not available	AA354685	EST	0.363	0.001
21	Hs.356547	NM_138799	Hypothetical protein BC016005	0.337	0.001
22	Hs.7780	AL049969	cDNA DKFZp564A072	0.186	0.001
23	Hs.123387	AW204707	ESTs	0.375	0.001
24	Hs.377879	AK055649	cDNA FLJ31087 fis	3.112	0.001
25	Hs.248056	NM_005306	G protein-coupled receptor 43	0.211	0.001
26	Hs.301947	NM_014509	Kraken-like serine hydrolase	0.212	0.001
27	Hs.11923	NM_018982	Hypothetical protein DJ167A19.1	0.155	0.001
28	Hs.247423	NM_001617	Adducin 2 (β) (ADD2)	2.044	0.001
29	Not available	D80630	EST	2.753	0.001
30	Hs.21293	NM_003115	UDP-*N*-acteylglucosamine pyrophosphorylase 1	0.185	0.001
31	Hs.292859	C19035	ESTs, moderately similar to VPP2_HUMAN	2.375	0.001
32	Hs.68864	AW845987	Lipase, member H (LIPH), mRNA	0.273	0.001
33	Hs.405944	X57819	Ig λ chain	2.388	0.002
34	Hs.110103	NM_018427	RNA polymerase I transcription factor RRN3	0.337	0.002
35	Hs.256150	NM_080654	NY-REN-41 antigen	2.718	0.002
36	Hs.76847	NM_014610	α Glucosidase II alpha subunit	0.135	0.002
37	Hs.109694	AI199981	Oxysterol binding protein-like 8 (OSBPL8), mRNA.	4.511	0.002
38	Hs.71119	NM_006765	Putative prostate cancer tumor suppressor (N33)	0.281	0.002
39	Hs.146162	AK075364	ESTs	2.151	0.002
40	Hs.333417	NM_004930	Capping protein (actin filament) muscle Z-line, β	0.291	0.002
41	Hs.410998	AA402587	ESTs, Highly similar to MLL septin-like fusion	1.507	0.002
42	Hs.79136	NM_012319	LIV-1 protein, estrogen regulated	0.210	0.002
43	Hs.109154	AW969025	ESTs	0.281	0.002
44	Hs.433622	NM_007085	Follistatin-like 1 (FSTL1)	0.233	0.002
45	Not available	T65456	EST	0.195	0.002
46	Hs.405946	NM_003877	Suppressor of cytokine signaling 2 (SOCS2)	0.448	0.002
47	Hs.127699	NM_001369	Dynein, axonemal, heavy polypeptide 5 (DNAH5)	0.284	0.002
48	Hs.422118	NM_001402	Eukaryotic translation elongation factor 1 alpha 1	0.175	0.002
49	Hs.92033	NM_030768	Integrin-linked kinase-associated serine/threonine phosphatase 2C	5.564	0.002
50	Hs.124895	AA907734	ESTs	3.399	0.002

[a] The risk of relapse is the IQR HR calculated for each probeset as described in "Materials and Methods."

hazards survival analysis predicting relapse. Pretreatment PSA concentration was also entered as a predictor in each analysis. From this analysis, 266 probesets were significant predictors of relapse at $P < 0.01$.

The pFDR (10) was calculated to estimate the number of false positives present among the 266 findings obtained at the selected 0.01 significance level. On the basis of 500 random permutations of the class labels, the final estimate for the pFDR was 23%. Thus, we can expect that ~61 of the 266 findings are false positives.

Another way to address the issue of significance across the entire dataset is to estimate the likelihood of obtaining 266 findings significant at $P < 0.01$. This can be accomplished using the 500 permuted simulations described above. Among the 500 permutations, only 20, or 4%, had 266 or more findings significant at $P < 0.01$. Thus, under simulated null conditions, it is estimated that the probability of finding at least 266 differentially expressed probesets is 4%.

Identification of the Candidate Marker Genes. The 266 probesets identified by survival analysis included both known genes and hypothetical genes of unknown function, as well as ESTs (Table 2; Supplementary Data). Cluster analysis performed in both dimensions on the 72 RP samples and these 266 probesets using Ward's minimum variance method (11) identified two gene expression subgroups (Fig. 1). Sixteen of the 17 patients known to have experienced a PSA relapse were clustered in one gene expression group characterized by a relative increase in expression of 85 genes (cluster 1) and loss of expression of 181 genes (cluster 2; Fig. 1). An additional 22 patients that were disease-free at the time of censoring were located in this expression cluster, and may suggest that these patients have an increased propensity for relapse in the future. Thirty-two patients who were disease-free at the time of censoring constituted the second expression group, which also included 1 patient who had experienced a PSA relapse.

Three of the ~200 probesets showing strongest correlation with relapse in our model were identified as the gene for the putative Ca^{2+} channel protein, *trp-p8* (Table 2; Supplementary Data; Ref. 14). For all three of the probesets, loss of expression of *trp-p8* mRNA was associated with a significantly shorter time to PSA relapse-free survival. The *trp-p8* probeset having the strongest association with relapse had an IQR HR of 0.26 (95% confidence interval, 0.12–0.54; $P < 0.001$), when PSA was included in the analysis (Fig. 2). To describe the magnitude of this effect, consider two hypothetical patients with identical PSA levels, but expression of *trp-p8* as measured

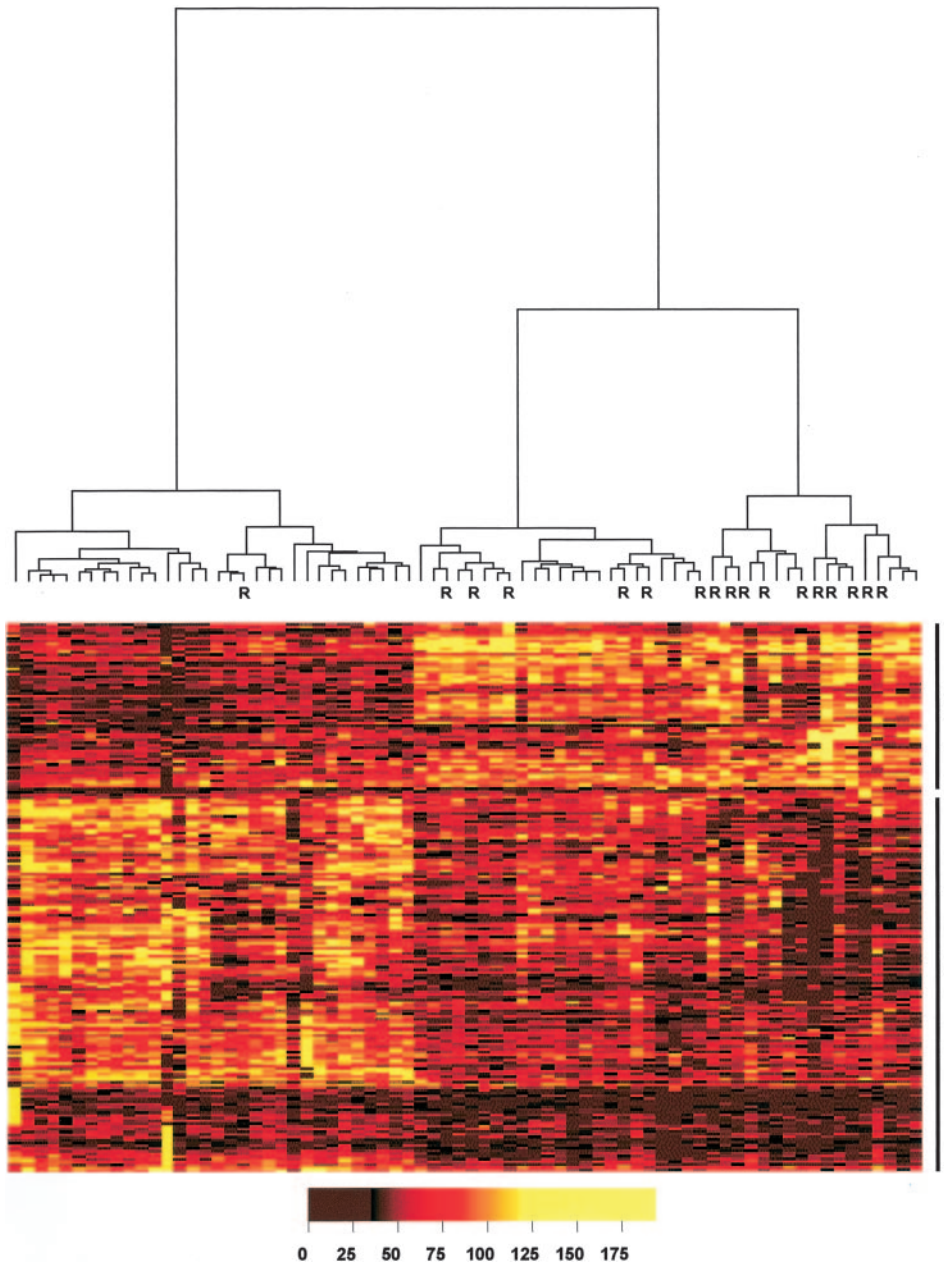

Fig. 1. Cluster analysis of prostate cancer samples from 72 patients treated for localized prostate cancer by RP. Each *column* represents a single RP specimen, and each *row* represents 1 of the 266 genes that demonstrated a strong association with PSA relapse in our model. The dendogram at the *top* shows the degree to which each prostate cancer is related to the others with respect to gene expression. The 17 patients known to have experienced a PSA relapse are indicated by *R*. The relative level of expression is indicated by the *color scale* at the bottom and is indicative of the normalized average intensity units of fluorescence signal detected by microarray analysis.

by this probeset differing by 118 expression units (which is the IQR for this probeset). The patient with lower *trp-p8* expression would have an ~4-fold increased risk of relapse. Loss of *trp-p8* remained a significant predictor of PSA relapse even when controlling for pretreatment PSA, Gleason score, and clinical stage simultaneously in a multivariate Cox proportional hazards model ($P = 0.0008$). Additional analysis of *trp-p8* was pursued subsequently, as this was the only gene within the top 10 ranked by *P* of which the expression was primarily restricted to the prostate (Fig. 3*A*). Low-level expression was detected in normal liver, and no detectable expression was seen in 32 distinct other normal tissues examined by oligonucleotide microarray analysis (Fig. 3*A*). These data confirm the findings of a recent study that also showed that *trp-p8* expression was prostate-specific (14). Analysis of 23 cancer cell lines showed that *trp-p8* is only expressed at very low levels in the androgen-dependent prostate cancer cell line LNCaP, but not in the androgen-independent prostate cancer cell lines, PC-3 and DU-145 (data not shown), consistent with previous data (14). Because this observation alone is not conclusive evidence that *trp-p8* expression is androgen regulated, we next used the androgen-dependent LuCaP-35 prostate cancer xenograft model to assess changes in *trp-p8* expression that occur during transition from androgen dependence to androgen independence of prostate cancer (7). Male LuCaP-35 mice were castrated, and tumors were harvested at several time points (0–100 days) after castration. High levels of *trp-p8* expression were detected on days 0–2 after castration, but not on days 5–100 after castration, and correlated significantly with PSA expression in the same mice (Pearson $P = 0.80$; Fig. 3, *B* and *C*). Because the mRNA expression of housekeeping genes such as β-actin and glyceraldehyde-3-phosphate dehydrogenase showed only minor changes throughout the experiment, these data suggest that *trp-p8* may be an androgen-regulated gene.

To gain additional insight into the putative association of *trp-p8* with androgen regulation, we examined the levels of *trp-p8* expression in the prostate tissue of patients who were treated with androgen deprivation therapy (NHT) before RP. ISH for *trp-p8* mRNA was performed on RP specimens from 13 patients who had received at

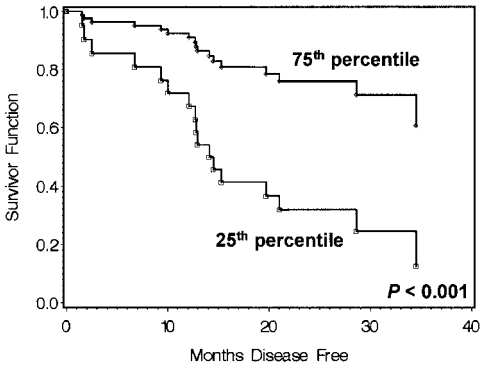

Fig. 2. Representative predicted survival curve for the *trp-p8* gene, which demonstrated a strong association ($P \leq 0.001$) with prostate cancer relapse in analyses using dichotomous preoperative serum PSA levels as a covariate. The survival curve represents the predicted survivorship function for a hypothetical patient who is PSA positive, and has a probeset intensity measure exactly at the 25th percentile of the observed data or at the 75th percentile. The survival curve indicates that loss of *trp-p8* expression is associated with an increased risk of PSA relapse.

least 3 months of preoperative NHT and the levels compared with tissue from 95 patients treated with RP alone (Fig. 4). These latter patients formed part of a large RP cohort described previously (13). Whereas *trp-p8* mRNA was detected in 80 of 95 (84%) prostate cancers from patients treated with RP alone, those patients who underwent NHT before RP demonstrated significantly less expression of *trp-p8*, with only 4 of 13 (31%) samples positive for *trp-p8* mRNA (Fisher's exact test, $P < 0.001$; Fig. 4). Taken together, these data from prostate cancer xenografts and clinical specimens are consistent with *trp-p8* being an androgen-regulated gene. We will now use a monoclonal antibody to trp-p8 to assess protein expression by immunohistochemistry in an independent cohort of formalin-fixed, paraffin-embedded prostate cancer specimens with known prostate cancer outcome that includes patients treated with antiandrogens before RP (13).

An analysis published recently to discover new markers of prostate cancer outcome used microarray analysis to classify 21 cancers where the recurrence status at 4 years postprostatectomy was known (15). That study found that no single gene was statistically associated with recurrence at $P < 0.05$, but identified a 5-gene k – nearest neighbor model that achieved 90% classification accuracy on leave-one-out cross-validation. Eleven genes in total appeared in the cross-validated models, which commonly included chromogranin A and IP3R, although none of these 11 probesets overlapped with the ~200 probesets that were significantly associated with relapse in the study presented here.

DISCUSSION

There are a considerable number of publications assessing the ability of biomarkers to predict an earlier time to relapse of prostate cancer after RP (reviewed in Ref. 16). Despite these data, there remain no molecular markers of routine clinical utility that differentiate localized prostate cancers with an aggressive phenotype, and clinicians still rely on conventional preoperative and postoperative prognostic indicators such as pretreatment PSA levels, pathological stage, and Gleason grade in routine decision-making. This most likely reflects the fact that studies that have correlated differences in gene expression with patient outcome have assessed candidate genes with limited predictive power that provide no additional prognostic information above the conventional variables. This accentuates the need to discover novel genes with strong predictive ability.

Here we describe for the first time the application of classical multivariable survival analysis to a prostate cancer microarray dataset incorporating the expression profiles of >46,000 genes to identify markers of disease outcome. This technique provides several significant advances over previous methods of analyses that have been used to discover markers of disease outcome from microarray data. In contrast to statistical methods described previously that rely on the classification of tumors based on known outcome (17) or known classifiers of patient outcome (*e.g.,* estrogen receptor status; Refs. 18, 19), this technique provides for censored data. This enables these analyses to proceed before the occurrence of all of the events, in this case, PSA relapse. Moreover, this survival analysis incorporates the time taken to PSA relapse and may also include covariates (*e.g.,* preoperative serum PSA levels) to identify genes that provide additional predictive value above conventional markers of outcome. The statistical analyses described herein have also incorporated a stringent method of estimating the pFDR that was described recently (10). This method is designed specifically for the analysis of microarray data where general dependence between hypotheses or "clumpy dependence" exists, where ≥50 genes interact in common pathways to produce some overall process (10). However, this is, to our knowledge, the first instance that it has been applied to microarray data from a survival analysis.

There is a discrepancy between our results and those reported by Singh *et al.* (15): in the previous study, no probesets were significantly associated with relapse in univariate analyses, whereas we identified ~200 probesets that predicted prostate cancer relapse, when pretreatment PSA levels were controlled. None of the probesets we identified overlapped with the 11 probesets used by Singh *et al.* (15) in their multivariate classifiers. This discrepancy may be because of several factors. First, we used the custom-designed Eos Hu03 Affymetrix array, rather than the Affymetrix U95Av2 arrays

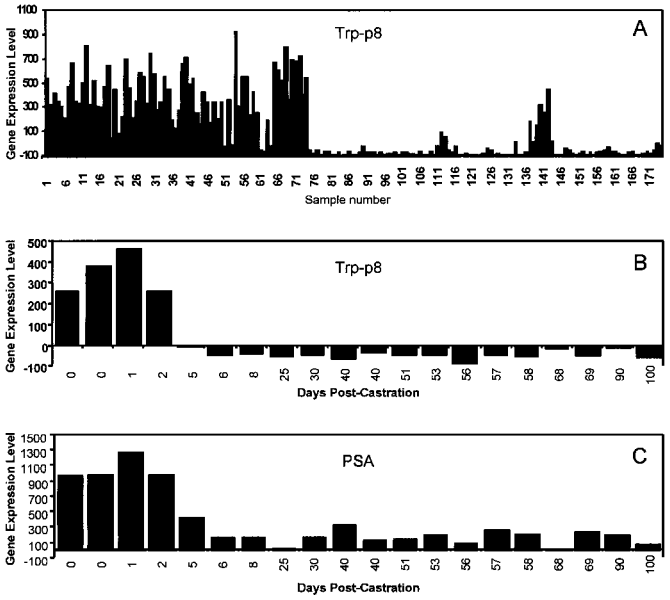

Fig. 3. *A,* expression of *trp-p8* mRNA detected by oligonucleotide microarray in prostate cancer samples and in normal body tissues. Samples are: prostate cancer 1–74, adrenal glands 75–77, aorta 78–80, artery 81–83, bladder 84–86, bone marrow 87–89, colonic epithelium 90–92, cerebral cortex 93–95, colon 96–98, colonic muscle 99–101, esophagus 102–104, heart 105–107, kidney 108–110, liver 111–113, lung 114–116, lymph node 117–119, muscle 120–122, oral mucosa 123–125, pharyngeal mucosa 126–128, pancreas 129–131, parathyroid glands 132–133, pituitary 134–136, prostate 137–143, retina 144–146, skin 147–149, small intestine 150–152, spleen 153–155, stomach 156–158, trachea 159–161, tongue 162–164, ureter 165–167, vagus nerve 168–170, and vein 171–174. B, expression of *trp-p8* mRNA, and *C, PSA* mRNA detected by oligonucleotide microarray in LuCaP-35 tumors at days 0 to 100 after castration. The expression level of *trp-p8* and *PSA* are shown as normalized average intensity units (*Y* axis) of fluorescence signal detected by microarray analysis.

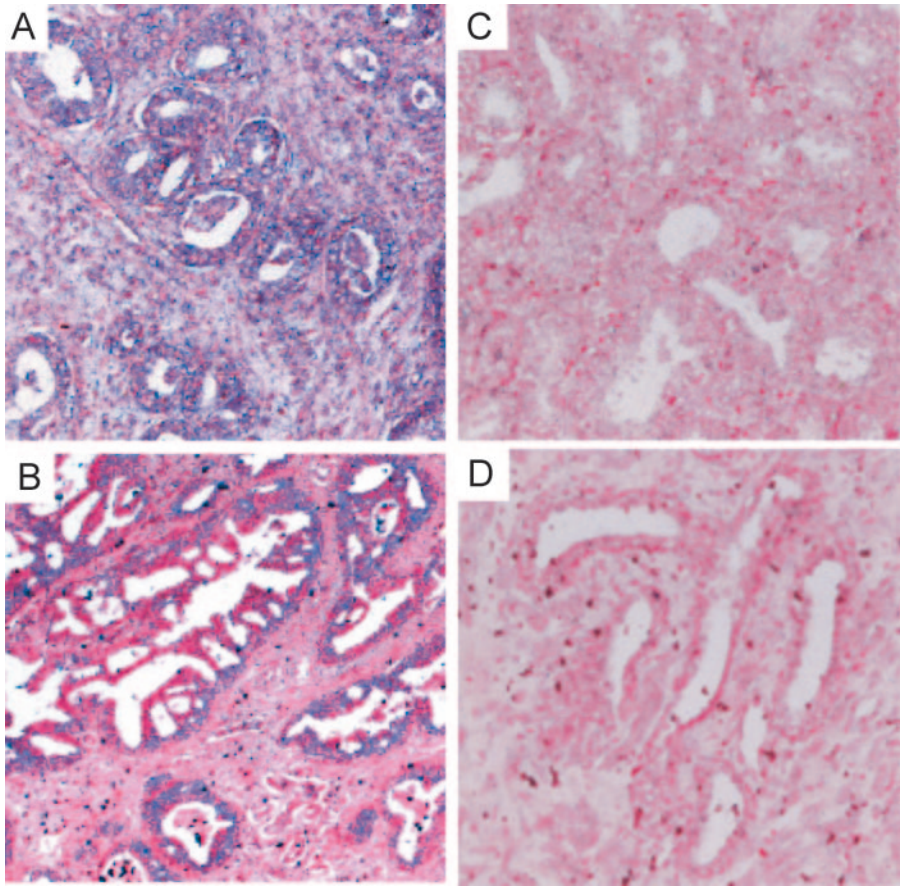

Fig. 4. *Trp-p8* mRNA expression detected by ISH in RP cases treated with or without NHT before surgery. *A,* a prostate cancer from a patient treated with RP only showing positive *trp-p8* mRNA expression in malignant prostate epithelium. *B,* a prostate cancer from a patient treated with RP and NHT showing positive *trp-p8* mRNA expression. *C,* a prostate cancer from a patient treated with RP only with no detectable *trp-p8* mRNA expression in the malignant epithelium, and *D,* a prostate cancer from a patient treated with preoperative NHT with no evidence of *trp-p8* expression.

used in the previous study. The Eos Hu03 arrays have approximately five times as many probesets as the U95Av2 arrays; *trp-p8* is an example of a gene that was not present in the oligonucleotide array used in the previous study. Also, even the probesets thought to measure the same gene on each array are often substantially different, in terms of the specific probe sequences and also the number of probes representing each probeset. Second, by using a statistical method that applies to censored data, we were able to take into account the various times to prostate cancer relapse in this model. Therefore, we were able to use our full data set in the analysis, rather than restricting the analysis to those patients with a specified length of follow-up. The larger data set and concomitant increase in statistical power may also contribute to our results differing from those of Singh *et al.* (15).

We have noted with interest the potential functional link between trp-p8 and calnexin identified in this study, and chromogranin A and IP3R identified by Singh *et al.* (15), because TRP channels are linked to the phosphatidylinositol signal transduction pathway (20). The TRP channels are subunits with six membrane-spanning domains with both COOH and NH_2 termini located intracellularly that probably form into tetramers to form nonselective cationic channels, which allow for the influx of Ca^{2+} ions into the cell. Trp-p8 or TRPM8 is a member of the TRPM (Transient Receptor Potential, Melastatin) subfamily of TRP ion channels that have potential roles in Ca^{2+}-dependent signaling, and control of cell proliferation, cell division, and cell migration (Fig. 5). Ligand binding to some membrane receptors initiates a sequence of events that lead to the activation of phospholipase C, generating inositol-1,4,5-triphosphate, which opens the intracellular ion channel IP3R and liberates Ca^{2+} from the endoplasmic reticulum. Activation of the TRP channels accompanies this chain of events, allowing the influx of Ca^{2+} ions into the cells, although their activation is not necessarily directly linked to Ca^{2+} depletion from internal stores (20). Chromogranin A, which is contained in large amounts in the secretory granules of some cell types, is believed to complex with the IP3R resulting in a conformational change that promotes inositol-1,4,5-triphosphate binding and subsequent Ca^{2+} release. Serum chromogranin A concentration has also been shown to be associated with poor prognosis after endocrine therapy for prostate cancer (21). Calnexin, also identified in this analysis as a marker of potential prognostic utility ($P = 0.004$), is believed to be a key chaperone involved in the folding, assembly, and oligomerization of newly synthesized IP3R receptors (22). Thus, both studies implicate an important role for this pathway in prostate cancer progression.

Our observation that loss of *trp-p8* is associated with a poor prognosis is also reminiscent of the prognostic role of another of the TRPM subfamily, TRPM1 or melastatin, in melanoma. Down-regulation of melastatin mRNA in primary cutaneous melanoma is a prognostic marker for metastasis in patients with localized melanoma and is independent of conventional clinicopathological predictors of metastases (23). The mechanistic action of the androgen-regulated *trp-p8* in prostate cancer progression remains unclear. Recent studies showed that the rat (24) and mouse (25) orthologues of trp-p8 are functional Ca^{2+} channels that respond to cold stimuli. Although cold is unlikely to be the natural stimulus for trp-p8 in the prostate, the implication that the human trp-p8 protein may be a functional Ca^{2+} channel suggests a role in the regulation of intracellular Ca^{2+} levels with possible effects on cell motility, cell proliferation, and resistance to apoptotic stimuli.

In summary, our analyses have identified a statistically significant group of genes that strongly correlate with prostate cancer relapse and contribute unique information to relapse prediction above preoperative PSA. The annotation and clinical validation of these genes will

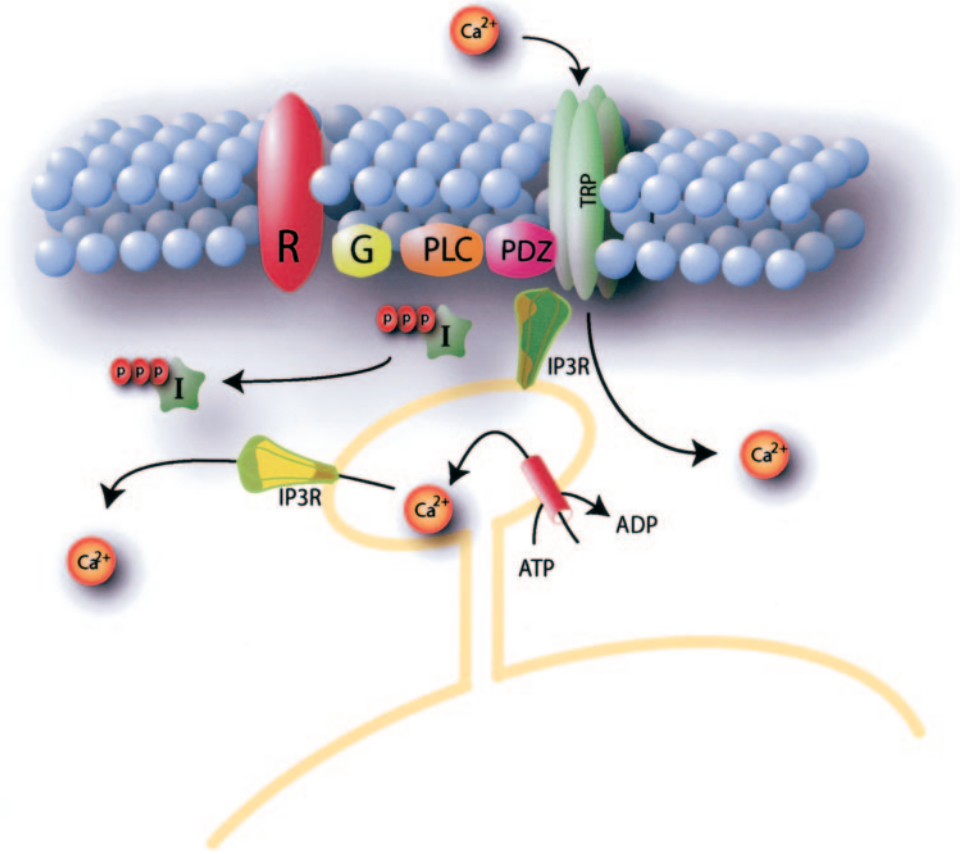

Fig. 5. The phosphatidylinositol signal transduction pathway. Ligand binding to some membrane receptors (*R*), perhaps through a heterotrimeric G protein (*G*), initiates a sequence of events that lead to the activation of phospholipase C (*PLC*), generating inositol-1,4,5-triphosphate (*I*), which opens the intracellular ion channel IP3R and liberates Ca^{2+} from the endoplasmic reticulum. Activation of the TRP channels accompanies this chain of events, allowing the influx of Ca^{2+} into the cells.

likely identify prognostic indicators of clinical utility in identifying prostate cancer patients with significant disease as well as potential targets for therapeutic intervention. Moreover, the functional characterization of genes that cosegregate with phenotype and disease progression is expected to lead to a much deeper understanding of the underlying biology of prostate cancer.

ACKNOWLEDGMENTS

We thank the following members of the St. Vincent's Hospital Campus Prostate Cancer Group for their contributions to this study: Drs. Jennifer Turner and Warick Delprado for the contribution of tissue and pathology expertise; Drs. David Golovsky, Phillip Brenner, Raji Kooner, and Gordon O'Neill for contributing patients; and Sr. Anne-Maree Haynes, Luis Winoto, and Catherine Langusch for their roles in the establishment and maintenance of the prostate cancer clinical database and tissue bank. We also thank Drs. Andrew Biankin and Ghassan Ghandour for critical discussion of this work.

REFERENCES

1. Jemal, A., Thomas, A., Murray, T., and Thum, M. Cancer statistics, 2002. CA Cancer J. Clin., *52:* 23–47, 2002.
2. Snow, P. B., Smith, D. S., and Catalona, W. J. Artificial neural networks in the diagnosis and prognosis of prostate cancer: a pilot study. J. Urol., *152:* 1923–1926, 1994.
3. Kattan, M. W., Eastham, J. A., Stapleton, A. M., Wheeler, T. M., and Scardino, P. T. A preoperative nomogram for disease recurrence following radical prostatectomy for prostate cancer. J. Natl. Cancer Inst., *90:* 766–771, 1998.
4. D'Amico, A. V., Whittington, R., Malkowicz, S. B., Fondurulia, J., Chen, M-H, Kaplan, I., Beard, C. J., Tomaszewski, J. E., Renshaw, A. A., Wein, A., and Coleman, C. N. Pretreatment nomogram for prostate-specific antigen recurrence after radical prostatectomy or external-beam radiation therapy for clinically localised prostate cancer. J. Clin. Oncol., *17:* 168–172, 1999.
5. Graefen, M., Noldus, J., Pichlmeier, A., Haese, P., Hammerer, S., Fernandez, S., Conrad, R., Henke, E., Huland, E., and Huland, H. Early prostate-specific antigen relapse after radical retropubic prostatectomy: prediction on the basis of preoperative and postoperative tumor characteristics. Eur. Urol., *36:* 21–30, 1999.
6. Kattan, M. W., Wheeler, T. M., and Scardino, P. T. Postoperative nomogram for disease recurrence after radical prostatectomy for prostate cancer. J. Clin. Oncol., *17:* 1499–1507, 1999.
7. Linja, M. J., Savinainen, K. J., Saramaki, O. R., Tammela, T. L. J., Vessella, R. L., and Visakorpi, T. Amplification and overexpression of androgen receptor gene in hormone-refractory prostate cancer. Cancer Res., *61:* 3550–3555, 2001.
8. Tukey, J. W. Exploratory Data Analysis. Reading, Massachusetts: Addison-Wesley, 1977.
9. Cox, D. R. Regression models and life tables (life tables). J. R. Stat. Soc., *34:* 187–189, 1972.
10. Storey, J. D., and Tibshirani, R. Estimating false discovery rates under dependence, with applications to DNA microarrays. Technical Report, Department of Statistics, Stanford University, CA, 2001.
11. Ward, J. H. Hierarchical grouping to optimize an objective function. J. Am. Stat. Assoc, *58:* 236–244, 1963.
12. Kononen, J., Bubendorf, L., Kallioniemi, A., Bårlund, M., Schraml, P., Leighton, S., Torhorst, J., Mihatsch, M. J., Sauter, G, and Kallioniemi, O-P. Tissue microarrays for high-throughput molecular profiling of tumor specimens. Nat. Med., *4:* 844–847, 1998.
13. Quinn, D. I., Henshall, S. M., Haynes, A-M, Brenner, P. C., Kooner, R., Golovsky, D., Mathews, J., O'Neil, G. F., Turner, J. J., Delprado, W., Finlayson, J. F., Sutherland, R. L., Grygiel, J. J., and Stricker, P. D. Prognostic significance of pathological features in localized prostate cancer treated with radical prostatectomy: implications for staging systems and predictive models. J. Clin. Oncol., *19:* 3692–3705, 2001.
14. Tsavaler, L., Shapero, M. H., Morkowski, S., and Laus, R. Trp-p8, a novel prostate-specific gene, is up-regulated in prostate cancer and other malignancies and shares high homology with transient receptor potential calcium channel proteins. Cancer Res., *61:* 3760–3769, 2001.
15. Singh, D., Febbo, P. G., Ross, K., Jackson, D. G., Manola, J., Ladd, C., Tamayo, P., Renshaw, A., D'Amico, A. V., Richie, J. P., Lander, E. S., Loda, M., Kantoff, P. W., Golub, T. R., and Sellers, W. R. Gene expression correlates of clinical prostate cancer behavior. Cancer Cell, *1:* 203–209, 2002.
16. Hamdy, F. C. Prognostic and predictive factors in prostate cancer. Cancer Treat. Rev., *27:* 143–151, 2001.
17. van 't Veer, L. J., Dai, H., van de Vijver, M. J., He, Y. D., Hart, A. A. M., Mao, M., Peterse, H. L., van der Kooy, K., Marton, M. J., Witteveen, A. T., Schreiber, G. J., Kerkhoven, R. M., Roberts, C., Linsley, P. S., Bernards, R., and Friend, S. H. Gene expression profiling predicts clinical outcome of breast cancer. Nature (Lond.), *415:* 530–536, 2002.

18. Perou, C. M., Sfrlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., Fluge, O., Perganrnenschikov, A., Williams, C., Zhu, S. X., Lenning, P. E., Borresen-Dale, A-L, Brown, P. O., and Botstein, D. Molecular portraits of human breast tumours. Nature (Lond.), *406:* 747–752, 2000.
19. Gruvberger, S., Ringnér, M., Chen, Y., Panavally, S., Saal, L. H., Borg, A., Ferno, M., Peterson, C., and Meltzer, P. S. Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns. Cancer Res., *61:* 5979–5984, 2001.
20. Clapham, D. E., Runnels, L. W., and Strubing, C. The trp ion channel family. Nat. Rev. Neurosci., *2:* 387–396, 2001.
21. Isshiki, S., Akakura, K., Komiya, A., Suzuki, H., Kamiya, N., and Ito, H. Chromogranin A concentration as a serum marker to predict prognosis after endocrine therapy for prostate cancer. J. Urol., *167:* 512–515, 2002.
22. Joseph, S. K., Boehning, D., Bokkala, S., Watkins, R., and Widjaja, J. Biosynthesis of inositol triphosphate receptors: selective association with the molecular chaperone calnexin. Biochem. J., *342:* 153–161, 1999.
23. Duncan, L. M., Deeds, J., Cronin, F. E., Donovan, M., Sober, A. J., Kauffman, M., and McCarthy, J. Melastatin expression and prognosis in cutaneous malignant melanoma. J. Clin. Oncol., *19:* 568–576, 2001.
24. McKemy, D. D., Neuhausser, W. N., and Julius, D. Identification of a cold receptor reveals a general role for TRP channels in thermosensation. Nature (Lond.), *416:* 52–58, 2002.
25. Peier, A. M., Moqrich, A., Hergarden, A. C., Reeve, A. J., Andersson, D. A., Story, G. M., Earley, T. J., Dragoni, I., McIntyre, P., Bevan, S., and Patapoutian, A. A TRP channel that senses cold stimuli and menthol. Cell, *108:* 705–715, 2002.