SUPPLEMENTARY DATA

Gene expression profiling predicts clinical outcome of prostate cancer

Gennadi V. Glinsky, Anna B. Glinskii, Andrew J. Stephenson, Robert M. Hoffman, William L. Gerald

Table of Contents

1. Protocol of discovery and validation of the prostate cancer recurrence predictor algorithm. We hypothesized that clinically relevant genetic signatures could be found by searching for clusters of co-regulated genes that display highly concordant transcript abundance behavior across multiple experimental models and clinical settings which are modeling or representing malignant phenotypes of interest (1-3). Thus, according to this model the primary criterion in transcript selection process should be the concordance of changes in expression rather than a magnitude of changes (e.g., fold change). One of the predictions of this model is that transcripts of interest would be expected to have a tightly controlled "rank order" of expression within a cluster of co-regulated genes reflecting a balance of up- and down-regulated mRNAs as a desired regulatory end-point in a cell. A degree of resemblance of the transcript abundance rank order within a gene cluster between a test sample and reference standard is measured by a Pearson correlation coefficient and designated as a phenotype association index (PAI). To identify genes with consistently concordant expression patterns across multiple data sets and various experimental conditions, we compared the expression profile of 218 genes (test samples) to the expression profiles of transcripts differentially regulated in multiple experimental models (reference standard) of human prostate cancer (1).

The transcripts comprising each signature were selected based on Pearson correlation coefficients ($r > 0.95$) reflecting a degree of similarity of expression profiles in clinical tumor samples (recurrent versus non-recurrent tumors) and experimental samples using the following protocol.

Step 1. Sets of differentially regulated transcripts were independently identified for each experimental conditions (see below) and clinical samples using the Affymetrix microarray processing and statistical analysis software package as described in Materials and Methods.

Step 2. Sub-sets of transcripts exhibiting concordant expression changes in clinical and experimental samples were identified using the Affymetrix MicroDB and DMT software. Sub-sets of transcripts were identified with concordant changes of transcript abundance behavior in recurrent versus non-recurrent clinical tumor samples (218 transcripts) and experimental conditions independently defined for each signature (Signature 1: PC-3MLN4 orthotopic versus s.c. xenografts; Signature 2: PC-3MLN4 versus PC-3M & PC-3 orthotopic xenografts; Signature 3: PC-3/LNCap consensus class, Ref. 1). Thus, from a set of 218 transcripts three concordant sub-sets of transcripts were identified corresponding to each binary comparison of clinical and experimental samples Table 4S, Supplement).

Step 3. Selection of small gene clusters was performed from sub-sets of genes exhibiting concordant changes of transcript abundance behavior in recurrent versus non-recurrent clinical tumor samples (218 transcripts) and experimental conditions defined for each signature (Signature 1: PC-3MLN4 orthotopic versus s.c. xenografts; Signature 2: PC-3MLN4 versus PC-3M & PC-3 orthotopic xenografts; Signature 3: PC-3/LNCap consensus class, Ref. 1). Expression profiles were presented as Log10 average fold changes for each transcript and processed for visualization and Pearson correlation analysis using Microsoft Excel software. Cut-off criterion for cluster selection for evaluation in Step 4 was set to exceed a Pearson correlation coefficient 0.95.

Step 4. Identified small gene clusters exhibiting highly concordant pattern of expression (Pearson correlation coefficient, $r > 0.95$) in clinical and experimental settings were evaluated for their ability to discriminate clinical samples with distinct outcome after the therapy. To assess a potential prognostic relevance of individual gene clusters, we calculated a Pearson correlation coefficient for each of 21 tumor samples (training data set) by comparing the expression profiles of individual samples to the reference expression profiles of relevant experimental samples defined for each signature and an "average" expression profile of recurrent versus non-recurrent tumors. Fold expression changes in the "average" expression profile of recurrent versus non-recurrent tumors were calculated for each gene as a ratio of the "average" expression value of a gene in recurrent tumors (8 samples in training set) to the "average" expression value in non-recurrent tumors (13 samples in training set). Fold expression changes in individual clinical samples were calculated for each gene as a ratio of the expression value in a given sample to the "average" expression value of the gene across the entire data set of 21 samples. Based on expected correlation of expression profiles of identified gene clusters with recurrent clinical behavior of prostate cancer, we named the corresponding correlation coefficients calculated for individual samples the phenotype association indices (PAIs). We evaluated the prognostic power of identified clusters of co-regulated transcripts based on their ability to segregate the patients with recurrent and non-recurrent prostate tumors into distinct sub-groups and selected a single best performing cluster for each binary conditions specified in the Table 4S, Supplement (Figure 1; Tables 1 & 2).

Step 5. We used the Kaplan-Meier survival analysis to assess the prognostic power of each best performing cluster in predicting the probability that patients would

remain disease-free after therapy (Figure 2). We selected the prognosis discrimination cut-off value for each signature based on highest level of statistical significance in patient's stratification into poor and good prognosis groups as determined by the log-rank test (lowest P value and highest hazard ratio; Table 2 & Figure 2; Table 6S, Supplement). Clinical samples having the Pearson correlation coefficient at or higher the cut-off value were identified as having the poor prognosis signature. Clinical samples with the Pearson correlation coefficient lower the cut-off value were identified as having the good prognosis signature.

Step 6. We developed a prostate cancer recurrence predictor algorithm taking into account calls from all three individual signatures. We selected the common prognosis discrimination cut-off value for all three signatures based on highest level of statistical significance in patient's stratification into poor and good prognosis groups as determined by the Kaplan-Meier survival analysis (lowest P value and highest hazard ratio defined by the log-rank test; Table 2 & Figure 2). Clinical samples having the Pearson correlation coefficient at or higher the cut-off value defined by at least two signatures were identified as having the poor prognosis signature. Clinical samples with the Pearson correlation coefficient lower the cut-off value defined by at least two signatures were identified as having the good prognosis signature. We found that the cut-off value of PAIs > 0.2 scored in two of three individual clusters allowed to achieve the 90% recurrence prediction accuracy (Table 2; Figure 2C).

Step 7. We validated the prognostic power of prostate cancer recurrence predictor algorithm alone and in combination with the established markers of outcome using an independent clinical set of 79 prostate cancer patients (Figures 3-6; Tables 3 & 4). Fold

expression changes in the “average” expression profile of recurrent versus non-recurrent tumors were calculated for each gene as a ratio of the “average” expression value of a gene in recurrent tumors (37 samples in validation set) to the “average” expression value in non-recurrent tumors (42 samples in validation set). Fold expression changes in individual clinical samples were calculated for each gene as a ratio of the expression value in a given sample to the “average” expression value of the gene across the entire data set of 79 samples.

References

1. Glinsky, G.V., Krones-Herzig, A., Glinskii, A.B., Gebauer, G. 2003. Microarray analysis of xenograft-derived cancer cell lines representing multiple experimental models of human prostate cancer. Molecular Carcinogenesis. **37:** 209-221.
2. Glinsky, G.V., Krones-Herzig, A., Glinskii, A.B. 2003. Malignancy-associated regions of transcriptional activation: gene expression profiling identifies common chromosomal regions of a recurrent transcriptional activation in human prostate, breast, ovarian, and colon cancers. Neoplasia. **5:** 21-228.
3. Glinsky, G.V., Ivanova, Y.A., Glinskii, A.B. 2003. Common malignancy-associated regions of transcriptional activation (MARTA) in human prostate, breast, ovarian, and colon cancers are targets for DNA amplification. Cancer Letters. **201:** 67-77.

2. Description of the prostate cancer recurrence predictor validation data set

Tissue samples were obtained from 79 patients (37 with recurrent and 42 with non-recurrent prostate cancer) who had undergone radical prostatectomy for clinically localized prostate cancer between 1993 and 1999 (Table S1). All patients had negative lymph nodes on final pathological evaluation and no patient received any neoadjuvant or adjuvant therapy before documented disease recurrence. Disease recurrence was defined as 3 consecutive increases in the level of PSA. All non-recurrent patients had maintained an undetectable PSA for a minimum of 5 years after radical prostatectomy.

The patients in our cohort do not represent consecutive patients with prostate cancer treated by radical prostatectomy at our institution between 1993 and 1999. Rather, we attempted to obtain tissue from an equal number of recurrent and non-recurrent patients for the purpose of analyzing gene expression differences between these two classes. As a result, the rate of positive surgical margins (63%), extracapsular extension (56%), and seminal vesicle invasion (13%) is higher than that reported in large radical prostatectomy series. Likewise, the median PSA level (7.6 ng/mL) is significantly higher than that reported in large radical prostatectomy series.

3. Supplementary Tables

Table 1S. Clinical and pathological characteristics of 79 patients.

Table 2S. Human prostate carcinoma cell lines and xenografts derived from androgen-dependent (LNCap) and androgen-independent (PC3) lineages through serial orthotopic re-implantation and recovery from primary and metastatic tumors in nude mice.

Table 3S. 218 genes differentially regulated in 8 recurrent versus 13 non-recurrent human prostate tumors.

Table 4S. Prostate cancer recurrence predictor signatures and overall classification accuracy in good-prognosis and poor-prognosis sub-groups of patients defined according to whether they had a good-prognosis or a poor-prognosis signature.

Table 5S. Expression profiles of genes comprising prostate cancer recurrence predictor signatures.

Table 6S. Phenotype association indices for individual tumor samples comprising 21-sample clinical set utilized for discovery of the prostate cancer recurrence predictor algorithm.

Table 7S. Cox multivariate proportional hazard analysis.

Table 1S. Clinical and pathological characteristics of 79 patients

	Number	Percent
Age (years)		
< 50	5	6%
50 - 60	30	38%
> 60	44	56%
Biochemical relapse		
Yes	37	47%
No	42	53%
Tumor stage (1992 TNM)		
T1C	34	43%
T2A	16	20%
T2B	20	25%
T2C	7	9%
T3A	2	3%
RP Gleason Sum		
4	1	1%
5	1	1%
6	15	19%
7	44	56%
8	10	13%
9	8	10%
Capsular invasion		
None	17	22%
Focal	6	8%
Invasive	18	23%
Established	38	48%
Surgical margins		
Negative	29	37%
Positive	50	63%
Seminal vesicle invasion		
Negative	69	87%
Positive	10	13%
Lymphe node		
Negative	76	96%
Positive	3	4%
Pre-RP PSA		
< 5.0	18	23%
5.0 - 10.0	31	39%
> 10.0	30	38%

RP, radical prostatectomy; PSA, prostate specific antigen; Median follow-up = 70 months

Table 2S. Human prostate carcinoma cell lines and xenografts derived from androgen-dependent (LNCap) and androgen-independent (PC3) lineages through serial orthotopic re-implantation and recovery from primary and metastatic tumors in nude mice. RNA from all conditions was prepared at least twice from independent experiments to assure reproducibility.

Cell Lines	Cycles of progression	Site of transplantation / recovery	Orthotopic tumorigenicity	Metastatic potential	RNA sources used
Normal Epithelia[1]	0	None	None	None	In vitro
PC3	0	None	High	Intermediate	In vitro, in vivo
PC3M	1	Prostate / liver	High	High	In vitro, in vivo
PC3M-LN4	4	Prostate / lymph nodes	High	Very high	In vitro, in vivo
PC3M-Pro4	4	Prostate / prostate	High	Intermediate	In vitro
LNCap	0	None	Intermediate	Low	In vitro
LNCap-LN3	3	Prostate / lymph nodes	High	High	In vitro
LNCap-Pro5	5	Prostate / prostate	High	Low	In vitro

[1]Two primary normal human prostate epithelial cell lines (normal epithelia) were obtained from Clonetics / BioWhittaker (San Diego, CA) and grown in complete prostate epithelial growth media provided by the supplier.

Table 3S. 218 genes differentially regulated in 8 recurrent versus 13 non-recurrent human prostate tumors

Affymetrix Probe Set	P value - T-Test	T-Test_Change	P value - MW-Test	MW_Change Direction
40642_at	0	Down	0.001	Down
1135_at	0.001	Down	0.007	Down
39748_at	0.001	Down	0.002	Down
37343_at	0.001	Down	0.007	Down
37806_at	0.001	Down	0.009	Down
41352_at	0.001	Down	0.004	Down
31881_at	0.002	Up	0.006	Up
34413_at	0.002	Down	0.005	Down
39671_at	0.003	Up	0.002	Up
31577_at	0.003	Up	0.004	Up
33922_at	0.003	Up	0.003	Up
37828_at	0.003	Up	0.006	Up
40130_at	0.003	Down	0.009	Down
40328_at	0.003	Down	0.002	Down
160027_s_at	0.004	Down	0.004	Down
38994_at	0.004	Down	0.011	Down
1124_at	0.004	Down	0.007	Down
36234_at	0.004	Down	0.006	Down
33431_at	0.004	Down	0.005	Down
36732_at	0.005	Down	0.005	Down
33306_at	0.005	Down	0.011	Down
36634_at	0.005	Down	0.017	Down
32786_at	0.005	Down	0.014	Down
41868_at	0.005	Down	0.002	Down
37630_at	0.005	Down	0.017	Down
35703_at	0.006	Down	0.006	Down
32502_at	0.006	Down	0.009	Down
36422_s_at	0.006	Down	0.036	Down
265_s_at	0.006	Down	0.036	Down
36203_at	0.006	Down	0.014	Down
35834_at	0.006	Down	0.014	Down
38575_at	0.007	Down	0.007	Down
39510_r_at	0.007	Down	0.03	Down
773_at	0.007	Down	0.025	Down
35249_at	0.008	Up	0.006	Up

38312_at	0.008	Up	0.011	Up
38774_at	0.008	Up	0.006	Up
35320_at	0.008	Down	0.005	Down
32563_at	0.008	Down	0.03	Down
160033_s_at	0.008	Down	0.006	Down
39733_at	0.008	Down	0.014	Down
32109_at	0.008	Down	0.02	Down
32855_at	0.008	Down	0.02	Down
40448_at	0.008	Down	0.036	Down
32870_g_at	0.009	Up	0.005	Up
36160_s_at	0.009	Down	0.03	Down
39253_s_at	0.009	Down	0.012	Down
32672_at	0.009	Down	0.012	Down
36711_at	0.009	Down	0.006	Down
41448_at	0.01	Up	0.005	Up
459_s_at	0.01	Down	0.006	Down
41120_at	0.01	Down	0.009	Down
31941_s_at	0.01	Down	0.014	Down
34300_at	0.01	Down	0.011	Down
32785_at	0.01	Down	0.033	Down
770_at	0.011	Down	0.03	Down
32907_at	0.011	Down	0.03	Down
39631_at	0.011	Down	0.017	Down
1915_s_at	0.011	Down	0.025	Down
33461_at	0.012	Up	0.002	Up
39648_at	0.012	Up	0.017	Up
41062_at	0.012	Up	0.008	Up
40077_at	0.012	Down	0.025	Down
33308_at	0.012	Down	0.043	Down
37393_at	0.012	Down	0.036	Down
37854_at	0.013	Up	0.004	Up
33228_g_at	0.013	Down	0.007	Down
131_at	0.013	Down	0.036	Down
38291_at	0.013	Down	0.025	Down
1081_at	0.013	Down	0.03	Down
984_g_at	0.014	Up	0.002	Up
33886_at	0.014	Down	0.02	Down

33436_at	0.014	Down	0.025	Down
37633_s_at	0.014	Down	0.017	Down
35019_at	0.015	Up	0.009	Up
41670_at	0.015	Up	0.014	Up
35256_at	0.015	Up	0.009	Up
38985_at	0.015	Up	0.009	Up
33304_at	0.015	Down	0.043	Down
35775_at	0.016	Up	0.011	Up
35557_at	0.016	Up	0.025	Up
35653_at	0.016	Down	0.014	Down
752_s_at	0.016	Down	0.017	Down
1934_s_at	0.017	Up	0.007	Up
35689_at	0.017	Up	0.014	Up
39702_at	0.017	Up	0.014	Up
35720_at	0.017	Up	0.006	Up
33374_at	0.017	Down	0.036	Down
36833_at	0.017	Down	0.025	Down
1622_at	0.017	Down	0.025	Down
2094_s_at	0.017	Down	0.02	Down
509_at	0.018	Down	0.036	Down
37136_at	0.018	Down	0.043	Down
1058_at	0.018	Down	0.036	Down
35649_at	0.018	Down	0.036	Down
34671_at	0.018	Down	0.014	Down
41536_at	0.018	Down	0.043	Down
35608_at	0.019	Up	0.017	Up
41411_at	0.019	Down	0.025	Down
39989_at	0.019	Down	0.025	Down
39385_at	0.019	Down	0.043	Down
33049_at	0.02	Up	0.002	Up
34676_at	0.02	Down	0.03	Down
31808_at	0.02	Down	0.039	Down
1194_g_at	0.021	Up	0.009	Up
39610_at	0.021	Up	0.014	Up
32589_at	0.021	Up	0.011	Up
40569_at	0.021	Down	0.02	Down
36127_g_at	0.021	Down	0.009	Down

41229_at	0.021	Down	0.017	Down
1662_r_at	0.021	Down	0.043	Down
41106_at	0.021	Down	0.043	Down
1126_s_at	0.021	Down	0.043	Down
287_at	0.021	Down	0.025	Down
38862_at	0.022	Up	0.006	Up
765_s_at	0.022	Up	0.015	Up
41343_at	0.022	Up	0.03	Up
33901_at	0.022	Up	0.03	Up
41585_at	0.022	Down	0.017	Down
41421_at	0.022	Down	0.03	Down
33429_at	0.022	Down	0.025	Down
36681_at	0.022	Down	0.03	Down
34732_at	0.022	Down	0.017	Down
40095_at	0.023	Up	0.011	Up
40674_s_at	0.023	Up	0.004	Up
32305_at	0.023	Up	0.036	Up
36456_at	0.023	Up	0.03	Up
33596_at	0.023	Down	0.036	Down
2049_s_at	0.023	Down	0.03	Down
31751_f_at	0.024	Up	0.011	Up
34211_at	0.024	Up	0.017	Up
35039_at	0.024	Up	0.009	Up
37888_at	0.024	Down	0.03	Down
35729_at	0.024	Down	0.017	Down
280_g_at	0.024	Down	0.03	Down
39275_at	0.024	Down	0.043	Down
35698_at	0.025	Up	0.02	Up
41804_at	0.025	Up	0.02	Up
38452_at	0.026	Up	0.036	Up
38471_r_at	0.026	Up	0.03	Up
2086_s_at	0.027	Down	0.043	Down
38383_at	0.027	Down	0.025	Down
1565_s_at	0.028	Up	0.017	Up
32480_at	0.028	Up	0.014	Up
37552_at	0.028	Up	0.036	Up
37906_at	0.028	Up	0.017	Up

33916_at	0.028	Down	0.02	Down
37026_at	0.028	Down	0.036	Down
40503_at	0.028	Down	0.043	Down
39204_at	0.028	Down	0.017	Down
35065_at	0.029	Up	0.036	Up
34545_at	0.029	Up	0.017	Up
39219_at	0.029	Up	0.017	Up
41183_at	0.029	Up	0.025	Up
1612_s_at	0.029	Down	0.02	Down
1458_at	0.029	Down	0.02	Down
35253_at	0.03	Up	0.025	Up
36860_at	0.03	Down	0.009	Down
36097_at	0.03	Down	0.025	Down
40935_at	0.031	Up	0.007	Up
39280_at	0.031	Down	0.043	Down
35009_at	0.032	Up	0.036	Up
33548_f_at	0.033	Up	0.014	Up
41369_at	0.033	Up	0.017	Up
32970_f_at	0.033	Up	0.02	Up
34694_at	0.033	Up	0.025	Up
1237_at	0.033	Down	0.02	Down
38371_at	0.034	Up	0.025	Up
35934_at	0.035	Up	0.017	Up
31533_s_at	0.035	Up	0.014	Up
31591_s_at	0.035	Up	0.025	Up
31807_at	0.035	Up	0.017	Up
35968_s_at	0.035	Down	0.036	Down
40071_at	0.036	Down	0.043	Down
1099_s_at	0.036	Down	0.036	Down
40301_at	0.036	Down	0.017	Down
1175_s_at	0.037	Up	0.036	Up
41105_s_at	0.037	Up	0.007	Up
40715_at	0.037	Up	0.011	Up
32778_at	0.037	Down	0.036	Down
33760_at	0.037	Down	0.043	Down
39995_s_at	0.038	Up	0.043	Up
41666_at	0.038	Down	0.025	Down

38506_at	0.038	Down	0.036	Down
794_at	0.039	Up	0.03	Up
41313_at	0.039	Down	0.043	Down
658_at	0.04	Up	0.036	Up
37707_i_at	0.04	Down	0.02	Down
37282_at	0.04	Down	0.043	Down
35786_at	0.041	Down	0.043	Down
32083_at	0.042	Up	0.014	Up
33948_at	0.042	Up	0.017	Up
38322_at	0.042	Down	0.03	Down
659_g_at	0.043	Up	0.025	Up
205_g_at	0.043	Up	0.02	Up
32606_at	0.043	Up	0.006	Up
31383_at	0.044	Up	0.025	Up
572_at	0.044	Up	0.036	Up
37830_at	0.044	Up	0.025	Up
33029_at	0.044	Up	0.017	Up
32638_s_at	0.045	Up	0.017	Up
35050_at	0.045	Up	0.017	Up
40113_at	0.045	Up	0.017	Up
31862_at	0.045	Up	0.03	Up
213_at	0.045	Down	0.043	Down
1435_f_at	0.046	Up	0.03	Up
1977_s_at	0.046	Up	0.036	Up
39092_at	0.046	Up	0.043	Up
31600_s_at	0.046	Up	0.025	Up
1234_at	0.047	Up	0.011	Up
34307_at	0.047	Down	0.03	Down
135_g_at	0.048	Up	0.036	Up
1650_g_at	0.048	Down	0.043	Down
988_at	0.048	Down	0.03	Down
37784_at	0.049	Up	0.036	Up
40200_at	0.049	Down	0.043	Down
40522_at	0.049	Down	0.036	Down
510_g_at	0.049	Down	0.036	Down

Table 4S. Prostate cancer recurrence predictor signatures and overall classification accuracy in good-prognosis and poor-prognosis sub-groups of patients defined according to whether they had a good-prognosis or a poor-prognosis signature.

Recurrence signature	Correlation coefficient	Experimental Setting	Clinical Setting	Overall Classification Performance	P value (Logrank test)
Signature 1	r = 0.999	PC-3MLN4 Orhtotopic vs. PC-3MLN4 sub-cutaneous xenografts	8 recurrent vs. 13 non-recurrent tumors	95% (20 of 21)	< 0.0001
Signature 2	r = 0.963	PC-3MLN4 Orhtotopic vs. PC-3M & PC-3 orthotopic xenografts	8 recurrent vs. 13 non-recurrent tumors	90% (19 of 21)	< 0.0001
Signature 3	r = 0.996	5 xenograft-derived cell lines vs. NPE in vitro (PC-3/LNCap consensus class)	8 recurrent vs. 13 non-recurrent tumors	86% (18 of 21)	0.001
Algorithm	NA	All three signatures	8 recurrent vs. 13 non-recurrent tumors	90% (19 of 21)	< 0.0001

Legend: 21 prostate cancer patients who provided tumor samples comprising a signature discovery (training) data set were classified according to whether they had a good-prognosis signature or poor-prognosis signature based on PAI values defined by either individual recurrence predictor signatures or recurrence predictor algorithm which is taking into account calls from all three signatures. Correlation coefficients reflect a degree of similarity of expression profiles in clinical setting (recurrent versus non-recurrent tumors) and experimental swettings (Signature 1: PC-3MLN4 orthotopic versus PC-3MLN4 s.c. xenografts; Signature 2: PC-3MLN4 orthotopic versus PC-3M & PC-3 orthotopic xenografts; Signature 3: PC-3/LNCap consensus class, Ref. 19). The number of correct predictions in poor-prognosis and good-prognosis groups is shown as a fraction of patients with the observed clinical outcome after therapy (8 patients developed relapse and 13 patients remained disease-free). P values were calculated with use of the log-rank test and reflect the statistically significant difference in the probability that patients would remain disease-free between poor-prognosis and good-prognosis sub-groups. NPE, primary normal human prostate epithelial cells.

Table 5S. Expression profiles of genes comprising prostate cancer recurrence predictor signatures

Signature 1	PC-3MLN4 orthotopic vs. sub-cutaneous xenografts	Clinical Samples, Recurrent vs. Non-recurrent tumors	
Gene Name	Log10 Fold Expression Changes	Log10 Fold Expression Changes	Correlation Coefficient
MGC5466	0.414589187	0.361872	0.99867454
Wnt5A	0.212352681	0.217576	
KIAA0476	-0.184524427	-0.12741	
ITPR1	-0.23858992	-0.18525	
TCF2	-0.344382734	-0.29267	
Signature 2	PC-3MLN4 orthotopic vs. PC-3&PC-3M orthotopic xenografts	Clinical Samples, Recurrent vs. Non-recurrent tumors	
Gene Name	Log10 Fold Expression Changes	Log10 Fold Expression Changes	Correlation Coefficient
MGC5466	0.361872	0.187749	0.963336
CHAF1A	0.232818	0.090371	
CDS2	0.172482	0.144277	
IER3	-0.20069	-0.12422	
Signature 3	Five PC-3&LNCap xenograft-derived cell lines vs. two NPE cell lines	Clinical Samples, Recurrent vs. Non-recurrent tumors	
Gene Name	Log10 Fold Expression Changes	Log10 Fold Expression Changes	Correlation Coefficient
PPFIA3	1.083503	0.153976	0.995802
COPEB	-0.6184	-0.2577	
FOS	-0.69839	-0.33464	
JUNB	-0.8278	-0.33492	
ZFP36	-1.04922	-0.38858	

Legend: The prostate tumor samples from 21 prostate cancer patients comprising a signature discovery (training) data set as well as xenografts and xenograft-derived cell lines were subjected to a microarray gene expression profiling analysis as decsribed in the Materials and Methods. Correlation coefficients reflect a degree of similarity of expression profiles in clinical setting (recurrent versus non-recurrent tumors) and experimental settings (Signature 1: PC-3MLN4 orthotopic versus PC-3MLN4 s.c. xenografts; Signature 2: PC-3MLN4 orthotopic versus PC-3M & PC-3 orthotopic xenografts; Signature 3: PC-3/LNCap consensus class, Ref. 19). The expression profile in clinical samples is presented as Log10 Fold expression changes of average gene expression value in recurrent vs. non-recurrent tumors. NPE, primary normal human prostate epithelial cells.

Table 6S. Phenotype association indices for individual tumor samples

Sample	Signature 1	Signature 2	Signature 3	Recurrence	DFI
T59	**0.920965328**	**0.62283823**	**0.976031988**	1	26
T04	**0.891655793**	**0.78388076**	**0.890977963**	1	46
T26	**0.885410025**	**0.953225**	**0.853505414**	1	14
T33	**0.794543542**	**0.87417819**	**0.509509129**	0	15
T57	**0.710948019**	**0.69165789**	**0.884248877**	1	4
T17	**0.652516655**	**0.9204434**	**0.912978554**	1	3
T62	**0.576621536**	**0.87910309**	**-0.922608457**	1	30
T23	**0.190439806**	**0.8562719**	**0.712833807**	1	37
T45	**0.062434855**	**-0.57024398**	**-0.353301599**	1	6
T46	**-0.037847471**	**-0.20100716**	**-0.326285749**	0	57
T01	**-0.151587251**	**0.57794677**	**-0.931739574**	0	55
T25	**-0.353643694**	**-0.50177771**	**-0.729354419**	0	52
T22	**-0.365270553**	**-0.13336072**	**-0.838207571**	0	54
T54	**-0.386411713**	**-0.8240244**	**-0.938184703**	0	51
T55	**-0.46357102**	**-0.89911482**	**-0.964914426**	0	66
T10	**-0.552811926**	**-0.2570071**	**-0.381668168**	0	50
T24	**-0.56194093**	**0.2618008**	**-0.370198423**	0	54
T13	**-0.643420256**	**-0.33145908**	**0.818307891**	0	54
T29	**-0.783661162**	**-0.77967421**	**-0.674430912**	0	51
T60	**-0.870481405**	**-0.39606961**	**-0.969165663**	0	55
T16	**-0.910897024**	**-0.49816985**	**-0.838207571**	0	49

Algorithm: 0.2 cut-off for individual indices & 2 out of 3 positive

Sample	Signature 1	Recurrence	DFI
T59	**0.920965328**	1	26
T04	**0.891655793**	1	46
T26	**0.885410025**	1	14
T33	**0.794543542**	0	15
T57	**0.710948019**	1	4
T17	**0.652516655**	1	3
T62	**0.576621536**	1	30
T23	**0.190439806**	1	37
T45	**0.062434855**	1	6
T46	**-0.037847471**	0	57
T01	**-0.151587251**	0	55
T25	**-0.353643694**	0	52
T22	**-0.365270553**	0	54
T54	**-0.386411713**	0	51
T55	**-0.46357102**	0	66
T10	**-0.552811926**	0	50
T24	**-0.56194093**	0	54
T13	**-0.643420256**	0	54
T29	**-0.783661162**	0	51
T60	**-0.870481405**	0	55
T16	**-0.910897024**	0	49

Signature 1: 0.0 cut-off

Sample	Signature 2	Recurrence	DFI
T26	**0.953225**	1	14
T17	**0.920443397**	1	3
T62	**0.879103094**	1	30
T33	**0.874178186**	0	15
T23	**0.856271899**	1	37
T04	**0.783880759**	1	46
T57	**0.691657892**	1	4
T59	**0.622838225**	1	26
T01	**0.577946771**	0	55
T24	**0.261800797**	0	54
T22	**-0.133360715**	0	54
T46	**-0.201007164**	0	57
T10	**-0.257007099**	0	50
T13	**-0.331459078**	0	54
T60	**-0.396069611**	0	55
T16	**-0.498169849**	0	49
T25	**-0.501777706**	0	52
T45	**-0.570243976**	1	6
T29	**-0.779674211**	0	51
T54	**-0.824024399**	0	51
T55	**-0.899114823**	0	66

Signature 2: 0.6 cut-off

Sample	Signature 3	Recurrence	DFI
T59	**0.976031988**	1	26
T17	**0.912978554**	1	3
T04	**0.890977963**	1	46
T57	**0.884248877**	1	4
T26	**0.853505414**	1	14
T13	**0.818307891**	0	54
T23	**0.712833807**	1	37
T33	**0.509509129**	0	15
T46	**-0.326285749**	0	57
T45	**-0.353301599**	1	6
T24	**-0.370198423**	0	54
T10	**-0.381668168**	0	50
T29	**-0.674430912**	0	51
T25	**-0.729354419**	0	52
T22	**-0.838207571**	0	54
T16	**-0.838207571**	0	49
T62	**-0.922608457**	1	30
T01	**-0.931739574**	0	55
T54	**-0.938184703**	0	51
T55	**-0.964914426**	0	66
T60	**-0.969165663**	0	55

Signature 3: 0.6 cut-off

Table 7S. Cox Proportional Hazards Survival Regression

Reference: *Statistical Models and Methods for Lifetime Data*, by J. F. Lawless. 1982, John Wiley & Sons, New York.

Variable:

PRE RP PSA	RP GLSN SUM	AGE	**Algorithm**

1 2 3 4

Descriptive Stats...

Variable	Avg	SD
1	11.2085	9.1544
2	7.0759	0.9382
3	60.6215	6.1450
4	0.4177	0.4932

Iteration History...

-2 Log Likelihood = 298.5693 (Null Model)
-2 Log Likelihood = 265.1709
-2 Log Likelihood = 264.8556
-2 Log Likelihood = 264.8550
-2 Log Likelihood = 264.8550 (Converged)

Overall Model Fit...

Chi Square= 33.7143; df=4; p= 0.0000

Coefficients, Std Errs, Signif, and Conf Intervs...

Var	Coeff.	StdErr	p	Lo95%	Hi95%
1	0.0361	0.0191	0.0593	-0.0014	0.0735
2	0.4297	0.1987	0.0306	0.0402	0.8191
3	0.0502	0.0301	0.0956	-0.0088	0.1092
4	1.3894	0.3556	0.0001	0.6924	2.0864

Risk Ratios and Confidence Intervs...

Var	Risk Ratio	Lo95%	Hi95%
1	1.0367	0.9986	1.0763
2	1.5367	1.0410	2.2684
3	1.0515	0.9912	1.1154
4	4.0124	1.9985	8.0556