

Microarray analysis identifies a death-from-cancer signature predicting therapy failure in patients with multiple types of cancer

Gennadi V. Glinsky, ... , Olga Berezovska, Anna B. Glinskii

J Clin Invest. 2005;115(6):1503-1521. <https://doi.org/10.1172/JCI23412>.

Research Article

Oncology

Activation in transformed cells of normal stem cells' self-renewal pathways might contribute to the survival life cycle of cancer stem cells and promote tumor progression. The *BMI-1* oncogene-driven gene expression pathway is essential for the self-renewal of hematopoietic and neural stem cells. We applied a mouse/human comparative translational genomics approach to identify an 11-gene signature that consistently displays a stem cell-resembling expression profile in distant metastatic lesions as revealed by the analysis of metastases and primary tumors from a transgenic mouse model of prostate cancer and cancer patients. To further validate these results, we examined the prognostic power of the 11-gene signature in several independent therapy-outcome sets of clinical samples obtained from 1,153 cancer patients diagnosed with 11 different types of cancer, including 5 epithelial malignancies (prostate, breast, lung, ovarian, and bladder cancers) and 5 nonepithelial malignancies (lymphoma, mesothelioma, medulloblastoma, glioma, and acute myeloid leukemia). Kaplan-Meier analysis demonstrated that a stem cell-like expression profile of the 11-gene signature in primary tumors is a consistent powerful predictor of a short interval to disease recurrence, distant metastasis, and death after therapy in cancer patients diagnosed with 11 distinct types of cancer. These data suggest the presence of a conserved BMI-1-driven pathway, which is similarly engaged in both normal stem cells and a highly malignant subset of human cancers diagnosed in a wide [...]

Find the latest version:

<https://jci.me/23412/pdf>





Microarray analysis identifies a death-from-cancer signature predicting therapy failure in patients with multiple types of cancer

Gennadi V. Glinsky, Olga Berezovska, and Anna B. Glinskii

Sidney Kimmel Cancer Center, San Diego, California, USA.

Activation in transformed cells of normal stem cells' self-renewal pathways might contribute to the survival life cycle of cancer stem cells and promote tumor progression. The *BMI-1* oncogene-driven gene expression pathway is essential for the self-renewal of hematopoietic and neural stem cells. We applied a mouse/human comparative translational genomics approach to identify an 11-gene signature that consistently displays a stem cell-resembling expression profile in distant metastatic lesions as revealed by the analysis of metastases and primary tumors from a transgenic mouse model of prostate cancer and cancer patients. To further validate these results, we examined the prognostic power of the 11-gene signature in several independent therapy-outcome sets of clinical samples obtained from 1,153 cancer patients diagnosed with 11 different types of cancer, including 5 epithelial malignancies (prostate, breast, lung, ovarian, and bladder cancers) and 5 nonepithelial malignancies (lymphoma, mesothelioma, medulloblastoma, glioma, and acute myeloid leukemia). Kaplan-Meier analysis demonstrated that a stem cell-like expression profile of the 11-gene signature in primary tumors is a consistent powerful predictor of a short interval to disease recurrence, distant metastasis, and death after therapy in cancer patients diagnosed with 11 distinct types of cancer. These data suggest the presence of a conserved *BMI-1*-driven pathway, which is similarly engaged in both normal stem cells and a highly malignant subset of human cancers diagnosed in a wide range of organs and uniformly exhibiting a marked propensity toward metastatic dissemination as well as a high probability of unfavorable therapy outcome.

Introduction

Recent studies indicate that the *Polycomb group* (PcG) gene *Bmi-1* determines the proliferative potential of normal and leukemic stem cells (1) and is required for the self-renewal of hematopoietic and neural stem cells (1–3). Self-renewal ability is an essential defining property of a pluripotent stem cell-like phenotype distinguishing stem cells from other cell types (4). An emerging concept of “tumor stem cells” argues that the presence of a rare stem cell-resembling population of cancer cells among the heterogeneous mix of cells constituting a tumor is essential for tumor progression and metastasis of epithelial malignancies (5–7). The concept of tumor stem cells implies that common genetic pathways might define critical stem cell-like functions in both normal and neoplastic stem cells (1, 5). *Bmi-1* oncogene is expressed in all primary myeloid leukemia and leukemic cell lines analyzed so far (1, 8), and overexpression of *Bmi-1* causes neoplastic transformation of lymphocytes (9, 10). Recently, *BMI-1* expression was reported in human non-small-cell lung cancer (11) and breast cancer cell lines (12); this suggests an oncogenic role for *BMI-1* activation in epithelial malignancies.

Expression profiling of prostate tumor samples using oligonucleotide or cDNA microarray technology rapidly emerged

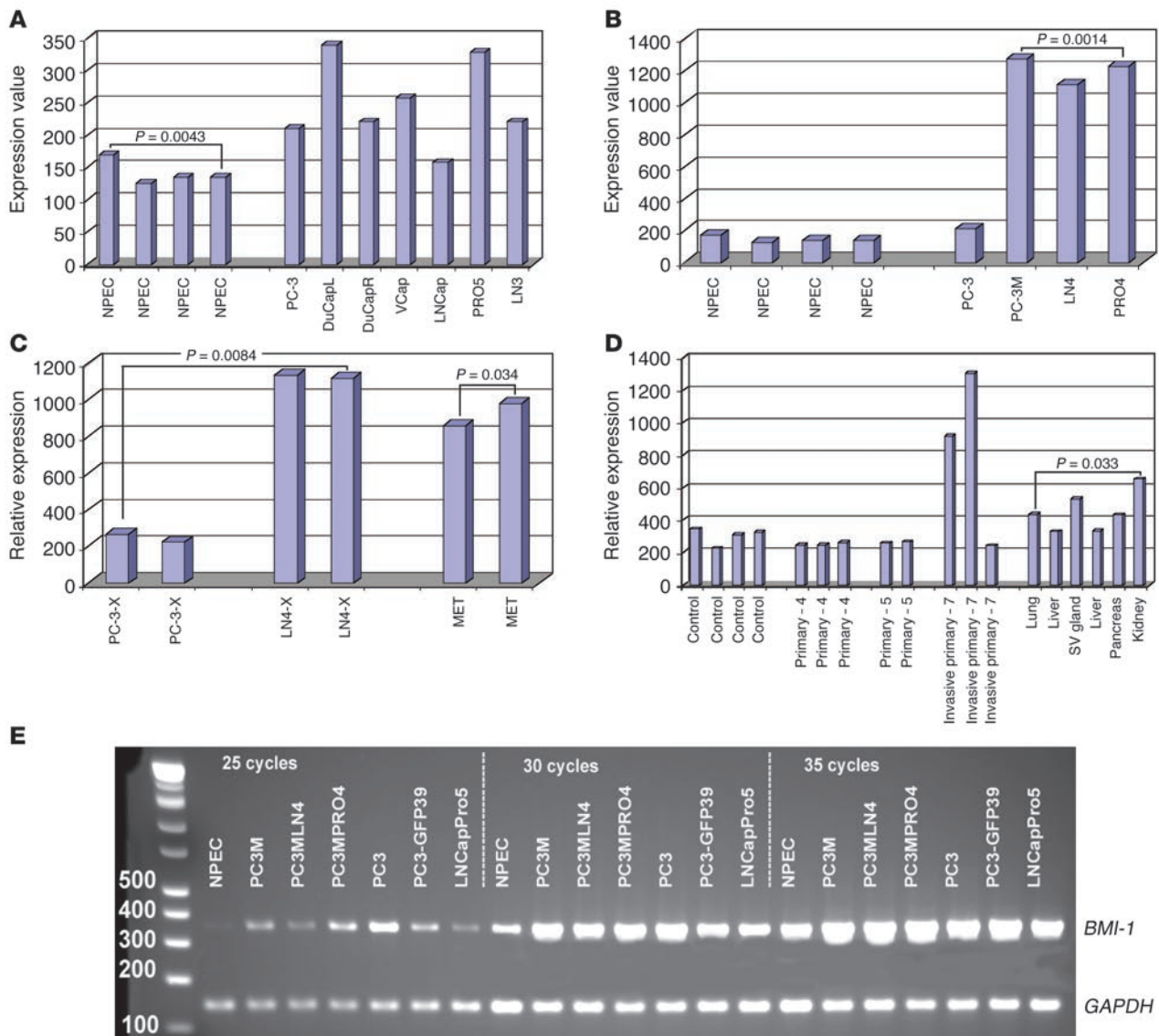
as a powerful tool to reveal multiple gene expression signatures associated with human prostate cancer (13–27), including potential prostate cancer prognosis markers (14, 15, 18, 24–26). However, one of the major limitations of these studies was that the same clinical data set was used for both signature discovery and validation. Furthermore, microarray analysis typically identifies vast data sets of candidate markers, and usually only a single hit or a few hits were validated using independent methods and independent clinical data sets; this diminished the potential advantage of the use of a panel of markers over a single marker in diagnostic and/or prognostic applications. Recently we attempted to address some of these limitations by using microarray-based gene expression profiling to identify molecular signatures that comprise small clusters of coregulated transcripts and distinguish subgroups of prostate cancer patients with differing outcomes after therapy (15). In these experiments for signature discovery and validation, we used 2 independent cohorts of prostate cancer patients (15). The proposed prostate cancer prognosis predictor algorithm uses a panel of 3 molecular signatures and appears to demonstrate high discrimination accuracy between subgroups of patients with distinct clinical outcome after therapy, providing additional predictive value over conventional markers of outcome (15).

Most recently, the global gene expression profiling approach was successfully used to identify molecular signatures associated with activation of oncogenic pathways (28), targeted genetic manipulations (29), and cellular responses to physiological stimuli (30) and to build robust transcriptional identifiers that reliably recognize the engagement of corresponding pathways within the highly complex patterns of gene expression in experimental and clinical samples.

Nonstandard abbreviations used: MSKCC, Memorial Sloan-Kettering Cancer Center; MTTs, metastatic TRAMP tumor sample; PAI, phenotype association index; PNS, peripheral nervous system; PSA, prostate-specific antigen; Q-RT-PCR, quantitative RT-PCR; RP, radical prostatectomy; siRNA, small interfering RNA; SPAI, stem cell-like PAI; SVM, support vector machine; TRN, terrain.

Conflict of interest: Gennadi V. Glinsky is a named inventor on a related patent application and serves as a consultant to Stratagene Inc. (La Jolla, California, USA) in the field of cancer biomarker development.

Citation for this article: *J. Clin. Invest.* 115:1503–1521 (2005). doi:10.1172/JCI23412.

**Figure 1**

Microarray (A–D) and RT-PCR (E) analyses reveal increased expression of *BMI-1* mRNA in multiple human prostate cancer cell lines established from metastatic tumors (PC-3, LNCap, DuCap, VCap, etc.) compared with normal human prostate epithelial cells (NPEC) (A and E); in xenograft-derived human prostate cancer cell line variants (PC-3M, PC-3MLN4, PC-3MPro4) compared with the plastic-maintained parental cells (PC-3) (B); in highly metastatic human prostate carcinoma xenografts (PC-3MLN4) compared with the less metastatic parental counterparts (PC-3) growing orthotopically in nude mice (C); in lymph node metastases of human prostate cancer growing in the prostate of nude mice (MET) (C); and in invasive primary prostate tumors and distant metastatic lesions in the TRAMP transgenic mouse model of prostate cancer (D). Prostate tissues from age-matched wild-type C57BL/6 mice served as control samples in Figure 1D. The numbers 4, 5, and 7 indicate the age of TRAMP mice (in months). Each sample represents a pool of tissues from 3–5 mice. *P* values were obtained using a 2-tailed *t* test. LN3, LNCapLN3; LN4, PC-3MLN4; PRO4, PC-3MPro4; PRO5, LNCapPRO5; SV, seminal vesicles; X, human xenograft tumors in nude mice.

We hypothesized that molecular signatures associated with activation of a normal stem cell's self-renewal program in metastatic cancer cells might be detectable by searching for genes exhibiting concordant patterns of regulation in metastatic lesions and normal stem cells in *Bmi-1*^{+/+} versus *Bmi-1*^{-/-} genetic backgrounds. Here we report that the expression of *Bmi-1* is elevated and a stem cell-like *BMI-1*-associated gene expression pathway is activated in metastatic prostate tumors. We provide evidence that the stem cell-resembling expression profile of the 11-gene signature in pri-

mary prostate tumors predicts therapy failure in prostate cancer patients. We show that expression of the 11-gene signature is a powerful predictor of a short interval to distant metastasis and poor survival after therapy in breast and lung cancer patients diagnosed with an early-stage disease. Finally, we extend our therapy-outcome analysis to include 1,153 cancer patients diagnosed with 11 different cancer types and demonstrate that prognostic power of the 11-gene signature is informative in all 11 different types of human cancer diagnosed in multiple organs.

**Table 1**Q-RT-PCR analysis of *Bmi-1* mRNA expression in human prostate carcinoma cell lines

Cell line	Average expression value ^A	SD	P value ^B
NPEC	0.090656645	0.0154152	
LNCap	0.216610094	0.0311867	0.0013481
LNCapPro5	0.292913482	0.0222714	1.472×10^{-5}
LNCapLN3	0.235569094	0.0429103	0.0038571
PC-3	1.030811318	0.1271548	0.000586
PC-3MLN4	0.635668126	0.0892679	0.0009314
PC-3MPro4	1.424229109	0.1758348	0.0005788
VCap	0.192483261	0.012621	6.494×10^{-5}
DuCap	0.128637764	0.012266	0.0092371

^ANormalized average expression value from 4 measurements. ^BTwo-tailed Student's *t* test compared with normal human prostate epithelial cells.

Results

***Bmi-1* oncogene expression is elevated in prostate cancer.** Recent experimental observations documented an increased *Bmi-1* expression in human non-small-cell lung cancer (11), human breast carcinomas (31), and established breast cancer cell lines (12), suggesting that an oncogenic role of *Bmi-1* activation may be extended beyond leukemia and, perhaps, may affect progression of the epithelial malignancies as well. Microarray gene expression analysis of established cancer cell lines representing multiple experimental models of human prostate cancer (16) revealed that *BMI-1* expression seems to be consistently elevated in human prostate cancer cell lines compared with the primary cultures of normal human prostate epithelial cells (Figure 1, A and B). To validate the results of the microarray experiments, we confirmed these observations using quantitative RT-PCR (Q-RT-PCR) analysis of *BMI-1* mRNA expression (Table 1; and see Supplemental Figure 1; supplemental material available online with this article; doi:10.1172/JCI23412DS1). Thus, results of expression profiling experiments appear to support the notion that transcriptional activation of the *BMI-1* gene is frequently associated with human prostate cancer.

Interestingly, microarray analysis shows markedly higher *BMI-1* expression levels in lymph node metastases and highly metastatic orthotopic xenografts of human prostate carcinoma in nude mice compared with the less metastatic counterparts (Figure 1C), implying that *BMI-1* activation might be associated with aggressive malignant behavior of prostate carcinoma cells. To test this hypothesis, we carried out expression profiling analysis of approximately 12,000 transcripts in a transgenic mouse model of metastatic prostate cancer (32). Microarray experiments detected increased levels of *Bmi-1* mRNA expression in late-stage invasive primary tumors and multiple distant metastatic lesions in the TRAMP transgenic mouse model of prostate cancer (Figure 1D), thus lending more credence to the idea that activation of a *BMI-1*-associated pathway is linked with prostate cancer metastasis. It should be pointed out that despite the apparently consistent pattern of increased *Bmi-1* expression in prostate cancer, there is considerable variability in the degree of elevation of *Bmi-1* expression at the distinct sites of malignant growth in vivo (Figure 1D). We carried out the Q-RT-PCR analysis of *Bmi-1* mRNA expression in 4 additional late-stage invasive primary prostate tumors of the 6- to 7-month old TRAMP mice and confirmed the 2- to 8-fold increase in *Bmi-1* expression in all 4 tumors (data not shown).

Identification of a *BMI-1*-pathway signature with concordant expression profiles in normal stem cells and distant metastatic lesions in a transgenic mouse model of prostate cancer. Recent experiments established that the *Bmi-1* gene is required for self-renewal of hematopoietic and neural stem cells (1–3) and identified *BMI-1*-regulated genes in neural stem cells that are presumably engaged in an execution of self-renewal programs in a state of both central nervous system (CNS) and peripheral nervous system (PNS) neurospheres (3). We hypothesized that molecular signatures associated with activation of a normal-stem cell's self-renewal program in metastatic cancer cells might be detectable by looking for genes manifesting concordant patterns of regulation in metastasis and normal stem cells in *Bmi-1*^{+/+} versus *Bmi-1*^{-/-} genetic backgrounds. Therefore, we set out to determine whether expression profiles of transcripts activated and suppressed in prostate cancer metastases would recapitulate the expression profile of the *BMI-1*-regulated genes in normal stem cells, by comparing the sets of differentially regulated genes in search of intersection of lists for both up- and downregulated transcripts (see Figures 2 and 3, Methods, and supplemental material for description of a signature discovery protocol). This analysis identified genes exhibiting highly concordant profiles of transcript-abundance behavior in prostate cancer metastases and *Bmi-1*^{+/+} versus *Bmi-1*^{-/-} PNS neurospheres, suggesting the presence of a conserved *BMI-1*-regulated pathway(s) similarly engaged in both normal stem cells and distant metastatic lesions of prostate carcinoma (Figures 2 and 3).

The metastatic TRAMP tumor sample (MTTS) signature is likely to be enriched for genes discriminative for the metastatic phenotype. It is reasonable to assume that many gene expression patterns wired into the MTTS signature would manifest the power to discriminate the metastatic phenotype and would have no relation to the transcriptional program of normal stem cells. We sought to use these features of the MTTS signature for identification of the gene expression components of a stem cell transcriptome that are coordinately expressed in metastatic cancer cells and might manifest discriminative diagnostic power for the malignant phenotype. Sets of differentially regulated transcripts were independently identified for distant metastatic lesions and primary prostate tumors versus age-matched control samples in a transgenic TRAMP mouse model of metastatic prostate cancer (MTTS signature) as well as PNS (PNS signature) and CNS (CNS signature) neurospheres in *Bmi-1*^{+/+} versus *Bmi-1*^{-/-} backgrounds. This analytical step defined 3 large parent signatures (Figure 3): MTTS signature comprising 868 upregulated and 477 downregulated transcripts; PNS signature comprising 885 upregulated and 1,088 downregulated transcripts; and CNS signature comprising 769 upregulated and 778 downregulated transcripts.

Next we intersected the MTSS signature with the stem cell signatures in the state of PNS and CNS neurospheres to identify concordant sets of genes and define the stem cell signatures embedded into MTSS signature (Figures 2 and 3). Subsets of transcripts exhibiting concordant expression changes in MTTS (MTTS signature) as well as PNS (PNS signature) and CNS (CNS signature) neurospheres in *Bmi-1*^{+/+} versus *Bmi-1*^{-/-} backgrounds were identified. Thus, 2 concordant subsets of transcripts were identified corresponding to each binary comparison of metastatic TRAMP tumors and neural stem cell samples in a state of PNS and CNS neurospheres (141 upregulated and 58 downregulated transcripts for PNS neurospheres [$r = 0.7593$; $P < 0.0001$; Figure 2A] and 40 upregulated and 24 downregulated transcripts for CNS

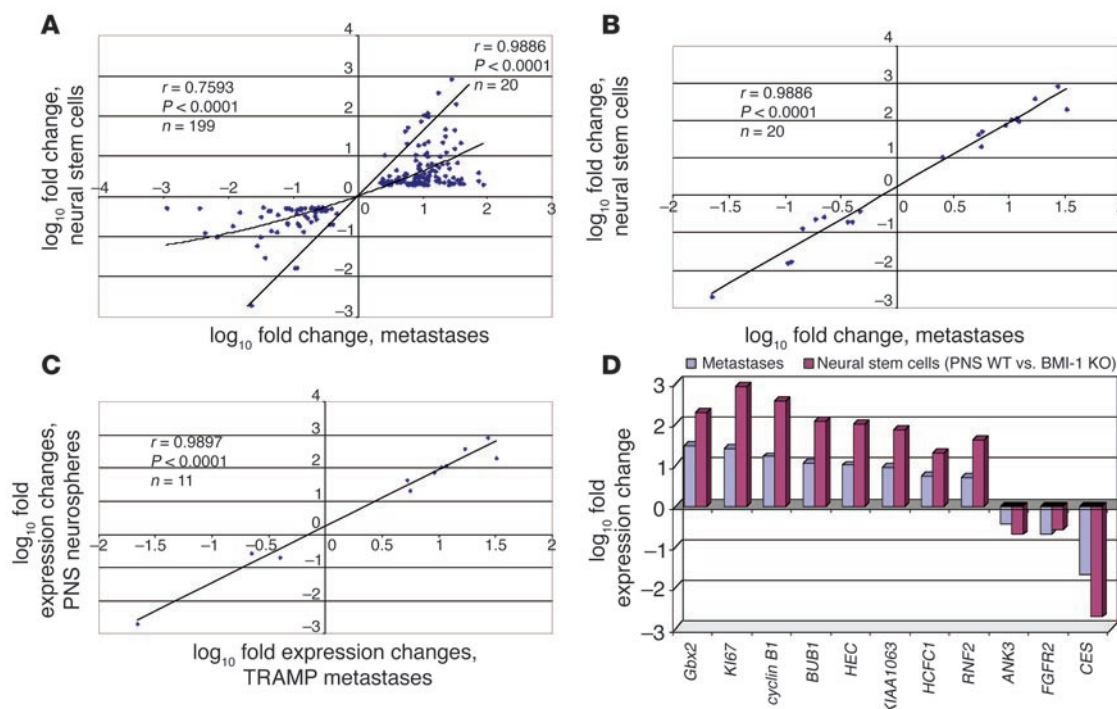


Figure 2

Distant metastatic lesions in the TRAMP transgenic mouse model of prostate cancer exhibit stem cell-like expression signatures of the BMI-1 pathway. Transcripts differentially regulated in distant metastatic lesions of 6-month-old TRAMP mice (MTTS signature) were compared with the BMI-1-regulated genes in neural stem cells (3) in search of intersection of lists. (A) Expression profiles and the corresponding Pearson correlation coefficient for 199 genes (141 upregulated and 58 downregulated) comprising concordant differentially regulated sets of transcripts in metastatic TRAMP samples and PNS neurospheres are shown. Small gene expression signatures comprising transcripts with a high level of expression correlation in metastatic cancer cells and stem cells (the selection threshold for small signatures was arbitrarily set at Pearson correlation coefficients greater than 0.95) were selected from large concordant sets. The reduction in the signature transcript number was terminated when further elimination of a transcript did not increase the value of the Pearson correlation coefficient. Using this approach, a single candidate prognostic gene expression signature was selected for each binary intersection of the MTTS signature and parent stem cell signatures (Figure 3). Consecutive steps of selection from the 199-gene concordant set of a subset of 20 genes (A and B) and a small MTTS/PNS 11-gene signature (C and D) are shown. In D, $r = 0.9897$, $P < 0.0001$ between gene groups, and $n = 11$ per group. Complete lists of genes and corresponding concordant subsets are shown in Supplemental Table 2. See text, Figure 3, and Table 3 for details.

neurospheres [$r = 0.7679$; $P < 0.0001$]). A third concordant subset of 27 genes comprising 15 upregulated and 12 downregulated transcripts was selected for intersection common to all 3 signatures ($r = 0.8002$; $P < 0.0001$).

This analysis also identified a stem cell-like expression profile for transcripts coordinately expressed in metastatic cancer cells and normal stem cells, which can be used as a consistent reference standard to interrogate independent data sets for possible presence of a stem cell-like expression signature (Figure 2). Practical considerations essential for future development of genetic diagnostic tests using an analytical platform most compatible with the state-of-the-art clinical laboratory practice prompted us to select from concordant gene sets small gene expression signatures comprising transcripts with a high level of expression correlation in metastatic cancer cells and stem cells (the selection threshold for small signatures was arbitrarily set at Pearson correlation coefficients greater than 0.95). The reduction in the signature transcript number was terminated when further elimination of a transcript did not increase the value of the Pearson correlation coefficient. Using this approach, a single candidate prognostic gene expression signature was selected for each binary intersection of the MTTS signature and parent stem cell signatures (Figure 3). Then

small signatures (1 11-gene signature for the PNS set, 1 11-gene signature for the CNS set, and one 14-gene signature for the common PNS/CNS set) were tested for the power to discriminate the metastatic phenotype (using 1 mouse prostate cancer data set and 1 human prostate cancer data set comprising primary and metastatic tumors) and therapy-outcome classification performance (using human prostate cancer therapy outcome set 1). Based on diagnostic and prognostic classification performance, a single best-performing 11-gene MTTS/PNS signature was selected for further validation analysis (Figures 3 and 4).

During the malignant-phenotype classification performance tests (Figure 4), we asked whether individual metastatic lesions and primary prostate tumors would exhibit a stem cell-like expression profile of the candidate prognostic signatures. We selected for this analysis 3 small signatures demonstrating the most significant correlation (Figures 2 and 3) of expression profiles in stem cells and prostate cancer metastasis. To assess a degree of similarity of the signature expression profiles in individual tumor samples and normal stem cells, we calculated a Pearson correlation coefficient for each sample by comparing signature expression profile in an individual sample to the stem cell-associated expression profile of the corresponding small signatures. Based on

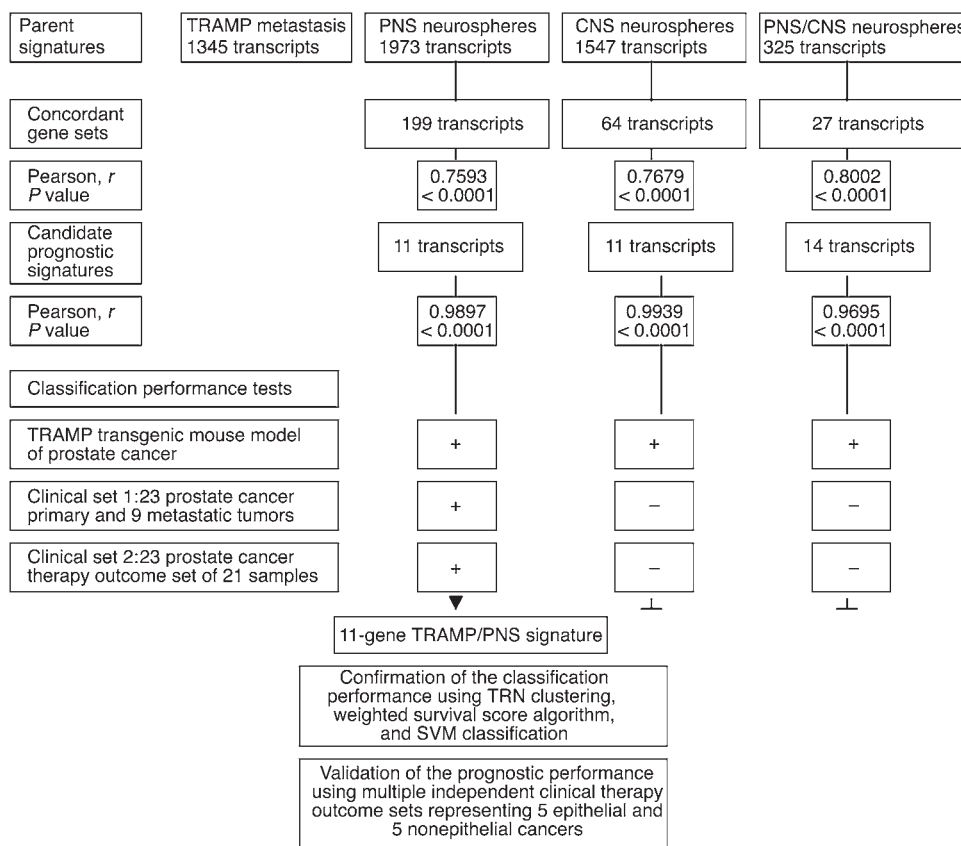


Figure 3

Sequential analytical steps used for identification, selection, and validation of the 11-gene death-from-cancer signature. The figure shows an overview of the approach used for the development and validation of a cancer survival predictor based on gene expression monitoring.

expected similarity of the prognostic signatures in stem cells and prostate cancer metastasis, we named the corresponding Pearson correlation coefficients measured for individual samples the stem cell-like phenotype association indices (SPAIs; see Methods for a description of definition and measurement of SPAIs). As shown in Figure 4, A, D, and G, 2 of 3 late-stage invasive primary tumors and all distant metastatic lesions in the TRAMP transgenic mouse model of prostate cancer have positive SPAIs, thus manifesting a stem cell-like expression profile of the small signatures.

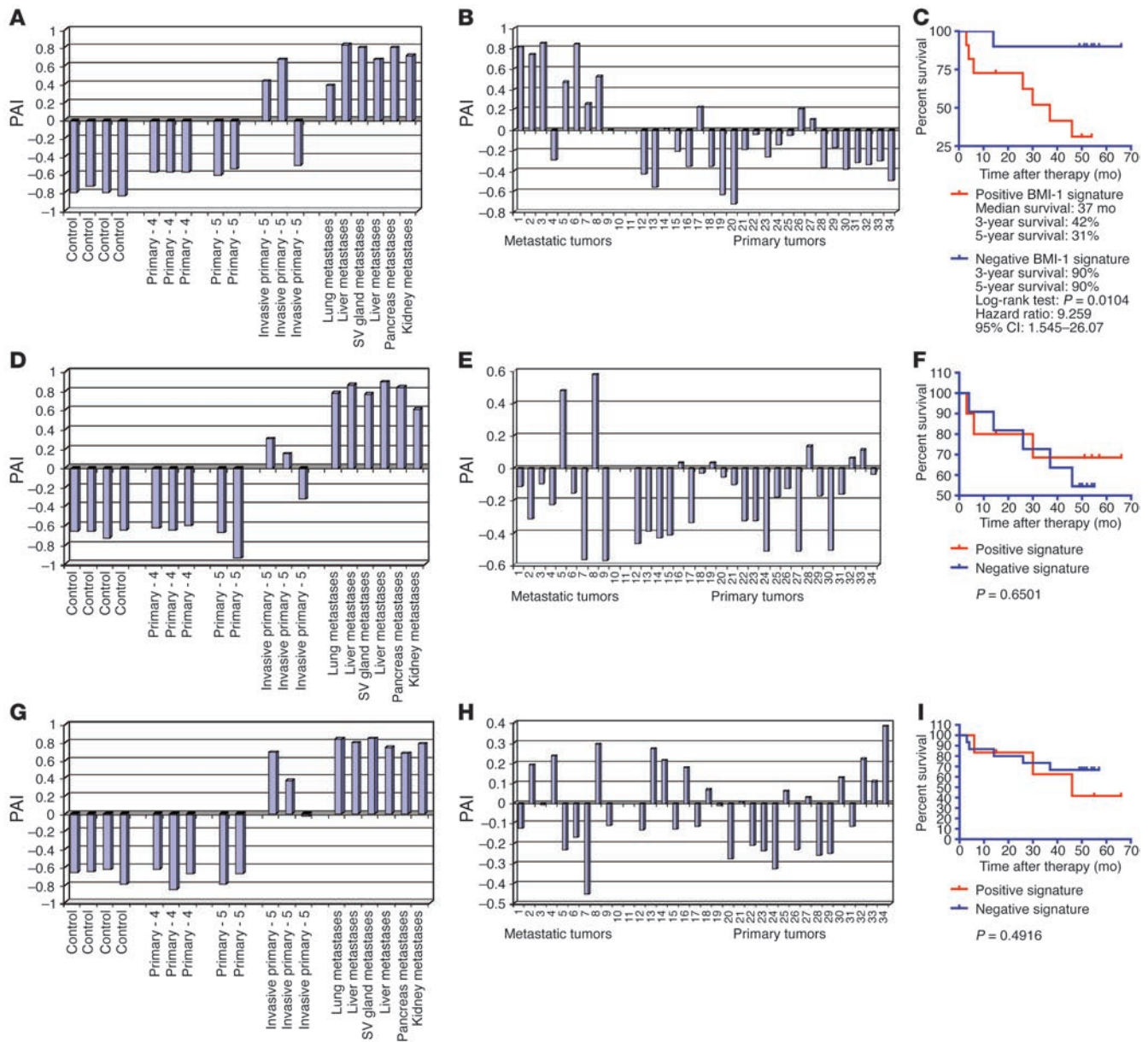
Distant metastatic lesions and primary prostate tumors from cancer patients with differing therapy outcome display distinct expression profiles of the 11-gene MTTs/PNS signature. To perform similar analysis for human tumors, we translated the murine small signatures into a list of human homologs using the NCBI UniGene database (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unigene>) and retrieved the expression data for corresponding Affymetrix probe sets (Tables 2 and 3 and Supplemental Table 2). We calculated the SPAIs for each of the 9 metastatic tumors and 23 primary prostate carcinomas and determined that 7 of 9 samples of distant metastatic lesions from prostate cancer patients exhibited a stem cell-like expression profile of the 11-gene MTTs/PNS signature (Figure 4B). In contrast, a majority of primary prostate tumors seemed to display a distinct expression profile of the 11-gene MTTs/PNS signature as manifested in negative SPAI values (Figure 4B). Interestingly, a subset of samples of primary prostate

carcinomas manifested expression profiles of the 11-gene MTTs/PNS signature similar to those of the metastatic tumors, as reflected in positive correlation coefficients (positive SPAI values in Figure 4B), suggesting that primary prostate tumors with distinct expression profiles of the PNS neurosphere-derived 11-gene signature (e.g., positive and negative SPAI values) may have different biological features and distinct clinical courses of disease progression. Validation analysis using the CNS neurosphere-derived MTTs/CNS 11-gene signature and MTTs/PNS/CNS 14-gene signature indicates that application of these signatures is less informative in distinguishing metastatic and primary human prostate tumors (Figure 4, E and H).

To evaluate the potential biological significance and clinical utility of the 11-gene MTTs/PNS signature expression in human prostate cancer, we set out to examine whether the detection of a stem cell-like expression profile in primary prostate tumors of individual cancer patients would help in patients' stratification at the

time of diagnosis into subgroups with distinct courses of disease progression based on differing therapy outcome after radical prostatectomy (RP). We assessed the prognostic power of the 11-gene signature based on ability to segregate the patients with recurrent and nonrecurrent course of disease progression after RP into distinct subgroups. We calculated a Pearson correlation coefficient for each of 21 tumor samples of outcome set 1 by comparing the 11-gene signature expression profiles of individual samples with the stem cell-like expression profile of the 11-gene BMI-1-pathway signature in PNS neurospheres (Figure 2). To determine the prognostic power of the 11-gene signature, we performed the Kaplan-Meier survival analysis using, as a clinical endpoint, the disease-free interval after therapy in prostate cancer patients with positive and negative SPAIs.

The Kaplan-Meier survival curves showed a highly significant difference in the probability that prostate cancer patients would remain disease-free after therapy between the groups with positive and negative SPAIs defined by the 11-gene MTTs/PNS signature (Figure 4C), suggesting that patients with positive SPAIs exhibit a poor outcome signature whereas patients with negative SPAIs manifest a good outcome signature. The estimated hazard ratio for disease recurrence after therapy in the group of patients with positive SPAIs as compared with the group of patients with negative SPAIs defined by the 11-gene MTTs/PNS signature (Figure 4C) was 9.259 (95% confidence interval of ratio, 1.545–26.07; $P = 0.0104$).

**Figure 4**

Selection of the best-performing small signature based on evaluation of the metastatic-phenotype-discrimination performance and therapy-outcome prediction power of candidate prognostic signatures. Expression profiles of the 3 small signatures (11-gene MTTs/PNS signature, **A–C**; 11-gene MTTs/CNS signature, **D–F**; and 14-gene MTTs/PNS/CNS signature, **G–I**) were evaluated in metastatic lesions at multiple distant target organs and primary prostate carcinomas in the TRAMP transgenic mouse model of prostate cancer (**A**, **D**, and **G**) and prostate cancer patients (**B**, **E**, and **H**) for presence of a stem cell-like expression profile. (**B**, **E**, and **H**) Data from the analysis of 9 distant metastatic lesions and 23 primary human prostate carcinoma samples. (**C**, **F**, and **I**) Kaplan-Meier analysis of the probability that patients would remain disease-free among 21 prostate cancer patients constituting clinical outcome set 1, according to whether they had a good-prognosis or a poor-prognosis signature as defined by the expression profiles of the small prognostic signatures. The y axes in **A**, **B**, **D**, **E**, **G**, and **H** show the SPAI values in corresponding metastatic and primary tumor samples (see Methods for a description of SPAI definition and calculation). CI, confidence interval.

Fifty-eight percent of patients with positive SPAIs had a disease recurrence within 3 years after therapy, whereas 90% of patients with negative SPAIs remained relapse-free (Figure 4C). Five years after therapy, 69% of patients with positive SPAIs had a disease recurrence, whereas 90% of patients with negative SPAIs remained relapse-free (Figure 4C). In contrast to the PNS neurosphere-derived signature, the CNS neurosphere-derived signature failed

to stratify the prostate cancer patients into prognostic subgroups with distinct probability of disease relapse after therapy ($P = 0.6501$; Figure 4F). Similarly, the 14-gene MTTs/PNS/CNS signature failed in both classification-performance tests using human cancer specimens ($P = 0.4916$; Figure 4, H and I). Based on this analysis, we identified the 11-gene MTTs/PNS signature as a best-performing malignant-phenotype classifier and proposed to identify the

**Table 2**

Cancer types and number of cancer patients in the therapy-outcome sets analyzed in this study

Cancer type	No. of patients	Reference
Prostate cancer	100	14, 15
Breast cancer	97	33
Lung cancer	211	34, 56
Ovarian cancer	81	57, 58
Bladder cancer	31	59
Lymphoma	298	60, 61
MCL	92	62
Mesothelioma	17	63
Medulloblastoma	60	64
Glioma	50	65
AML	116	66
Total	1,153	

MCL, mantle cell lymphoma; AML, acute myeloid leukemia.

group of prostate cancer patients with positive values of the PNS neurosphere-derived 11-gene signature as a poor-prognosis group and the group of prostate cancer patients with negative values of the 11-gene signature as a good-prognosis group.

The identified signature genes were defined based on a strong correlative behavior in multiple independent sets of experimental and clinical samples obtained from 2 species (mice and humans). To test by independent methods the suspected association of the expression of BMI-1–pathway target genes with the expression of the BMI-1 gene product in the context of human cancer cells, we subjected human prostate carcinoma cells to small interfering RNA-mediated (siRNA-mediated) silencing of expression of the endogenous *BMI-1* gene. The PC-3-32 human prostate carcinoma cells were transfected with *BMI-1* or control siRNAs and continuously monitored for mRNA expression levels of *BMI-1* and a selected set of genes using RT-PCR and Q-RT-PCR methods (data not shown). Q-RT-PCR and RT-PCR analyses showed that the siRNA-mediated *BMI-1*–silencing protocol allowed for approximately 90% inhibition of the endogenous *BMI-1* mRNA expression. We validated the effect of siRNA-mediated *BMI-1* silencing at the BMI-1 protein expression level using immunofluorescent analysis. The *BMI-1* silencing was specific, since the expression levels of 9 unrelated transcripts (such as *GAPDH*, *EZH2*, and several other genes) were not altered (data not shown). Consistent with the hypothesis that expression of genes comprising the BMI-1–pathway signature is associated with the expression of the BMI-1 gene product, mRNA abundance levels of 8 of 11 interrogated BMI-1–pathway target genes were altered in the human prostate carcinoma cells with approximately 90% silenced *BMI-1* gene.

Reduction of the *BMI-1* mRNA and protein expression in human prostate carcinoma metastasis precursor cells did not alter significantly the viability of adherent cultures grown at the optimal growth condition and in serum-starvation experiments (data not shown) and had only modest inhibitory effect on proliferation (an approximately 25–30% reduction in the number

of cells during the 3-day silencing protocol). However, the ability of human prostate carcinoma cells to survive in a nonadherent state was severely affected after siRNA-mediated reduction of the *BMI-1* expression. Fluorescence-activated cell sorting (FACS) analysis revealed an approximately 3-fold increase of apoptosis in the *BMI-1* siRNA-treated human prostate carcinoma cells cultured in nonadherent conditions. These data suggest that human prostate carcinoma cells expressing a high level of the BMI-1 protein are more resistant to apoptosis induced in cells of epithelial origin in response to attachment deprivation (anoikis) and, perhaps, would survive better in blood during metastatic dissemination, thus forming a pool of metastasis precursor cells that can survive circulatory stress. Thorough follow-up experiments would be required to establish to a full extent the biological and functional role of BMI-1 overexpression and BMI-1–pathway activation in the various epithelial cancers.

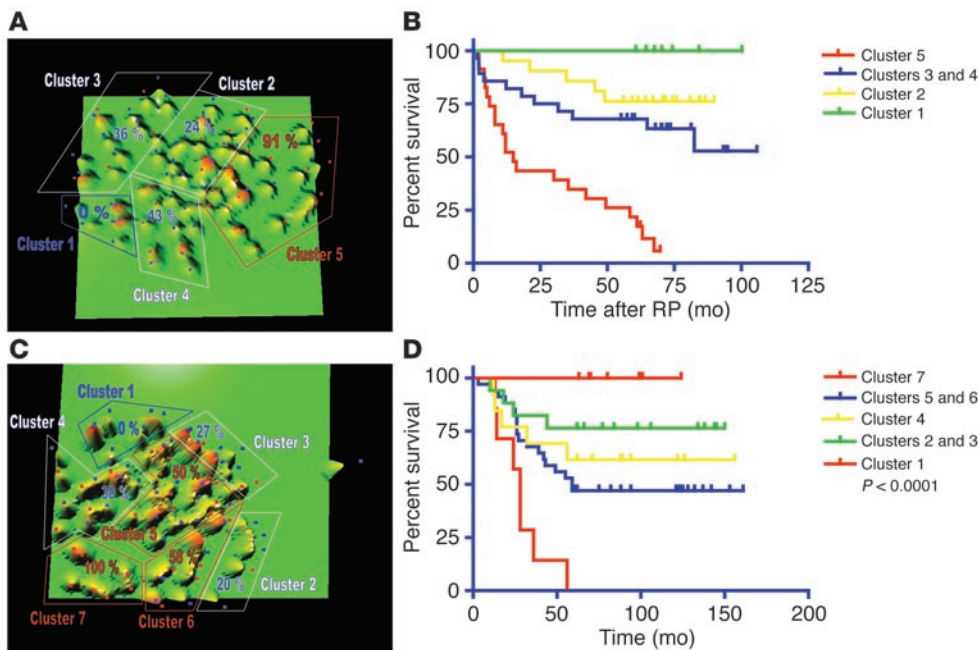
Expression of the 11-gene MTTs/PNS signature in primary prostate tumors is a predictor of a therapy failure in prostate cancer patients. To validate a survival prediction model based on the 11-gene MTTs/PNS signature, we tested the prognostic performance of the model in multiple independent therapy-outcome data sets representing 5 epithelial and 5 nonepithelial cancers (Table 3). We divided patients within individual cohorts into a training set, which was used to select the cutoff threshold and to test the model, and a test set, which was used to evaluate the reproducibility of the classification performance. Using the training set of samples, we selected the prognosis-discrimination cutoff value for a signature based on the highest level of statistical significance in patients' stratification into poor- and good-prognosis groups as determined by the log-rank test (lowest *P* value and highest hazard ratio in the training set). Clinical samples having the Pearson correlation coefficient at or higher than the cutoff value were identified as having the poor-prognosis signature. Clinical samples with the Pearson correlation coefficient lower than the cutoff value were identified as having the good-prognosis signature. The same discrimination cutoff value was then applied to evaluate the reproducibility of the prognostic performance in the test set of patients. Lastly, we applied the model to the entire outcome set using the same cutoff threshold to confirm the classification performance. The training and test sets were balanced with respect to the total number of

Table 3

The 11-gene signature associated with poor prognosis of cancer patients diagnosed with multiple types of cancer

Gene	UniGene ^A (<i>Homo sapiens</i>)	Affymetrix MG-U74A probe set	GenBank accession no.	UniGene ^A (<i>Mus musculus</i>)
<i>Gbx2</i>	Hs.184945	94200_at	Z48800	Mm.204730
<i>Klf6</i>	Hs.80976	99457_at	X82786	Mm.4078
<i>CCNB1</i>	Hs.23960	160159_at	X64713	Mm.318364
<i>BUB1</i>	Hs.469649	104097_at	AF002823	Mm.2185
<i>KNTC2</i>	Hs.414407	93441_at	AI595322	Mm.225956
<i>USP22</i>	Hs.462492	97960_at	AW125800	Mm.30602
<i>HCFC1</i>	Hs.83634	100901_at	U80821	Mm.248353
<i>RNF2</i>	Hs.124186	93164_at	Y12783	Mm.31512
<i>ANK3</i>	Hs.499725	98477_s_at	L40632	Mm.235960
<i>FGFR2</i>	Hs.533683	93090_at	M23362	Mm.16340
<i>CES1</i>	Hs.499222	101538_i_at	AW226939	Mm.292803

^AUniGene IDs were updated to correspond with the UniGene cluster IDs in build 183 (NCBI UniGene website; <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unigene>).

**Figure 5**

TRN analysis within the mRNA abundance space of genes constituting the 11-gene MTTs/PNS signature reveals clustering patterns, among prostate cancer (A and B) and breast cancer (C and D) patients, that are associated with distinct frequencies of therapy failure (A and B) and differing probability of disease-free survival after therapy (C and D). A TRN clustering algorithm was applied to the 79 samples (A) constituting prostate cancer therapy outcome set 2 and the 97 samples (C) constituting the breast cancer therapy outcome set. Kaplan-Meier analysis (B and D) was applied to subgroups of patients defined by the TRN clustering algorithm as shown in A and C.

patients, negative and positive therapy outcomes, and the length of survival. At this stage of the analysis, we did not carry out additional model training, development, or optimization steps, except for selection of a prognostic cutoff threshold in the training set. Throughout the study, we consistently used the same MTTs/PNS expression profile as a reference standard to quantify the Pearson correlation coefficients of the individual samples.

In addition to this analysis, we confirmed the model performance using various sample-stratification approaches, such as terrain (TRN) clustering (Figure 5), support vector machine (SVM) classification (Supplemental Table 4), and weighted survival score algorithm (Figure 6E and Figure 7D). Finally, we evaluated the therapy outcome-predictive power of the 11-gene model in a prostate cancer setting using a prognostic test based on an independent method of gene expression analysis, namely the Q-RT-PCR method (Figure 6F).

To further validate the potential clinical utility of the 11-gene MTTs/PNS signature, we evaluated the prognostic power of the 11-gene signature applied to an independent set of 79 clinical samples (prostate cancer outcome set 2) obtained from 37 prostate cancer patients who developed recurrence after the therapy and 42 patients who remained disease-free. In this cohort of patients, the Kaplan-Meier survival analysis demonstrated a highly significant difference in the probability that prostate cancer patients would remain disease-free after therapy between the groups with positive and negative SPAIs defined by the 11-gene BMI-1-pathway signature (Figure 6, A–C). The estimated hazard ratio for disease recurrence after therapy in the group of patients with positive SPAIs as compared with the group of patients with negative SPAIs defined by the 11-gene MTTs/PNS signature (Figure 6C) was 3.74 (95% confidence interval of ratio, 3.010–25.83; $P < 0.0001$). Sixty-seven percent of patients with positive SPAIs had a disease recurrence within 3 years after therapy, whereas 70% of patients with negative SPAIs remained relapse-free (Figure 6C). Five years after therapy, 83% of patients with positive SPAIs had a disease recurrence, whereas 64% of patients with negative SPAIs remained relapse-free (Figure 6C).

The standard Kaplan-Meier log-rank statistic assesses the difference in the survival curves. However, it does not test for multiple hypotheses or account for random co-occurrence; this represents an inherent problem of the gene expression profiling experiments. We attempted to partly mitigate this problem by using an alternative biological endpoint to the patients' survival during the signature selection process and by applying the survival analysis to a single signature, thus eliminating the multiple comparisons from the survival model building protocol. The MTTs signature is likely to carry many gene expression patterns displaying the power to discriminate the metastatic phenotype that have no relation to the transcriptional program of normal stem cells. One of our main goals was to identify the stem cell signature that is associated with the pluripotency self-renewal phenotype and is embedded into MTTs signature. This approach implies that a candidate marker signature would have a defined stem cell-like expression profile that can be used in the subsequent follow-up validation analyses as a reference standard to look for expression of a stem cell-like signature in clinical samples. To further assess the statistical validity of the 11-gene stem cell-like profile, we performed 1,000 random permutations of the 11-gene stem cell profiles randomly selected from the 1,973-gene PNS signature. For each random 11-gene stem cell profile, we assessed its metastatic phenotype-discriminative performance in the TRAMP transgenic mouse model at the discriminative confidence levels of the 11-gene BMI-1-pathway MTTs/PNS signature. Only 1 random 11-gene stem cell profile of the 1,000 permutations demonstrated classification power matching the metastatic phenotype-discriminative performance of the 11-gene MTTs/PNS signature. We performed 10,000 permutations to test the likelihood that small 11-gene signatures derived from the large MTTs signature would display high discrimination power to assess the significance at the 0.1% level. We carried out 10,000 permutations of small 11-gene signatures derived from the large 1,345-gene MTTs signature and compared their sample-stratification power with that of the 11-gene MTTs/PNS signature. The classification-performance cutoff P values were established by

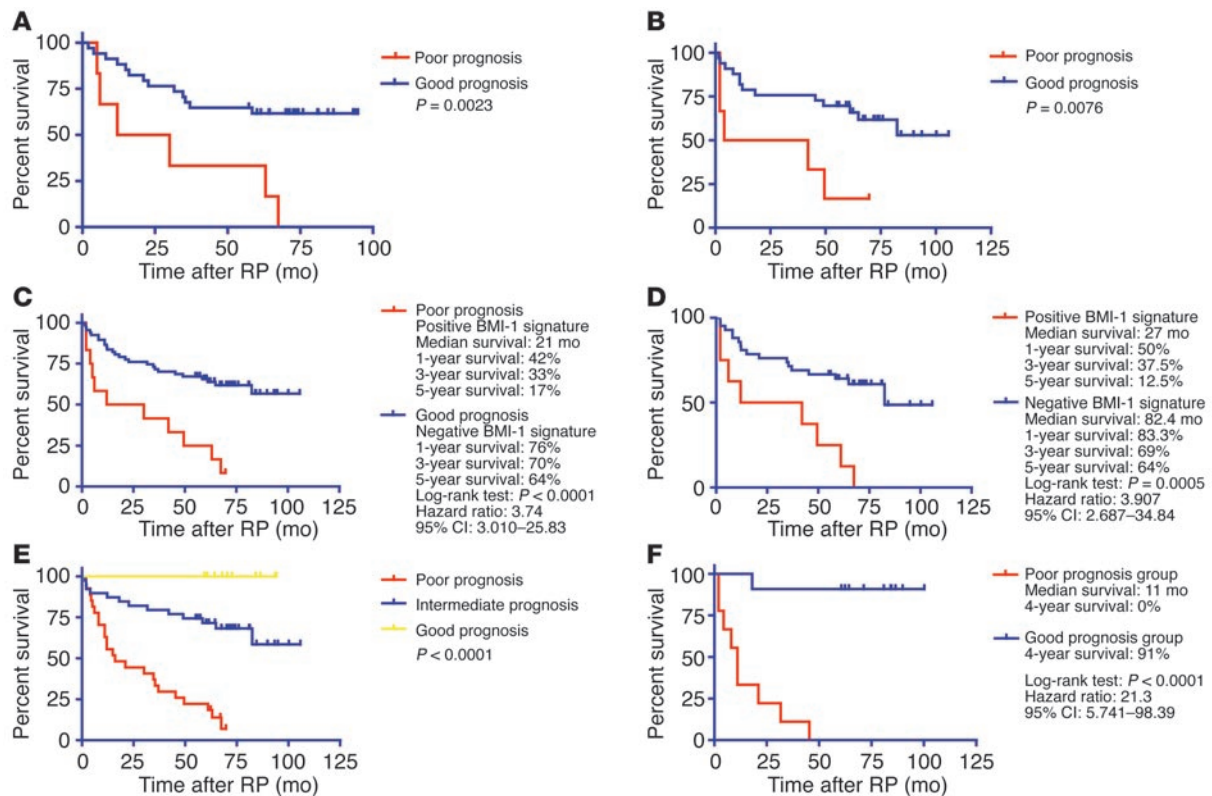


Figure 6

Classification of prostate cancer patients into subgroups with distinct therapy outcome based on expression profile of the 11-gene MTTs/PNS signature. (A–C) Kaplan-Meier analysis of the probability that patients would remain disease-free among 79 prostate cancer patients constituting clinical outcome set 2, according to whether they had a good-prognosis or a poor-prognosis signature as defined by the expression profiles of the 11-gene MTTs/PNS signature. The patients' stratification cutoff value of 0.4 was defined in the training set of 40 patients (19 poor prognosis and 21 good prognosis; A), validated in a test set of 39 patients (18 poor prognosis and 21 good prognosis; B) and confirmed in an entire cohort of 79 patients (C). (D) Kaplan-Meier survival curves for distinct subgroups of prostate cancer patients diagnosed with early-stage disease (stages 1C and 2A). (E) Kaplan-Meier survival curves for 79 prostate cancer patients stratified into distinct subgroups using a weighted survival predictor score algorithm. (F) Kaplan-Meier survival curves for 20 prostate cancer patients stratified into distinct subgroups using Q-RT-PCR assay of the 11-gene signature.

application of a 2-tailed Student's *t* test to the 11-gene MTTs/PNS signature ($P = 0.0005$ for metastasis versus primary prostate cancer data set and $P = 0.026$ for recurrent versus nonrecurrent prostate cancer data set). We found that 10,000 permutations generated 7 random 11-gene signatures performing at the sample-classification level of the 11-gene MTTs/PNS signature.

Cox proportional hazard survival regression analysis. To ascertain the incremental statistical power of the individual covariates as predictors of therapy outcome and unfavorable prognosis, we performed both univariate and multivariate Cox proportional hazard survival analyses (Table 4). Several individual gene members of the 11-gene signature, such as *KI67* and *Cyclin B1*, have been described previously as significant predictors of prognosis and may reflect correlation between proliferative fraction and poor therapy outcome as has been shown recently for the lymphoma survival predictor signature. However, our analysis appears to indicate that the 11-gene signature is a more uniform therapy-outcome predictor across the multiple data sets compared with the individual genes (see below) and, perhaps, is a better “integrator” and “sensor” of the biological diversity across the spectrum of human cancers. We performed both univariate and multivariate Cox proportional hazard survival analyses to compare the prognostic performance of the entire

11-gene signature and individual genes (Table 4 and Supplemental Table 3). In the univariate analysis, prognostic performance of *KI67* expression as a predictor of therapy outcome varied in different outcome data sets. It was highly significant in the prostate cancer therapy outcome set 2 (Memorial Sloan-Kettering Cancer Center [MSKCC] data set); however, it showed only a trend toward statistical significance in the prostate cancer outcome set 1 ($P = 0.1$; Harvard data set) and the breast cancer outcome data set ($P = 0.0533$). In prostate cancer, the significant prognosis predictors in univariate Cox regression analysis were *KI67*, *ANK3*, *FGFR2*, *CES*, and the 11-gene MTTs/PNS signature. In breast cancer, the significant prognosis predictors in univariate analysis were *Cyclin B1*, *BUB1*, *HEC*, and the 11-gene signature. Thus, our analysis seems to indicate that individual genes demonstrate a variable performance across multiple outcome data sets, and we were unable to identify a single gene uniformly predictive of the poor therapy outcome.

In the multivariate analysis (Table 5), the most significant prostate cancer recurrence predictor was the model that included 11 covariates (11-gene signature; 4 individual genes [*KI67*, *ANK3*, *FGFR2*, and *CES1*]; and 6 clinico-pathological features [pre-RP PSA, RP Gleason sum, surgical margins, seminal vesicle invasion, age, and extracapsular extension]). Interestingly, several covari-

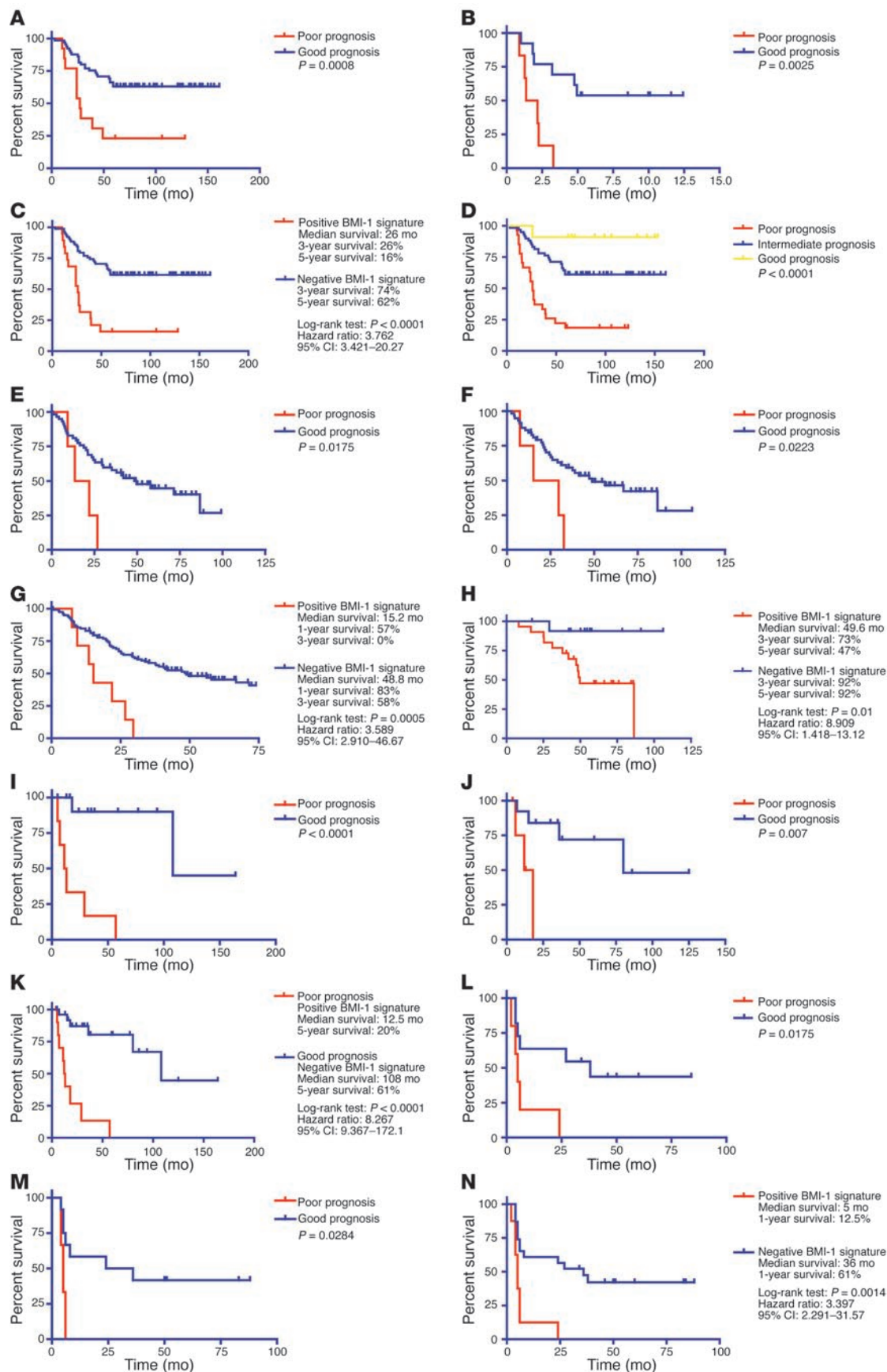


Figure 7

Classification of patients diagnosed with 4 different types of epithelial cancer into subgroups with distinct therapy outcome based on expression profile of the 11-gene MTTs/PNS signature. Kaplan-Meier analysis of the probability that patients would remain metastasis-free (for the breast cancer group) or survive after therapy (for the other groups) among 97 early-stage breast cancer patients (A–D), 125 lung adenocarcinoma patients of all stages (E–G), 35 lung adenocarcinoma patients diagnosed with stage 1A disease (H), 37 ovarian cancer patients of all stages (I–K), and 31 bladder cancer patients (L–N), according to whether they had a good-prognosis or a poor-prognosis signature as defined by the expression profiles of the 11-gene MTTs/PNS signature. For each type of cancer, the patient's stratification cutoff value was defined in the training set, validated in a test set, and confirmed in an entire cohort. D and I–K show the Kaplan-Meier survival curves for 97 breast cancer patients and 37 ovarian cancer patients, respectively, stratified into distinct subgroups using a weighted survival predictor score algorithm.



Table 4
Cox proportional hazard survival regression analysis

Prostate cancer		
Covariates	Statistics	Remarks
<i>Gbx2</i>	$\chi^2 = 1.5817$; df = 1; $P = 0.2085$	
<i>KI67</i>	$\chi^2 = 9.9016$; df = 1; $P = 0.0017$	
<i>cyclin B1</i>	$\chi^2 = 0.1370$; df = 1; $P = 0.7113$	
<i>BUB1</i>	$\chi^2 = 0.9193$; df = 1; $P = 0.3377$	
<i>HEC</i>	$\chi^2 = 2.3450$; df = 1; $P = 0.1257$	
<i>KIAA1063</i>	$\chi^2 = 0.1376$; df = 1; $P = 0.7106$	
<i>HCFC1</i>	$\chi^2 = 2.2379$; df = 1; $P = 0.1347$	
<i>RNF2</i>	$\chi^2 = 1.6235$; df = 1; $P = 0.2026$	
<i>ANK3</i>	$\chi^2 = 8.9237$; df = 1; $P = 0.0028$	
<i>FGFR2</i>	$\chi^2 = 7.7985$; df = 1; $P = 0.0052$	
<i>CES</i>	$\chi^2 = 9.3565$; df = 1; $P = 0.0022$	
Signature	$\chi^2 = 3.9990$; df = 1; $P = 0.0455$	
5 Covariates	$\chi^2 = 26.6628$; df = 5; $P = 0.0001$	Signature + 4 genes
6 Covariates	$\chi^2 = 26.9003$; df = 6; $P = 0.0002$	Signature + 5 genes
11 Covariates	$\chi^2 = 26.9684$; df = 11; $P = 0.0046$	11 Genes
12 Covariates	$\chi^2 = 29.2850$; df = 12; $P = 0.0036$	Signature + 11 genes
11 Covariates	$\chi^2 = 50.7039$; df = 11; $P = 0.0000$	Signature + 4 genes + 6 clinico-pathological features
Breast cancer		
Covariates	Statistics	Remarks
<i>Gbx2</i>	$\chi^2 = 0.0021$; df = 1; $P = 0.9631$	
<i>KI67</i>	$\chi^2 = 3.7357$; df = 1; $P = 0.0533$	
<i>cyclin B1</i>	$\chi^2 = 4.6430$; df = 1; $P = 0.0312$	
<i>BUB1</i>	$\chi^2 = 10.4330$; df = 1; $P = 0.0012$	
<i>HEC</i>	$\chi^2 = 15.6837$; df = 1; $P = 0.0001$	
<i>KIAA1063</i>	$\chi^2 = 0.5386$; df = 1; $P = 0.4630$	
<i>HCFC1</i>	$\chi^2 = 0.7418$; df = 1; $P = 0.3891$	
<i>RNF2</i>	$\chi^2 = 0.0360$; df = 1; $P = 0.8495$	
<i>ANK3</i>	$\chi^2 = 2.5573$; df = 1; $P = 0.1098$	
<i>FGFR2</i>	$\chi^2 = 0.2834$; df = 1; $P = 0.5945$	
<i>CES</i>	$\chi^2 = 0.0477$; df = 1; $P = 0.8272$	
Signature	$\chi^2 = 7.1372$; df = 1; $P = 0.0076$	
4 Covariates	$\chi^2 = 16.4355$; df = 4; $P = 0.0025$	Signature + 3 genes
5 Covariates	$\chi^2 = 16.7995$; df = 5; $P = 0.0049$	Signature + 4 genes
11 Covariates	$\chi^2 = 28.7740$; df = 11; $P = 0.0025$	11 Genes
12 Covariates	$\chi^2 = 29.3656$; df = 12; $P = 0.0035$	Signature + 11 genes

df, degrees of freedom.

ates, such as the 11-gene signature, *KI67*, *CES1*, pre-RP prostate-specific antigen (PSA) level, surgical margins, and extracapsular extension, remained statistically significant prognostic markers in the multivariate analysis (Table 5). Thus, while prognostic performance of individual gene members of the 11-gene signature varied greatly in different outcome data sets, the identified 11-gene signature seems to perform as the most consistent predictor of poor therapy outcome across multiple independent outcome data sets comprising over 1,000 clinical samples and representing 11 distinct types of human cancer (see below). Yet the statistically best-performing multivariate cancer type-specific model seems to require a combination of calls based on expression levels of individual genes, a gene expression signature, and clinico-pathological covariates (Tables 4 and 5).

We sought to use an alternative statistical metric to further evaluate the prognostic power of the genes constituting the 11-gene signature. We implemented the weighted survival score analysis to reflect the incremental statistical power of the individual

covariates as predictors of therapy outcome based on a multicomponent prognostic model (Figure 6E). Final survival predictor score comprises a sum of scores for individual genes and reflects the relative contribution of each of the 11 genes in the multivariate analysis. The negative weighting values imply that higher expression correlates with longer survival and favorable prognosis, whereas the positive score values indicate that higher expression correlates with poor outcome and shorter survival. Application of the weighted survival predictor model based on a cumulative score of the weighted expression values of 11 genes confirmed the prognostic power of the identified 11-gene signature in stratification of prostate cancer patients into subgroups with statistically distinct probability of relapse-free survival after RP (Figure 6E).

Expression of the 11-gene MTTs/PNS signature is a predictor of a short relapse-free survival after therapy in prostate cancer patients with an early-stage disease. Identification of patients with high likelihood of poor outcome after therapy would be particularly desirable in a cohort of patients diagnosed with a seemingly localized early-stage prostate cancer. Next we determined whether the 11-gene MTTs/PNS signature would be useful in defining subgroups of patients diagnosed with an early-stage prostate cancer and having a statistically significant difference in the likelihood of disease relapse after therapy. In the group of patients diagnosed with stage 1C or 2A prostate cancer (Figure 6D), the median relapse-free survival after therapy in the poor-prognosis subgroup defined by the 11-gene BMI-1-pathway signature was 27 months. In contrast, the median relapse-free survival after therapy in the good-prognosis group was 82.4 months. Eighty-eight percent of patients in the poor-prognosis subgroup had a disease recurrence within 5 years after therapy. Conversely, 64% of patients in the good-prognosis subgroup remained relapse-free (Figure 6D). The estimated hazard ratio for disease recurrence after therapy in the poor-prognosis subgroup as compared with the good-prognosis subgroup of patients defined by the 11-gene signature was 3.907 (95% confidence interval of ratio, 2.687–34.84; $P = 0.0005$).

Validation of the prognostic performance of the 11-gene BMI-1-pathway signature using a Q-RT-PCR-based assay. Routine clinical use of prognostic tests based on microarray-derived gene expression signatures would require prospective validation study of the utility of identified markers in an experimental setting highly compatible with state-of-the-art clinical laboratory practice. Since microarray-based assay format is not readily available for application in the clinical laboratory, we considered the Q-RT-PCR-based test as an alternative clinically compatible analytical platform suitable for measurements of mRNA expression level of marker genes. Expression of mRNA for 11 genes (Supplemental Table 1) and an endogenous control gene (*GAPDH*) was measured by real-time PCR in 20 specimens of primary prostate cancer obtained from patients with documented PSA recurrence within 5 years after RP and patients who remained disease-free for at least 5 years after RP (10 patients



Table 5
11-Covariate prostate cancer recurrence predictor model

Covariate	Coefficient	SE	Significance, <i>P</i>	Confidence interval, low 95%	Confidence interval, high 95%
Signature	-2.3537	0.9858	0.0170	-4.2858	-0.4215
<i>KI67</i>	2.2832	0.7823	0.0035	0.7499	3.8166
<i>ANK3</i>	-0.1563	0.7197	0.8280	-1.5670	1.2543
<i>FGFR2</i>	-0.8295	0.4955	0.0941	-1.8007	0.1418
<i>CES</i>	-1.6403	0.8113	0.0432	-3.2303	-0.0502
Pre-RP PSA	0.0493	0.0251	0.0495	0.0001	0.0985
RP GLSN SUM	0.2850	0.2385	0.2322	-0.1825	0.7525
SM	1.0609	0.4648	0.0225	0.1499	1.9720
Sem Ves Inv	0.6016	0.5064	0.2348	-0.3909	1.5941
Age	0.0311	0.0351	0.3755	-0.0377	0.0999
ECE	0.9296	0.4360	0.0330	0.0751	1.7842

RP, radical prostatectomy; PSA, prostate-specific antigen; GLSN SUM, Gleason sum; SM, surgical margins; Sem Ves Inv, seminal vesicle invasion; ECE, extracapsular extension.

in each group). As shown in Figure 6F, a prostate cancer therapy outcome test based on measurements of mRNA expression levels of 11 genes using Q-RT-PCR discriminates prostate cancer patients into subgroups with statistically distinct probability of relapse-free survival after RP.

The Kaplan-Meier survival analysis demonstrated that application of the 11-gene Q-RT-PCR-based prostate cancer therapy outcome test segregates prostate cancer patients into subgroups with statistically significant difference in the probability of remaining relapse-free after the therapy (Figure 6F). The estimated hazard ratio for disease recurrence after therapy in the poor-prognosis group of patients as compared with the good-prognosis group defined by the test was 21.3 (95% confidence interval of ratio, 5.741–98.39; $P < 0.0001$). One hundred percent of patients in the poor-prognosis group had a disease recurrence within 4 years after RP, whereas 91% of patients in the good-prognosis group remained relapse-free (Figure 6F).

Expression of the 11-gene MTTs/PNS signature predicts metastatic recurrence and poor survival after therapy in breast cancer and lung adenocarcinoma patients diagnosed with an early-stage disease. BMI-1 expression was previously implicated in human breast and lung cancers (11, 12, 31), which suggests that activation of BMI-1-associated pathway(s) might be relevant to these types of carcinomas as well. We therefore sought to investigate whether measurements of expression of the 11-gene MTTs/PNS signature would be informative in the prediction of the patients' prognosis in a group of 97 young women diagnosed with sporadic lymph node-negative early-stage breast cancer who were analyzed in a recent expression profiling study of early-stage breast cancer (33). (This group comprises 46 patients who developed distant metastases within 5 years and 51 patients who continued to be disease-free at least 5 years after therapy; they constitute clinically defined poor-prognosis and good-prognosis groups, correspondingly.) Kaplan-Meier analysis indicates that breast cancer patients with tumors displaying a stem cell-like expression profile of the 11-gene signature have a significantly higher probability of developing distant metastases within 5 years after therapy and therefore can be identified as a poor-prognosis subgroup (Figure 7, A–D). Median metastasis-free survival after therapy in the poor-prognosis subgroup of breast cancer patients defined by the 11-gene

signature was 26 months. Eighty-four percent of patients in the poor-prognosis subgroup were diagnosed with distant metastasis within 5 years after therapy (Figure 7C). In contrast, 62% of patients in the good-prognosis subgroup remained metastasis-free (Figure 7C). The estimated hazard ratio for metastasis-free survival after therapy in the poor-prognosis subgroup as compared with the good-prognosis subgroup of patients defined by the 11-gene signature was 3.762 (95% confidence interval of ratio, 3.421–20.27; $P < 0.0001$). Thus, the expression pattern of the 11-gene MTTs/PNS signature is strongly predictive of a short postdiagnosis and post-treatment interval to distant metastases in early-stage breast cancer patients.

Next we asked whether expression analysis of the 11-gene signature would be informative in patients' stratification into sub-

groups with distinct survival probability after therapy in a group of 125 patients diagnosed with lung adenocarcinoma (34). Similarly to the prostate and breast cancer patients, the Kaplan-Meier analysis shows that patients with tumors displaying a stem cell-like expression profile of the 11-gene signature have significantly higher risk of death after therapy and therefore can be defined as a poor-prognosis subgroup (Figure 7, E–H). Median survival after therapy in the poor-prognosis subgroup of lung adenocarcinoma patients defined by the 11-gene BMI-1-pathway signature was 15.2 months (Figure 7G). In contrast, the median survival after therapy in the good-prognosis subgroup was 48.8 months. One hundred percent of patients in the poor-prognosis subgroup died within 3 years after therapy. Conversely, 58% of patients in the good-prognosis subgroup remained alive (Figure 7G). The estimated hazard ratio for death after therapy in the poor-prognosis subgroup as compared with the good-prognosis subgroup of patients defined by the 11-gene signature was 3.589 (95% confidence interval of ratio, 2.910–46.67; $P = 0.0005$).

Next we examined whether the 11-gene MTTs/PNS signature would be useful in defining subgroups of patients diagnosed with an early-stage lung adenocarcinoma and having a statistically significant difference in survival probability after therapy. In the group of patients diagnosed with stage 1A lung adenocarcinoma (Figure 7H), the median survival after therapy in the poor-prognosis subgroup defined by the 11-gene signature was 49.6 months. Fifty-three percent of patients in the poor-prognosis subgroup died within 5 years after therapy. In contrast, 92% of patients remained alive in the good-prognosis subgroup (Figure 7H). The estimated hazard ratio for death after therapy in the poor-prognosis subgroup as compared with the good-prognosis subgroup of patients defined by the 11-gene signature was 8.909 (95% confidence interval of ratio, 1.418–13.12; $P = 0.01$).

Based on this analysis, we concluded that detection of a stem cell-like expression profile of the 11-gene MTTs/PNS signature in primary tumors from patients diagnosed with early-stage prostate, breast, and lung carcinomas is associated with a high propensity toward metastatic dissemination and significantly higher risk of poor therapy outcome. Interestingly, therapy outcome in cancer patients diagnosed with other types of epithelial

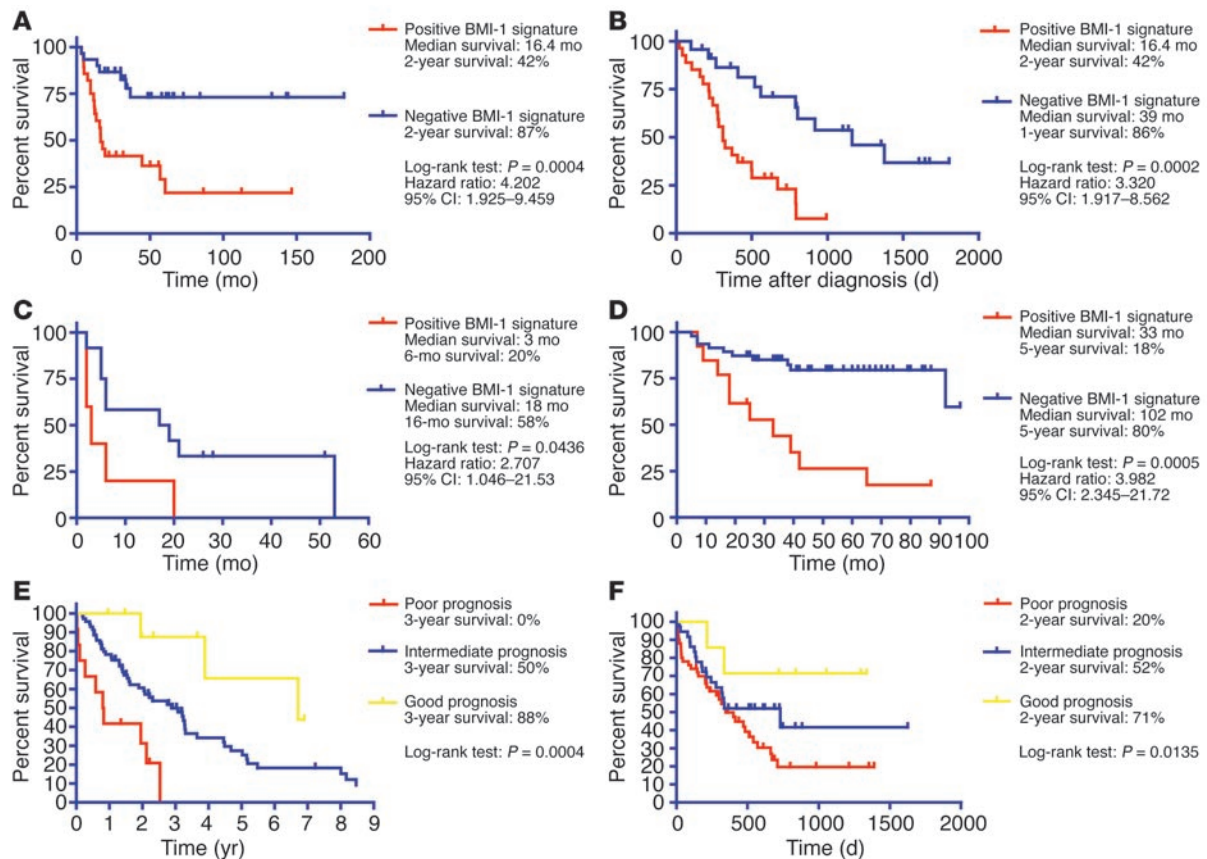


Figure 8

Classification of cancer patients diagnosed with different types of nonepithelial malignancies into subgroups with distinct therapy outcome based on expression profile of the 11-gene MTTs/PNS signature. Kaplan-Meier survival analysis of the probability of a therapy failure in cancer patients diagnosed with different types of nonepithelial cancers and having distinct expression profiles of the 11-gene MTTs/PNS signature is shown. Data from lymphoma patients (A), malignant glioma patients (B), mesothelioma patients (C), medulloblastoma patients (D), mantle cell lymphoma patients (E), and acute myeloid leukemia patients (F) are shown.

cancers, such as ovarian and bladder cancers, seems to manifest similar association with distinct patterns of expression of the 11-gene signature (Figure 7, I–N).

Expression of the 11-gene signature predicts therapy outcome in patients diagnosed with nonepithelial malignancies. Altered BMI-1 expression was implicated recently in several nonepithelial malignancies, such as B cell non-Hodgkin lymphoma (35) and pediatric brain tumors (36). We therefore sought to analyze whether the 11-gene MTTs/PNS signature would be useful in defining subgroups of patients diagnosed with nonepithelial cancers and having a statistically significant difference in survival probability after therapy. Using Kaplan-Meier method, we analyzed the prognostic power of the 11-gene signature in patients diagnosed with diffuse large B cell lymphoma, mantle cell lymphoma, acute myeloid leukemia, mesothelioma, medulloblastoma, and glioma (Table 2). Kaplan-Meier analysis demonstrates that a stem cell-like expression profile of the 11-gene signature in primary tumors is a consistent powerful predictor of a therapy failure and short survival in cancer patients diagnosed with 5 distinct types of nonepithelial cancers (Figure 8, A–F). Consistent with our findings, an increased BMI-1 expression in human medulloblastomas was demonstrated in a recent study (37). Taken together, these data seem to imply the presence of a conserved BMI-1-associated pathway(s) similarly engaged in both

neural stem cells and a highly malignant subset of human cancers diagnosed in a wide range of organs and uniformly exhibiting a marked propensity toward metastatic dissemination as well as a high probability of unfavorable therapy outcome.

Discussion

A growing number of expression profiling studies provide experimental evidence indicating the presence of a transcriptionally distinct subtype of human solid tumors manifesting a marked propensity toward metastatic dissemination, highly malignant clinical behavior, and a high probability of poor therapy outcome in cancer patients diagnosed with early-stage carcinomas of various origins (refs. 15, 38–40; this study). These results are consistent with the idea that, at least in a subset of human solid tumors, the acquisition of full metastatic potential, including an emergence and seeding of potent metastasis precursor cells, is a relatively early event in tumor progression. Collectively, these data suggest an early involvement, in development of this transcriptionally defined subtype of human carcinomas, of a highly malignant combination of mutant alleles conferring the proclivity to metastasize (40) and/or an engagement of unique unconventional cellular targets such as stem cells and/or early progenitor cells in transformation and tumor progression.



One of the hallmark biological features of normal stem cells is the ability to fuse spontaneously *in vitro* and *in vivo* with other cell types, leading to formation of reprogrammed viable somatic cell hybrids (41–44). Accumulation of normal stem cells in experimental tumors *in vivo* has been demonstrated in several studies (45, 46). Furthermore, most recent studies demonstrated that committed myelomonocytic cells such as macrophages can produce functional epithelial cells by *in vivo* fusion (47), thus extending the number of cell types that might serve as hypothetically “eligible” fusion partners for tumor cells. It would be of interest to study how cancer cells co-opt stem cell-like transcriptome into progression pathways and whether some human carcinomas could attract stem cells by mimicking a stem cell “niche” microenvironment, thus directly engaging normal stem cells into malignant process via cell fusion. One interesting endpoint of our analysis is that a relatively small set of coregulated transcripts appears to predict clinical outcome in a large number of human tumors representing 10 distinct types of cancer. Perhaps inclusion of a relevant biological model in the signature discovery protocol was an essential component of the successful hit selection, since recent metaanalysis of cancer microarray data that was based solely on statistical approaches did not identify an outcome signature common to multiple cancer types (48).

It has been suggested that sets of coordinately expressed genes defined as cancer-associated gene expression signatures might reflect the cell of origin of cancer (49). Unlike stem cells in the state of CNS neurospheres that are recovered from the CNS, stem cells in the state of PNS neurospheres might be present in many (if not all) peripheral tissues and therefore are more likely and readily accessible cellular targets for direct involvement into malignant process. It remains to be elucidated whether the precision of analytical protocols used in this study was sufficient to identify the broadly applicable gene expression markers of the *BMI-1*-pathway activation and normal stem cell engagement in malignant progression of human cancers. Protein products of 2 genes upregulated in the MTTs/PNS signature profile (BUB1 and HEC1) are known to play an important role in the spindle assembly mitotic checkpoint. A recent study suggested a novel mechanism leading to development of frequent aneuploidy in human cancer due to aberrant expression of Mad2 protein, inappropriate activation of the spindle checkpoint, and, eventually, aneuploidy (50). Both BUB1 and HEC1 proteins play a key role in the assembly of checkpoint proteins, being required for Mad2 recruitment to the kinetochores (51, 52); this suggests that aberrant BUB1 and HEC1 expression might contribute to and/or reflect the altered function of the mitotic checkpoint in metastatic cancer cells.

In conclusion, using a mouse/human comparative translational genomics approach, we identified an 11-gene signature that consistently displays a stem cell-like expression pattern in metastatic lesions of prostate carcinomas recovered from multiple distant target organs. Our results indicate that a stem cell-resembling expression profile of the 11-gene signature is associated with a highly malignant clinical course of disease progression and predicts high likelihood of therapy failure in multiple types of human cancer. Statistically significant negative prognostic value of a stem cell-like expression of the 11-gene signature in early-stage primary solid tumors of diverse origin suggests the presence of a genetically distinct subtype of human carcinomas with high propensity toward metastatic dissemination

even at the early stage of disease progression. Further elucidation of possible causal relationships between activation of a stem cell-resembling gene expression program and malignant behavior of human carcinoma cells should have considerable theoretical and practical implications.

Methods

Clinical samples. Expression profiling data of primary tumor samples obtained from 1,122 cancer patients representing therapy-outcome cohorts for 10 types of human cancer (Table 2) were analyzed in this study. Microarray analysis and associated clinical information for 32 clinical samples (23 primary prostate tumors and 9 distant metastatic lesions) used to delineate the expression profiles of human prostate cancer metastases were reported previously (13). Two clinical outcome sets comprising 21 (outcome set 1) and 79 (outcome set 2) samples were used for analysis of the association of the therapy outcome with distinct expression profiles of the 11-gene signature. Original gene expression profiles of the 21 clinical samples analyzed in this study were reported elsewhere (14). Primary gene expression data files of clinical samples as well as associated clinical information can be found at <http://www-genome.wi.mit.edu/cancer/>.

The prostate tumor tissues constituting the second clinical outcome set were obtained from 79 prostate cancer patients undergoing therapeutic or diagnostic procedures performed as part of routine clinical management at the MSKCC (New York, New York, USA). Clinical and pathological features of 79 prostate cancer cases constituting the validation outcome set are presented elsewhere (15). Median follow-up after therapy in this cohort of patients was 70 months. Samples were snap-frozen in liquid nitrogen and stored at -80°C . Each sample was examined histologically using H&E-stained cryostat sections. Care was taken to remove nonneoplastic tissues from tumor samples. Cells of interest were manually dissected from the frozen block and other tissues trimmed away. All of the studies were approved by the MSKCC Institutional Review Board.

Expression analysis data for tumor samples obtained from 125 lung adenocarcinoma patients as well as associated clinical information were reported elsewhere (34). Original work describing gene expression profiles of the set of 97 clinical samples of early-stage breast cancer was reported elsewhere (33). Primary gene expression data files of clinical samples as well as associated clinical information have been previously described (33). To date, our analysis includes 1,153 therapy-outcome samples from patients diagnosed with 11 distinct types of cancer (Table 2): prostate cancer (100 patients); breast cancer (97 patients); lung adenocarcinoma (211 patients); ovarian cancer (50 patients); bladder cancer (31 patients); diffuse large B cell lymphoma (298 patients); mantle cell lymphoma (MCL, 92 patients); mesothelioma (17 patients); medulloblastoma (60 patients); glioma (50 patients); and acute myeloid leukemia (116 patients).

Cell culture. Cell lines used in this study were previously described (16). The LNCap- and PC-3-derived cell lines were developed by consecutive serial orthotopic implantation, either from metastases to the lymph node (for the LN series) or reimplanted from the prostate (Pro series). This procedure generated cell variants with differing tumorigenicity, frequency, and latency of regional lymph node metastasis (16). Except where noted, cell lines were grown in RPMI1640 supplemented with 10% FBS and gentamycin (GIBCO; Invitrogen Corp.) to 70–80% confluence and subjected to serum starvation as described previously (16), or maintained in fresh complete media, supplemented with 10% FBS.

Anoikis assay. Cells were harvested by 5-minute digestion with 0.25% trypsin/0.02% EDTA (Irvine Scientific), washed, and resuspended in serum-free medium. Cells at a concentration of 1.7×10^5 cells per well in 1 ml of serum-free medium were plated in 24-well Ultra Low Attach-



ment polystyrene plates (Corning Inc.) and incubated at 37°C and 5% CO₂ overnight. Viability of cell cultures subjected to anoikis assays was greater than 95% in a trypan blue dye exclusion test.

Apoptosis assay. Apoptotic cells were identified and quantified using the Annexin V-FITC kit (BD Biosciences — Pharmingen) according to the manufacturer's instructions. The following controls were used to set up compensation and quadrants: (a) unstained cells; (b) cells stained with Annexin V-FITC (no propidium iodide); and (c) cells stained with propidium iodide (no Annexin V-FITC). Each measurement was carried out in quadruplicate, and each experiment was repeated at least twice. Annexin V-FITC-positive cells were scored as early apoptotic cells; both Annexin V-FITC- and propidium iodide-positive cells were scored as late apoptotic cells; unstained Annexin V-FITC- and propidium iodide-negative cells were scored as viable or surviving cells. In selected experiments, apoptotic cell death was documented using the TUNEL assay.

Flow cytometry. Cells were washed in cold PBS and stained using the Annexin V-FITC Apoptosis Detection Kit (BD Biosciences) according to the manufacturer's instructions. Flow analysis was performed with a FACSCalibur instrument (BD Biosciences). CellQuest software (BD Biosciences) was used for data acquisition and analysis. All measurements were performed under the same instrument setting, analyzing 10³–10⁴ cells per sample.

Orthotopic xenografts. Orthotopic xenografts of human prostate PC-3 cells and sublines used in this study were developed by surgical orthotopic implantation as previously described (15, 16). Briefly, 2 × 10⁶ cultured PC-3 cells or cells of the PC-3M or PC-3MLN4 subline were injected s.c. into male athymic mice, and allowed to develop into firm palpable and visible tumors over the course of 2–4 weeks. Intact tissue was harvested from a single s.c. tumor and surgically implanted in the ventral lateral lobes of the prostate gland in a series of 6 athymic mice per cell line subtype as described earlier (16).

Transgenic mouse model of prostate cancer. A breeding colony of TRAMP (transgenic adenocarcinoma of the mouse prostate) mice is maintained on C57BL/6 background in the Animal Care Facility at the Sidney Kimmel Cancer Center (53). The TRAMP mouse colony is based on a breeding pair of TRAMP mice kindly provided by Norman Greenberg (Baylor College of Medicine, Houston, Texas, USA). Standard PCR assay was carried out to monitor the presence of the SV-40 large T antigen in new litters. Twenty-one PCR-confirmed male TRAMP mice were defined for microarray analysis carried out in this study. Animals were killed at different ages according to the established time course of the disease progression (32), and prostates as well as primary and metastatic tumors were immediately removed and snap-frozen in liquid nitrogen. Prostate tissues from age-matched wild-type C57BL/6 mice served as control samples in our microarray analysis of the TRAMP model of prostate cancer. Necropsies with gross microscopic examination were carried out. All procedures were approved by the Sidney Kimmel Cancer Center Institutional Animal Care and Use Committee and followed Sidney Kimmel Cancer Center Standard Operating Procedures in accordance with the NIH Guide for the Care and Use of Laboratory Animals.

Tissue processing for mRNA and RNA isolation. Fresh-frozen orthotopic and transgenic primary tumors, metastases, and mouse prostates were examined by use of H&E-stained frozen sections. Orthotopic tumors of all sublines exhibited similar morphology consisting of sheets of monotonous closely packed tumor cells with little evidence of differentiation, interrupted by only occasional zones of largely stromal components, vascular lakes, or lymphocytic infiltrates. Fragments of tumor judged free of these nonepithelial clusters were used for mRNA preparation. Frozen tissue (1–3 mm × 1–3 mm) was submerged in liquid nitrogen in a ceramic mortar and ground to powder. The frozen tissue powder

was dissolved and immediately processed for mRNA isolation using a FastTrack kit for mRNA extraction (Invitrogen Corp.) according to the manufacturer's instructions.

RNA and mRNA extraction. For gene expression analysis, cells were harvested in lysis buffer 2 hours after the last media change at 70–80% confluence, and total RNA or mRNA was extracted using the RNeasy (QIAGEN) or FastTrack kit (Invitrogen Corp.). Cell lines were not split more than 5 times before RNA extraction, except where noted.

Affymetrix arrays. The protocol for mRNA quality control and gene expression analysis was that recommended by Affymetrix. In brief, approximately 1 µg of mRNA was reverse-transcribed with an oligo-dT primer that has a T7 RNA polymerase promoter at the 5' end. Second-strand synthesis was followed by complementary RNA (cRNA) production incorporating a biotinylated base. Hybridization to Affymetrix U95Av2 arrays representing 12,625 transcripts overnight for 16 hours was followed by washing and labeling using a fluorescently labeled antibody. The arrays were read and data processed using Affymetrix equipment and software as reported previously (13, 15, 16).

Data analysis. Detailed protocols for data analysis and documentation of the sensitivity, reproducibility, and other aspects of the quantitative statistical microarray analysis using Affymetrix technology have been reported previously (15, 16). Forty to fifty percent of the surveyed genes were called present by Affymetrix Microarray Suite version 5.0 software in these experiments. The concordance analysis of differential gene expression across the data sets was performed using Affymetrix MicroDB version 3.0 and DMT version 3.0 software as described previously (13, 15, 16). We processed the microarray data using the Affymetrix Microarray Suite version 5.0 software and performed statistical analysis of the expression data set using the Affymetrix MicroDB and Affymetrix DMT software. The Pearson correlation coefficient for individual test samples and the appropriate reference standard were determined using Microsoft Excel version 2002 (Microsoft Corp.) and GraphPad Prism version 4.00 software (GraphPad Software). We calculated the significance of the overlap between the lists of stem cell-associated and prostate cancer-associated genes by using the hypergeometric distribution test (54). Analytical protocol of identification and validation of the 11-gene BMI-1-pathway signature is described below and presented in Figure 3. We used MultiExperiment Viewer (MEV) software version 3.0.3 of the Institute for Genomic Research for SVM classification and TRN clustering algorithm data analysis and visualization.

Protocol of discovery and validation of the 11-gene BMI-1-pathway signature. We hypothesized that molecular signatures associated with activation of a normal stem cell's self-renewal and/or survival program in metastatic cancer cells might be detectable by looking for genes manifesting concordant patterns of regulation in distant metastatic lesions and stem cells in *Bmi-1*^{+/+} versus *Bmi-1*^{-/-} genetic backgrounds. Therefore, we sought to determine whether expression profiles of transcripts activated and suppressed in prostate cancer metastases would recapitulate the expression profile of the BMI-1-regulated genes in neural stem cells, by comparing the sets of differentially regulated genes in search of intersection of lists for both up- and downregulated transcripts. Thus, according to this model the primary criterion for transcript selection 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 coregulated genes, reflecting a balance of up- and downregulated mRNAs as a desired regulatory endpoint 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). Samples with stem cell-resembling expression profiles



of PAI (SPAI) are expected to have positive values of Pearson correlation coefficients. Detailed prognostic signature identification and validation protocol are described below and shown in Figure 3.

Step 1. Sets of differentially regulated transcripts were independently identified for distant metastatic lesions and primary prostate tumors versus age-matched control samples in a transgenic TRAMP mouse model of metastatic prostate cancer (MTTS signature) as well as PNS (PNS signature) and CNS (CNS signature) neurospheres in *BMI-1*^{+/+} versus *BMI-1*^{-/-} backgrounds using the Affymetrix microarray processing and statistical analysis software package (Affymetrix Microarray Suite version 5.0, MicroDB version 3.0, and DMT version 3.0) as described above and in previous publications (15, 16). Transcripts with negative signal-intensity values in both experimental and control sets were eliminated from further consideration. At least 2-fold changes of the mRNA abundance levels in experimental versus control samples for both upregulated and downregulated genes were required for inclusion in the lists of differentially regulated transcripts. Fold expression changes of the mRNA abundance levels for each transcript were calculated as ratios of the average intensity values for a given transcript in experimental versus control samples for both upregulated and downregulated genes and log₁₀-transformed for further analysis. Thus, this analytical step defined 3 large parent signatures (see Figure 3): MTTS signature comprising 868 upregulated and 477 downregulated transcripts; PNS signature comprising 885 upregulated and 1,088 downregulated transcripts; and CNS signature comprising 769 upregulated and 778 downregulated transcripts.

Step 2. Subsets of transcripts exhibiting concordant expression changes in metastatic TRAMP tumor samples (MTTS signature) as well as PNS (PNS signature) and CNS (CNS signature) neurospheres in *BMI-1*^{+/+} versus *BMI-1*^{-/-} backgrounds were identified. Concordant lists of transcripts were obtained by intersecting the 2 lists each of upregulated and downregulated genes. Thus, 2 concordant subsets of transcripts were identified corresponding to each binary comparison of metastatic TRAMP tumors and neural stem cell samples in a state of PNS and CNS neurospheres (141 upregulated and 58 downregulated transcripts for PNS neurospheres [$r = 0.7593$; $P < 0.0001$; Figure 2A] and 40 upregulated and 24 downregulated transcripts for CNS neurospheres [$r = 0.7679$; $P < 0.0001$]). A third concordant subset of 27 genes comprising 15 upregulated and 12 downregulated transcripts was selected for intersection common to all 3 signatures ($r = 0.8002$; $P < 0.0001$).

Step 3. Selection of small gene clusters was performed from subsets of genes exhibiting concordant changes of transcript-abundance behavior in metastatic TRAMP tumor samples and PNS and CNS neurospheres in *BMI-1*^{+/+} versus *BMI-1*^{-/-} backgrounds. Expression profiles were presented as log₁₀ average fold changes for each transcript and processed for visualization and Pearson correlation analysis using Microsoft Excel software (Microsoft Corp.). For the concordant differentially expressed genes, vectors of log₁₀ average fold change were determined for both experimental settings, and the correlation between 2 vectors was determined. Practical considerations essential for future development of genetic diagnostic tests prompted us to select from concordant gene sets small gene expression signatures comprising transcripts with a high level of expression correlation in metastatic cancer cells and stem cells. The concordant list of differentially expressed genes was reduced by removing those genes whose removal led to the largest increase in the correlation coefficient. The reduction in the signature transcript number was terminated when further elimination of a transcript did not increase the value of the Pearson correlation coefficient. The cutoff criterion for signature reduction was arbitrarily set to exceed a Pearson correlation coefficient of 0.95 ($P < 0.0001$). Using this approach, a single candidate prognostic gene expression signature was selected for each intersection of the MTTS signature and parent stem cell

signatures (Figure 3). Thus, 3 highly concordant small signatures were identified corresponding to 3 concordant subsets of genes defined in step 2 (a set of 11 genes comprising 8 upregulated and 3 downregulated transcripts for PNS neurospheres, i.e., the 11-gene MTTS/PNS signature; a set of 11 genes comprising 7 upregulated and 4 downregulated transcripts for CNS neurospheres, i.e., the 11-gene MTTS/CNS signature; and a set of 14 genes comprising 8 upregulated and 6 downregulated transcripts, i.e., the MTTS/PNS/CNS signature).

Step 4. The small signatures identified in step 3 (one 11-gene signature for the PNS set, one 11-gene signature for the CNS set, and one 14-gene signature for the common PNS/CNS set) were tested for the power to discriminate the metastatic phenotype (using 1 mouse prostate cancer data set and 1 human prostate cancer data set comprising primary and metastatic tumors) and therapy-outcome classification performance (using human prostate cancer therapy outcome set 1). Three identified small signatures were evaluated for their ability to discriminate metastatic and primary prostate tumors in a TRAMP mouse model of prostate cancer, and clinical samples of 9 metastatic versus 23 primary prostate tumors as well as primary prostate tumors from 21 patients with distinct outcome after the therapy (8 recurrent and 13 nonrecurrent samples). To assess a potential diagnostic and prognostic relevance of small signatures, we calculated a Pearson correlation coefficient for each individual tumor sample by comparing the expression profiles of individual samples with the reference expression profile in either PNS or CNS neurospheres in *BMI-1*^{+/+} versus *BMI-1*^{-/-} backgrounds. 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 clinical samples. For each data set, the vector (X) of average gene expression was determined, and then the relative expression vector (R) was determined for each sample ($R = X/X$). The relative expression vectors were log₁₀-transformed and correlated with the fixed vectors of gene expression determined in step 3. Negative expression values were treated as missing data. Based on the expected correlation of expression profiles of identified gene clusters with stem cell-like expression profiles, we named the corresponding correlation coefficients calculated for individual samples the SPAIs. We evaluated the prognostic power of identified small signatures based on their ability to discriminate metastatic versus primary tumors (criterion 1) and to segregate the patients with recurrent and nonrecurrent prostate tumors into distinct subgroups (criterion 2) and selected a single best-performing small signature for subsequent validation analysis (Figures 3 and 4). Based on diagnostic and prognostic classification performance, a single best-performing 11-gene MTTS/PNS signature was selected for further validation analysis (Figures 3 and 4). The fixed numerical vector describing the stem cell-like signature in 11 genes is shown in Supplemental Tables 2, 3, and 5.

Step 5. To assess the incremental statistical power of the individual genetic and clinical covariates as predictors of therapy outcome and unfavorable prognosis in prostate cancer patients, we performed both univariate and multivariate Cox proportional hazard survival analyses (Table 4).

Step 6. To validate a survival prediction model based on the 11-gene MTTS/PNS signature, we tested the prognostic performance of the model in the multiple independent therapy-outcome data sets representing 5 epithelial and 5 nonepithelial cancers. We divided the patients within individual cohorts into a training set, which was used to select the cutoff threshold and to test the model, and a test set, which was used to evaluate the reproducibility of the classification performance. We used the training set to select the prognosis-discrimination cutoff value for a signature based on the highest level of statistical significance in patients' stratification into poor- and good-prognosis groups as determined by the log-rank test (lowest P value and highest hazard ratio in the training set).



Clinical samples having the Pearson correlation coefficient at or higher than the cutoff value were identified as having the poor-prognosis signature. Clinical samples with the Pearson correlation coefficient below the cutoff value were identified as having the good-prognosis signature. Each training set was used to estimate a threshold of the correlation coefficients before a survival analysis was performed. These thresholds are shown in the legends to Figures 6 and 7. The same discrimination cutoff value was then applied to evaluate the reproducibility of the prognostic performance in the test set of patients. Lastly, we applied the model to the entire outcome set using the same cutoff threshold to confirm the classification performance. The average gene expression vectors were determined for each gene and applied separately on the training, test, and combined data sets. The training and test sets were balanced with respect to the total number of patients, negative and positive therapy outcomes, and the length of survival. For the breast cancer data set, we maintained the patients' distribution among training and test data sets described in the original publication (33). At this stage of the analysis, we did not carry out additional model training, development, or optimization steps, except for selection of a prognostic cutoff threshold in the training set. The same MTTS/PNS expression profile was consistently used throughout the study as a reference standard to quantify the Pearson correlation coefficients of the individual samples. The distribution of the patients in the training and validation sets as well as the threshold values that separate the samples into good- and poor-prognostic groups are shown in the legends to Figures 6 and 7.

The cohort of 97 breast cancer patients (shown together in Figure 7C) was divided into a training set (78 patients; 34 poor prognosis and 44 good prognosis) (Figure 7A) and a test set (19 patients; 12 poor prognosis and 7 good prognosis) (Figure 7B) as described in the original publication (33). One hundred twenty-five lung cancer patients (shown together in Figure 7G) were divided into a training set (63 patients; 36 poor prognosis and 27 good prognosis) (Figure 7E) and a test set (62 patients; 35 poor prognosis and 27 good prognosis) (Figure 7F). Thirty-seven ovarian cancer patients (shown together in Figure 7K) were divided into a training set (19 patients; 8 poor prognosis and 11 good prognosis) (Figure 7I) and a test set (18 patients; 7 poor prognosis and 11 good prognosis) (Figure 7J). Thirty-one bladder cancer patients (shown together in Figure 7N) were divided into a training set (16 patients; 11 poor prognosis and 5 good prognosis) (Figure 7L) and a test set (15 patients; 10 poor prognosis and 5 good prognosis) (Figure 7M). The cutoff values used for classification were 0.56, 0.75, 1.16, and 0.35 for breast, lung, ovarian, and bladder cancer data sets, respectively.

Step 7. We tested the model performance using various sample-stratification approaches, such as TRN clustering (Figure 5), SVM classification (Supplemental Table 4), and weighted survival score algorithm (Figures 6E and 7D). We evaluated the therapy outcome–predictive power of the 11-gene model in a prostate cancer setting using a prognostic test based on an independent method of gene expression analysis, namely Q-RT-PCR (Figure 6F). In order to facilitate the evaluation of the 11-gene model, the average expression vector for different cancer types and the thresholds that separate the samples into good- and poor-prognostic groups are shown in Supplemental Table 6.

SPAI. Definition of the Pearson correlation coefficient as a phenotype association index is based on highly concordant behavior of the 11-gene signature between neural stem cells in the state of PNS neurospheres and prostate cancer metastasis (Figure 2, C and D; $r = 0.9897$; $P < 0.0001$). Values for a standard PNS neurosphere and for TRAMP metastasis were established as described in the signature discovery protocol and shown in Figure 2D. They were used consistently throughout the study as uniform reference standards for measurements of Pearson correlation coefficients

for clinical samples. 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 PAI. Samples with stem cell–resembling expression profiles of PAI (SPAI) are expected to have positive values of Pearson correlation coefficients.

Random co-occurrence test. We performed 10,000 permutations to test the likelihood that small 11-gene signatures derived from the large MTTS signature would display high discrimination power to assess the significance at the 0.1% level. We carried out 10,000 permutations of small 11-gene signatures derived from the large 1,345-gene MTTS signature and compared their sample-stratification power with that of the 11-gene MTTS/PNS signature. The classification-performance cutoff P values were established by application of a 2-tailed Student's t test to the 11-gene MTTS/PNS signature ($P = 0.0005$ for metastasis versus primary prostate cancer data set and $P = 0.026$ for recurrent versus nonrecurrent prostate cancer data set). Random concordant gene sets comprising approximately 200 transcripts were generated using a mouse transcriptome data set representing expression profiling data of approximately 12,000 transcripts across 38 normal tissues (55). Inter- and intraspecies probe set match between different array types was performed at 95% or greater identity level using the Affymetrix database (www.affymetrix.com). To assess discrimination of random 11-gene signatures derived from the 1,345-gene MTTS signature, a 2-tailed Student's t test was carried out for metastatic versus primary prostate cancer data set (32 samples) and recurrent versus nonrecurrent prostate cancer data set (21 samples). The signatures were ranked based on P values, and ranking metrics of each random 11-gene signature were compared with the 11-gene MTTS/PNS signature P values. We found that 10,000 permutations generated 7 random 11-gene signatures performing at the sample-classification level of the 11-gene MTTS/PNS signature.

Weighted survival predictor score algorithm. We implemented the weighted survival score analysis to reflect the incremental statistical power of the individual covariates as predictors of therapy outcome based on a multi-component prognostic model. The microarray-based or Q-RT-PCR–derived gene expression values were normalized and log-transformed on a base 10 scale. The log-transformed normalized expression values for each data set were analyzed in a multivariate Cox proportional hazard regression model, with overall survival or event-free survival as the dependent variable. To calculate the survival/prognosis predictor score for each patient, we multiplied the log-transformed normalized gene expression value measured for each gene by a coefficient derived from the multivariate Cox proportional hazard regression analysis. Final survival predictor score comprises a sum of scores for individual genes and reflects the relative contribution of each of the 11 genes in the multivariate analysis. The negative weighting values indicate that higher expression correlates with longer survival and favorable prognosis, whereas the positive score values indicate that higher expression correlates with poor outcome and shorter survival. Thus, the weighted survival predictor model is based on a cumulative score of the weighted expression values of 11 genes. For example, the following equation describes the relapse-free survival predictor score for prostate cancer patients (see Table 5): relapse-free survival score = $(-0.403 \times Gbx2) + (1.2494 \times KI67) + (-0.3105 \times cyclin\ B1) + (-0.1226 \times BUB1) + (0.0077 \times HEC) + (0.0369 \times KIAA1063) + (-1.7493 \times HCF1) + (-1.1853 \times RNF2) + (1.5242 \times ANK3) + (-0.5628 \times FGFR2) + (-0.4333 \times CES1)$.

BMI-1 siRNA experiments. The target siRNA SMART pools for *BMI-1* and control luciferase siRNAs were purchased from Dharmacon Research Inc. They were transfected into PC-3-32 human prostate carcinoma cells according to the manufacturer's protocols. Cell cultures were continuously monitored for growth and viability and assayed for mRNA expression levels of *BMI-1* and selected sets of genes (Table 2 and Figure 7) using RT-PCR and Q-RT-PCR methods.



Q-RT-PCR analysis. The real-time PCR method measures the accumulation of PCR products with a fluorescence detector system and allows for quantification of the amount of amplified PCR products in the log phase of the reaction. Total RNA was extracted using RNeasy Mini Kit (QIAGEN) according to the manufacturer's instructions. A measure of 1 µg (tumor samples), or 2 µg and 4 µg (independent preparations of reference cDNA samples), of total RNA was used then as a template for cDNA synthesis with SuperScript II (Invitrogen Corp.). Q-RT-PCR primer sequences were selected for each cDNA with the aid of Primer Express software (Applied Biosystems). PCR amplification was performed with the gene-specific primers listed in Supplemental Table 1.

Q-RT-PCR reactions and measurements were performed with SYBR Green and ROX (Applied Biosystems) as a passive reference, using the ABI 7900HT Sequence Detection System (Applied Biosystems). Conditions for the PCR were as follows: 1 cycle of 10 minutes at 95°C; and 40 cycles of 0.20 minutes at 94°C, 0.20 minutes at 60°C, and 0.30 minutes at 72°C. The results were normalized to the relative amount of expression of an endogenous control gene, *GAPDH*.

Expression of mRNA for 11 genes (Supplemental Table 1) and an endogenous control gene (*GAPDH*) was measured by real-time PCR on an ABI PRISM 7900HT Sequence Detection System (Applied Biosystems) in 20 specimens of primary prostate cancer obtained from patients with documented PSA recurrence within 5 years after RP and patients who remained disease-free for at least 5 years after RP (10 patients in each group). For each gene, at least 2 sets of primers were tested, and the set-up with highest amplification efficiency was selected for the assay used in this study. Specificity of the assay for mRNA measurements was confirmed by the absence of the expected PCR products when genomic DNA was used as a template. *GAPDH* (5'-CCCTCAACGACCACTTTGTCA-3' and 5'-TTCCTCTTGCTCTTGCTGG-3') was used as the endogenous RNA and cDNA quantity normalization control. For calibration and generation of standard curves, we used several reference cDNAs: cDNA prepared from primary in vitro cultures of normal human prostate epithelial cells (15, 16), cDNA derived from the PC-3M human prostate carcinoma cell line (15, 16), and cDNA prepared from normal human prostate (15, 16). Expression analysis of all genes was assessed in 2 independent experiments using reference cDNAs to control for variations among different Q-RT-PCR experiments. Before statistical analysis, the normalized gene expression values were log-transformed (on a base 10 scale) similarly to the transformation of the array-based gene expression data.

Survival analysis. The Kaplan-Meier survival analysis was carried out using GraphPad Prism version 4.00 software (GraphPad Software). The endpoint for survival analysis in prostate cancer was the biochemical recurrence defined by the serum PSA increase after therapy. Disease-free interval was defined as the time period between the date of RP and the date of PSA relapse (for the recurrence group) or the date of last follow-up (for the nonrecurrence group). Statistical significance of the difference between the survival curves for different groups of patients was assessed using χ^2 and log-rank tests. To evaluate the incremental statistical power of the individual covariates as predictors of therapy outcome and unfavorable prognosis, we performed both univariate and multivariate Cox proportional hazard survival analyses.

Acknowledgments

We thank C. Pettaway (M.D. Anderson Cancer Center, Houston, Texas, USA) for providing human prostate cancer cell lines and W. Sellers (Dana-Farber Cancer Institute, Boston, Massachusetts, USA) and W. Gerald (MSKCC, New York, New York, USA) for providing the Affymetrix CEL files of human prostate tumors and associated clinical information. We thank M. McClelland and J. Welsh for critical comments and technical and material assistance. We thank D. Mercola and A. Krones-Herzig for technical and logistical help with tissue acquisition and RNA extraction, and A. Sawyers and Y. Ivanova for excellent technical assistance. We thank W. Gerald (MSKCC, New York, New York, USA) and R. Bast (M.D. Anderson Cancer Center, Houston, Texas, USA) for providing the original microarray data files and associated clinical information. This work was supported in part by NIH/National Cancer Institute grant 5R01 CA89827 (to G.V. Glinsky). This work was greatly facilitated by the use of previously published and publicly accessible research data.

Received for publication September 21, 2004, and accepted in revised form March 1, 2005.

Address correspondence to: Gennadi V. Glinsky, Sidney Kimmel Cancer Center, 10835 Altman Row, San Diego, California 92121, USA. Phone: (858) 450-5990; Fax: (858) 623-2740; E-mail: gglinisky@skcc.org.

1. Lessard, J., and Sauvageau, G. 2003. *Bmi-1* determines the proliferative capacity of normal and leukaemic stem cells. *Nature*. **423**:255–260.
2. Park, I.-K., et al. 2003. *Bmi-1* is required for maintenance of adult self-renewing haematopoietic stem cells. *Nature*. **423**:302–305.
3. Molofsky, A.V., et al. 2003. *Bmi-1* dependence distinguishes neural stem cell self-renewal from progenitor proliferation. *Nature*. **425**:962–967.
4. Dick, J.E. 2003. Self-renewal writ in blood. *Nature*. **423**:231–233.
5. Pardoll, R., Clarke, M.F., and Morrison, S.J. 2003. Applying the principles of stem-cell biology to cancer. *Nat. Rev. Cancer*. **3**:895–902.
6. Al-Hajj, M., Wicha, M.S., Benito-Hernandez, A., Morrison, S.J., and Clarke, M.F. 2003. Prospective identification of tumorigenic breast cancer cells. *Proc. Natl. Acad. Sci. U. S. A.* **100**:3983–3988.
7. Smalley, M., and Ashworth, A. 2003. Stem cells and breast cancer: a field in transit. *Nat. Rev. Cancer*. **3**:832–844.
8. Lessard, J., Baban, S., and Sauvageau, G. 1998. Stage-specific expression of polycomb group genes in human bone marrow cells. *Blood*. **91**:1216–1224.
9. Haupt, Y., Bath, M.L., Harris, A.W., and Adams, J.M. 1993. *Bmi-1* transgene induces lymphomas and collaborates with Myc in tumorigenesis. *Oncogene*. **8**:3161–3164.
10. Alkema, M.J., Jacobs, H., van Lohuizen, M., and Berns, A. 1997. Perturbation of B and T cell development and predisposition to lymphomagenesis in Eμ-*Bmi-1* transgenic mice require the *Bmi-1* RING finger. *Oncogene*. **15**:899–910.
11. Vonlanthen, S., et al. 2001. The *Bmi-1* oncoprotein is differentially expressed in non-small-cell lung cancer and correlates with INK4A-ARF locus expression. *Br. J. Cancer*. **84**:1372–1376.
12. Dimri, G.P., et al. 2002. The *Bmi-1* oncogene induces telomerase activity and immortalizes human mammary epithelial cells. *Cancer Res.* **62**:4736–4745.
13. LaTulippe, E., et al. 2002. Comprehensive gene expression analysis of prostate cancer reveals distinct transcriptional programs associated with metastatic disease. *Cancer Res.* **62**:4499–4506.
14. Singh, D., et al. 2002. Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell*. **1**:203–209.
15. Glinsky, G.V., Glinskii, A.B., Stephenson, A.J., Hoffmann, R.M., and Gerald, W.L. 2004. Expression profiling predicts clinical outcome of prostate cancer. *J. Clin. Invest.* **113**:913–923. doi:10.1172/JCI200420032.
16. Glinsky, G.V., Krones-Herzig, A., Glinskii, A.B., and Gebauer, G. 2003. Microarray analysis of xenograft-derived cancer cell lines representing multiple experimental models of human prostate cancer. *Mol. Carcinog.* **37**:209–221.
17. Magee, J.A., et al. 2001. Expression profiling reveals hepsin overexpression in prostate cancer. *Cancer Res.* **61**:5692–5696.
18. Dhanasekaran, S.M., et al. 2001. Delineation of prognostic biomarkers in prostate cancer. *Nature*. **412**:822–826.
19. Welsh, J.B., et al. 2001. Analysis of gene expression identifies candidate markers and pharmacological targets in prostate cancer. *Cancer Res.* **61**:5974–5978.
20. Luo, J., et al. 2001. Human prostate cancer and benign prostatic hyperplasia: molecular dissection by gene expression profiling. *Cancer Res.* **61**:4683–4688.
21. Stamey, T.A., et al. 2001. Molecular genetic profiling of Gleason grade 4/5 prostate cancers compared to benign prostatic hyperplasia. *J. Urol.* **166**:2171–2177.
22. Luo, J., et al. 2002. Gene expression signature of benign prostatic hyperplasia revealed by cDNA microarray analysis. *Prostate*. **51**:189–200.



23. Rhodes, D.R., Barrette, T.R., Rubin, M.A., Ghosh, D., and Chinnaiyan, A.M. 2002. Meta-analysis of microarrays: interstudy validation of gene expression profiles reveals pathways dysregulation in prostate cancer. *Cancer Res.* **62**:4427–4433.
24. Varambally, S., et al. 2002. The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature.* **419**:624–629.
25. Henshall, S.M., et al. 2002. Survival analysis of genome-wide gene expression profiles of prostate cancers identifies new prognostic targets of disease relapse. *Cancer Res.* **63**:4196–4203.
26. Lapointe, J., et al. 2004. Gene expression profiling identifies clinically relevant subtypes of prostate cancer. *Proc. Natl. Acad. Sci. U. S. A.* **101**:811–816.
27. Stuart, R.O., et al. 2004. In silico dissection of cell-type-associated patterns of gene expression in prostate cancer. *Proc. Natl. Acad. Sci. U. S. A.* **101**:615–620.
28. Huang, E., et al. 2003. Gene expression phenotypic models that predict the activity of oncogenic pathways. *Nat. Genet.* **34**:226–230.
29. Lamb, J., et al. 2003. A mechanism of cyclin D1 action encoded in the patterns of gene expression in human cancer. *Cell.* **114**:323–334.
30. Chang, H.Y., et al. 2004. Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. *PLoS Biol.* **2**:1–9.
31. Raaphorst, F.M., et al. 2003. Poorly differentiated breast carcinoma is associated with increased expression of the human polycomb group EZH2 gene. *Neoplasia.* **5**:481–488.
32. Gingrich, J.R., et al. 1996. Metastatic prostate cancer in a transgenic mouse. *Cancer Res.* **56**:4096–4102.
33. van 't Veer, L.J., et al. 2002. Gene expression profiling predicts clinical outcome of breast cancer. *Nature.* **415**:530–536.
34. Bhattacharjee, A., et al. 2001. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc. Natl. Acad. Sci. U. S. A.* **98**:13790–13795.
35. van Kemenade, F.J., et al. 2001. Coexpression of BMI-1 and EZH2 polycomb-group proteins is associated with cycling cells and degree of malignancy in B-cell non-Hodgkin lymphoma. *Blood.* **97**:3896–3901.
36. Hemmati, H.D., et al. 2003. Cancerous stem cells can arise from pediatric brain tumors. *Proc. Natl. Acad. Sci. U. S. A.* **100**:15178–15183.
37. Leung, C., et al. 2004. BMI1 is essential for cerebellar development and is overexpressed in human medulloblastomas. *Nature.* **428**:337–341.
38. Ramaswamy, S., Ross, K.N., Lander, E.S., and Golub, T.R. 2003. A molecular signature of metastasis in primary solid tumors. *Nat. Genet.* **33**:49–54.
39. van de Vijver, M.J., et al. 2002. A gene expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* **347**:1999–2009.
40. Bernards, R., and Weinberg, R.A. 2002. A progression puzzle. *Nature.* **418**:823.
41. Wang, X., et al. 2003. Cell fusion is the principal source of bone-marrow-derived hepatocytes. *Nature.* **422**:897–901.
42. Vassilopoulos, G., Wang, P.-R., and Russell, D.W. 2003. Transplanted bone marrow regenerates liver by cell fusion. *Nature.* **422**:901–904.
43. Alvarez-Dolado, M., et al. 2003. Fusion of bone-marrow-derived cells with Purkinje neurons, cardiomyocytes and hepatocytes. *Nature.* **425**:968–973.
44. Weimann, J.M., Johansson, C.B., Trejo, A., and Blau, H.M. 2003. Stable reprogrammed heterokaryons form spontaneously in Purkinje neurons after bone marrow transplant. *Nat. Cell Biol.* **5**:959–966.
45. Aboudy, K.S., et al. 2000. Neural stem cells display extensive tropism for pathology in adult brain: evidence from intracranial gliomas. *Proc. Natl. Acad. Sci. U. S. A.* **97**:12846–12851.
46. Brown, A.B., et al. 2003. Intravascular delivery of neural stem cell lines to target intracranial and extracranial tumors of neural and non-neural origin. *Hum. Gene Ther.* **14**:1777–1785.
47. Willenbring, H., et al. 2004. Myelomonocytic cells are sufficient for therapeutic cell fusion in liver. *Nat. Med.* **10**:744–748.
48. Rhodes, D.R., et al. 2004. Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression. *Proc. Natl. Acad. Sci. U. S. A.* **101**:9309–9314.
49. Shaffer, A.I., et al. 2001. Signatures of the immune response. *Immunity.* **15**:375–385.
50. Hernando, E., et al. 2004. Rb inactivation promotes genomic instability by uncoupling cell cycle progression from mitotic control. *Nature.* **430**:797–802.
51. Johnson, V.L., Scott, M.I., Holt, S.V., Hussein, D., and Taylor, S.S. 2004. Bub1 is required for kinetochore localization of BubR1, Cenp-E, Cenp-F and Mad2, and chromosome congression. *J. Cell Sci.* **117**:1577–1589.
52. Martin-Bluesma, S., Stucke, V.M., and Nigg, E.A. 2002. Role of Hec1 in spindle checkpoint signaling and kinetochore recruitment of Mad1/Mad2. *Science.* **297**:2267–2270.
53. Baron, V., et al. 2003. Inhibition of Egr-1 expression reverses transformation of prostate cancer cells in vitro and in vivo. *Oncogene.* **22**:4194–4204.
54. Tavazoie, S., Hughes, J.D., Campbell, M.J., Cho, R.J., and Church, G.M. 1999. Systematic determination of genetic network architecture. *Nat. Genet.* **22**:281–285.
55. Su, A.I., et al. 2002. Large-scale analysis of the human and mouse transcriptomes. *Proc. Natl. Acad. Sci. U. S. A.* **99**:4465–4470.
56. Beer, D.G., et al. 2002. Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat. Med.* **8**:816–824.
57. Lu, K.H., et al. 2004. Selection of potential markers for epithelial ovarian cancer with gene expression arrays and recursive descent partition analysis. *Clin. Cancer Res.* **10**:3291–3300.
58. Lancaster, J.M., et al. 2004. Gene expression patterns that characterize advanced stage serous ovarian cancers. *J. Soc. Gynecol. Investig.* **11**:51–59.
59. Dyrskjot, L., et al. 2003. Identifying distinct classes of bladder carcinoma using microarrays. *Nat. Genet.* **33**:90–96.
60. Shipp, M.A., et al. 2002. Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nat. Med.* **8**:68–74.
61. Rosenwald, A., et al. 2002. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N. Engl. J. Med.* **346**:1937–1947.
62. Rosenwald, A., et al. 2003. The proliferation gene expression signature is a quantitative integrator of oncogenic events that predicts survival in mantle cell lymphoma. *Cancer Cell.* **3**:185–197.
63. Gordon, G.J., et al. 2003. Using gene expression ratios to predict outcome among patients with mesothelioma. *J. Natl. Cancer Inst.* **95**:598–605.
64. Pomeroy, S.L., et al. 2002. Prediction of central nervous system embryonal tumour outcome based on gene expression. *Nature.* **415**:436–442.
65. Nutt, C.L., et al. 2003. Gene expression-based classification of malignant gliomas correlates better with survival than histological classification. *Cancer Res.* **63**:1602–1607.
66. Bullinger, L., et al. 2004. Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. *N. Engl. J. Med.* **350**:1605–1616.