Original Article   **Open Access**

Malignant field expression signatures in biopsy samples at diagnosis predict the likelihood of lethal disease in patients with localized prostate cancer

Gennadi V. Glinsky

Institute of Engineering in Medicine, University of California, San Diego, CA 92093, USA.

Correspondence to: Dr. Gennadi V. Glinsky, Institute of Engineering in Medicine, University of California, San Diego, 9500 Gilman Dr. MC 0435, La Jolla, CA 92093, USA. E-mail: gglinskii@ucsd.edu

How to cite this article: Glinsky GV. Malignant field expression signatures in biopsy samples at diagnosis predict the likelihood of lethal disease in patients with localized prostate cancer. *J Cancer Metastasis Treat* 2017;3:177-89.

ABSTRACT

Aim: Overtreatment of early-stage low-risk prostate cancer patients represents a significant problem in disease management and has significant socio-economic implications. Changes in prostate cancer screening and treatment practices in the United States have been associated with the recent decline in overall incidence and concomitant significant increase of the annual incidence of metastatic prostate cancer has been documented. Therefore, development of genetic and molecular markers of clinically significant disease in patients diagnosed with low grade localized prostate cancer would have a major impact in disease management. **Methods:** Identification of gene expression signatures (GES) associated with lethal prostate cancer has been performed using microarray analyses of biopsy specimens obtained at the time of diagnosis from 281 patients with Gleason 6 (G6) and G7 tumors in a Swedish watchful waiting cohort with up to 30 years follow-up. The performance of GES has been validated in independent cohort of 568 prostate cancer patients of the Cancer Genome Anatomy Project Prostate Cancer database. **Results:** GES comprising 98 genes identified 89% and 100% of all death events 4 years after diagnosis in G7 and G6 patients, respectively. At 6 years follow-up, 83% and 100% of all deaths events were captured in G7 and G6 patients, respectively. Remarkably, the 98-gene signature appears to perform successfully in patients stratification with as little as 2% of cancer cells in a specimen, strongly indicating that it captures a malignant field effect in human prostates harboring cancer cells of different degrees of aggressiveness. In G6 and G7 tumors from prostate cancer patients of age 65 or younger, GES identified 86% of all death events during the entire follow-up period. In G6 and G7 tumors from prostate cancer patients of age 70 or younger, GES identified 90% of all death events 6 years after diagnosis. **Conclusion:** Classification performance of the reported in this study 98-genes GES of lethal prostate

Quick Response Code:

cancer appeared suitable to meet design and feasibility requirements of a prospective 4 to 6 years clinical trial, which is essential for regulatory approval of diagnostic and prognostic tests in clinical setting. Prospectively validated GES of lethal PC in biopsy specimens of G6 and G7 tumors will help physicians to identify, at the time of diagnosis, patients who should be considered for exclusion from active surveillance programs and who would most likely benefit from immediate curative interventions.

INTRODUCTION

In the United States, widespread implementation of the prostate-specific antigen (PSA) screening programs enabled diagnosis of more than 200,000 cases of prostate cancer each year.[1] Clinically localized prostate cancer represents the vast majority of new cases.[2] Therefore, one of the most significant benefits of the widespread use of PSA screening is that the prevalence of the late stage, advanced and high grade prostate cancer at diagnosis has declined dramatically and the vast majority of newly diagnosed prostate cancers are early stage and low grade tumors.

The natural history of early stage clinically localized prostate cancer is considered favorable[3] and other types of cancer such as lung cancer are considered hundreds times as deadly. Despite this seemingly "indolent" nature, prostate cancer is the second leading cause of cancer-related deaths and accounts for 3.5% of all male deaths.[4] Development of clear, consensus guidelines for physicians' decision-making process in clinical management of early stage localized prostate cancer is one of the most significant public healthcare problems. Inevitable and fast approaching demographic changes in the Western world underscore the critical economic and logistical needs for a rational, evidence-based approach to the clinical management of the early stage localized prostate cancer. A path to solutions to this problem is complicated by a multitude of competing positions attempting to emphasize the perceived shortcomings and benefits of different approaches and need to balance multiple variables such as public health care costs, individual patients' benefits, interests, socio-economic status, ethical and professional responsibilities of the medical personnel, and humanitarian considerations.

Conclusive statistical evidence of the life-saving therapeutic benefits of radical prostatectomy versus watchful waiting in early prostate cancer have been documented in a randomized multicenter clinical trial: radical prostatectomy reduces disease-specific mortality, overall mortality, and the risks of metastasis and local progression.[5-7] Immediate curative interventions are the predominant therapy choice and 168,000 prostatectomies are performed each year to treat prostate cancer.[8] It seems reasonable to conclude, that early detection of prostate cancer facilitated by PSA screening and aggressive use of

radical prostatectomy for treatment of early prostate cancer have contributed to a significant extent to the reported 98-100% 5-year survival rates since 1998 in the United States (SEER 13 areas statistics).

However, there is a lack of consensus regarding the benefits of a population-scale PSA screening and a controversy about the potential for overdiagnosis and overtreatment of clinically insignificant disease that would not likely to become life-threatening in a man's lifetime.[9] Further socio-economic arguments in support of significant overdiagnosis and overtreatment have been presented in studies indicating that prevention of one prostate cancer death would require active treatment of 48 men for 9 years or 12 men for 14 years.[10,11] Outcome studies from contemporary population-based cohorts reported cumulative 10-year prostate cancer-specific mortality in patients with low-risk disease 2.4% and 0.7% in the surveillance group and curative intent groups, respectively,[12] which indicates that the surveillance may be a suitable treatment option for majority of patients with low-risk prostate cancer. Clinical evidence that active surveillance may be a safe, perhaps preferred option for older men diagnosed with a very low-grade or small-volume form of prostate cancer were published recently by Tosoian *et al*.[13] Therefore, active surveillance with curative intent for low-risk prostate cancer is under active consideration as a potentially safe alternative to immediate curative intervention with the expectations that it may reduce overtreatment and therapy-associated adverse events. It certainly would reduce the escalating economic burden of cost of prostate cancer treatment. The major limitation of these studies is a short follow-up time [for example, in the John Hopkins study,[13] the total cohort has a median follow-up of 2.7 years (range 0.01 to 15)] which requires the use of biochemical recurrence or other "proxy" end-points for disease-specific mortality. This limitation is particularly relevant for early prostate cancer because the overall survival benefits of radical prostatectomy versus watchful waiting are not statistically apparent until 10 years follow-up[5-7] due to the fact that a majority of death events in the watchful waiting cohorts of early prostate cancer occurs at or after 10 years follow-up (this study).[5-7] Furthermore, significantly longer follow-up data are required because most patients currently diagnosed with localized prostate cancer are aged 60-70 years and have a life expectancy of more than 15 years.[12]

Most importantly, there are no genetic or molecular methods prospectively defining low-risk or indolent prostate cancer at diagnosis with sufficient specificity and selectivity to ensure the safety of patients and allow physicians to make informed, ethical, evidence-based disease management decision of not treating prostate cancer. Given the natural history of early prostate cancer and long-term survival data from watchful waiting cohorts, conclusive prospective validation of laboratory methods defining low-risk indolent disease in Gleason 6 and 7 patients would require at least 10 years. Based on the analysis of the long-term survival data of prostate cancer patients from watchful waiting cohorts with up to 30 years follow-up, we reasoned that more feasible and clinically-relevant approach would be an attempt to identify genetic markers of lethal prostate cancer in patients with Gleason 6 and 7 tumors which would capture a vast majority of all cancer-related death events 4-6 years after diagnosis. Here we report identification of gene expression signatures (GES) of lethal prostate cancer in biopsy specimens obtained at the time of diagnosis from patients with Gleason 6 and 7 tumors in a Swedish watchful waiting cohort with up to 30 years follow-up. In retrospective analysis, best-performing GES of lethal prostate cancer identify 89% and 100% of all death events 4 years after diagnosis in Gleason 7 and Gleason 6 patients, respectively. GES appear to perform successfully in patients' stratification with as little as 2% of cancer cells in a specimen. In Gleason 6 and 7 prostate cancer patients of age 65 or younger, GES identifies 86% of all death events during the follow-up. In Gleason 6 and 7 prostate cancer patients of age 70 or younger, GES identifies 90% of all death events 6 years after diagnosis. Reported in this study GES of lethal prostate cancer in biopsy specimens of Gleason 6 and 7 tumors should help practicing physicians to identify at the time of diagnosis prostate cancer patients who should be considered for exclusion from the active surveillance programs and who would most likely benefit from immediate curative interventions.

METHODS

Patients

This study is based on prostate cancer patients from the population-based Swedish Watchful Waiting cohort of men with localized prostate cancer.[5-7,14] Distinguishing feature of this cohort is that it represents patients diagnosed with symptomatic early prostate cancer at the time when no PSA screening programs were in place: these men had symptoms of benign prostatic hyperplasia (lower urinary tract symptoms) and were subsequently diagnosed with prostate

cancer. All men in this study were determined at the time of diagnosis to have clinical stage T1 and T2, Mx, and N0, according to the 2002 American Joint Commission Committee TNM staging system.[5-7,14] The prospective follow-up time in this cohort is now up to 30 years and the study cohort was followed for cancer-specific and all-cause mortality until March 1, 2006.[11] Deaths were classified as cancer-specific when prostate cancer was the primary cause of death as determined through a complete review of medical records by a study end-point committee.[5-7,14] Importantly, that in addition to the histopathological examination at the time of diagnosis, slides and corresponding paraffin-embedded formalin-fixed blocks were subsequently retrieved and re-reviewed to confirm cancer status and to assess Gleason scores using review, examination, and grading procedures blinded with regard to disease outcome.[14]

Gene expression analysis, evaluation, and selection of GES

GES were developed based on a publicly available microarray analysis of a Swedish Watchful Waiting cohort with up to 30 years of clinical follow up using a novel method for gene expression profiling (cDNA-mediated annealing, selection, ligation, and extension method) which enabled the use of formalin-fixed paraffin-embedded transurethral resection of prostate (TURP) samples taken at the time of the initial diagnosis. Details of the experimental procedure can be found in a recent publication[14] and in Gene Expression Omnibus (GEO: http://www.ncbi.nlm.nih.gov/geo/) with platform accession number: GPL5474. Full data set and associated clinical information is available at GEO with accession number: GSE16560.

Feature selection was performed without assessment of differential gene expression between deceased and surviving patients. All 6,144 genes were evaluated for association with clinical and pathological variables (except survival status) using correlation analysis. Different thresholds on the *P*-values (0.05; 0.01; 0.001) were used for selection of gene sets with common patterns of association and concordance analysis was performed using expression profiling data of snpRNA-driven cell line-based models of prostate cancer predisposition[15,16] to identify concordant and discordant GES in cell lines and clinical samples.[17-20] GES were built based on selection of co-regulated transcripts in various experimental conditions and clinically-relevant models, including prostate cancer predisposition and longevity models.[16-20] Underlying concept at this stage of the analysis was to identify GES with concordant expression profiles across multiple data sets.[17-20] Cox regression analysis was

carried out to identify statistically significant candidate GES associated with patients' survival status. Cut-off threshold of *P*-values was set based on the *P*-value of the best-performing clinico-pathological parameter (Gleason score) in univariate Cox regression analysis (*P* = 0.0113). Genes from statistically significant GES were split, combined, and permutated using random iteration process to find novel statistically significant combinations based on univariate Cox regression analysis. GES scores were derived directly from measurements of expression values of each gene by calculating a single numerical value for each patient. GES scores represent the difference between sums of expression values of genes with common co-regulation profiles which is defined by up-regulation and/or positive correlation values versus down-regulation and/or negative correlation values. GES with *P* values < 0.01 were selected for further evaluation using multivariate Cox regression analysis of classification models which include GES and clinico-pathological co-variants (age and Gleason score). Cut-off threshold of *P*-values for candidate GES selection was set based on the *P*-value of the best-performing clinico-pathological model (age and Gleason score) in multivariate Cox regression analysis (*P* = 0.0052). Candidate GES that outperformed clinico-pathological models in multivariate Cox regression analysis were selected for further consideration using a split-sample validation procedure for classification threshold selection and GES classification performance evaluation as previously described.[17-20]

Gene expression-based classification models were designed and evaluated through a split-sample validation procedure which enables the unbiased estimation of the performance of a classifier since the evaluation is performed on an independent data set.[21] Specifically, the entire data set of 281 patients was split into training and test sets (141 and 140 patients, respectively), with approximately equal proportion

of men with lethal and indolent prostate cancer and statistically undistinguishable clinical and pathological variables, e.g. age and time of diagnosis, follow up time, Gleason scores, percent of cancer cells in specimens [Table 1]. The training set of 141 samples was utilized to identify and select the best classifier, whose performance was evaluated on the test set of 140 samples without any further adjustments to the threshold selection and classification protocols using Kaplan-Meyer survival analysis essentially as previously described.[17-20] Best-performing GES classifiers were further evaluated in various clinically-relevant patients' sub-groups, including only Gleason 6 patients (*n* = 83), only Gleason 7 patients (*n* = 117), Gleason 6 and 7 patients (*n* = 200), with further sub-division of patients in additional validation screens based on age at diagnosis (age 65 and younger; age 70 and younger) and percent of cancer cells in the samples (2%; 5% or less; 10% or less; 20% or less; 40% or less; and 50% or more). In all these secondary validation screens no further adjustments to the threshold selection and classification protocols were made. Ninety-eight genes classifier that remains statistically significant in all these validation screens is reported in this paper.

Statistical significance of the Pearson correlation coefficients for individual test samples, clinical variables, and the appropriate reference standard were determined using GraphPad Prism version 4.00 software. We calculated the significance of the differences in the numbers of death events and surviving patients between the groups using two-sided Fisher's exact test and the significance of the overlap between the lists of differentially-regulated genes using the hypergeometric distribution test.[22]

Validation analyses of GES were performed using the most recent release of web-based tools, the UCSC Xena (http://xena.ucsc.edu/) to explore and visualize

Table 1: Clinical characteristics of prostate cancer patients in the training and test sets

Characteristic	Training set (*n* = 141)	Test set (*n* = 140)
Years of diagnosis, range (years)	1977-1998	1977-1998
Years of diagnosis, mean ± SD (years)	1991 ± 4.1	1991 ± 4.0
Age at diagnosis, range (years)	51-91	55-91
Age at diagnosis, mean ± SD (years)	74.5 ± 7.5	73.5 ± 7.0
Follow-up time, range (months)	6-274	7-259
Follow-up time, mean ± SD (months)	102.3 ± 57.2	101.9 ± 55.7
Percent of cancer in samples, range (%)	2-90%	2-90%
Percent of cancer in samples, mean ± SD (%)	22.9 ± 22.7	24.0 ± 25.5
Gleason scores, *n* (%)		
Gleason 6	42 (29.8)	41 (29.3)
Gleason 7	62 (44)	55 (39.3)
Gleason 8-10	37 (26.2)	44 (31.4)
Clinical outcomes, *n* (%)		
Deceased	105 (74.5)	101 (72.1)
Alive	36 (25.5)	39 (27.9)

SD: standard deviation

Table 2: Receiver operating characteristic area under the curve analysis of training and test data sets

Data sets and survival time	10 years	7 years	6 years	5 years	4 years
Training set (n = 141)	0.85	0.854	0.814	0.788	0.794
Test set (n = 140)	0.826	0.801	0.786	0.758	0.759

Table 3: Percent of all death events at different follow-up time in lethal prostate cancer groups of training and test data sets

Data sets and survival time	10 years	7 years	6 years	5 years	4 years
Training set (n = 141)	75%	83%	82%	84%	84%
Test set (n = 140)	83%	88%	87%	84%	84%

the comprehensive functional cancer genomics datasets of thousands annotated clinical samples of the Cancer Genome Anatomy Project (TCGA) (https://xenabrowser.net/datapages/). The classification performance of the 98-genes GES was further validated using TCGA Prostate Cancer cohort of 568 clinical samples with known therapy outcomes after the initial treatment. Importantly, in contrast to biopsy samples analyzed in the population-based Swedish Watchful Waiting cohort, tumors tissues of the cotemporary TCGA Prostate Cancer cohort comprise the prostatectomy samples which were analyzed using the state of the art Illumina Next Generation Sequencing technology.

RESULTS

Clinical characteristics of the training and test sets are provided in Table 1, and further details for the entire Swedish Watchful Waiting cohort are available in a recent publication[14] and in Gene Expression Omnibus (GEO: http://www.ncbi.nlm.nih.gov/geo/) with accession number GSE16560. All of the 281 patients in the Swedish cohort had clinical symptoms and were diagnosed from TURP or adenoma enucleation samples and thus were staged depending on the proportion of the tissue that was cancerous either T1a or T1b.[14] Analysis of survival data in the entire cohort of 281 patients indicates that prostate cancer patients with different Gleason scores have markedly distinct timelines of death events during the extended up to 30 years follow-up [Figure 1]. Most striking indicator is that only 6% of untreated Gleason 6 prostate cancer

patients died at 5 years; 14% died between 5 to 10 years; and a majority of deaths (~ 35%) occurs 10-23 years after diagnosis. This analysis suggests that a majority of all death events (> 60%) in untreated Gleason 6 prostate cancer patients is occurring more than 10 years after diagnosis and during the sufficiently long follow-up period more than 50% of these patients will die [Figure 1]. Long-term survival timelines for untreated Gleason 7 prostate cancer patients with symptomatic prostate cancer appear even more alarming: 27% died at 5 years follow-up; 22% of deaths occurred between 5 to 10 years; and > 70% died during the entire follow-up period [Figure 1]. When compared with active surveillance patients from the PSA screening era these mortality figures seem very high, particularly because the survival references were made to prostate cancer-specific mortality. At least in part, it might be attributed that in this cohort all Gleason 7 prostate cancer patients were clinically symptomatic in contrast to predominantly asymptomatic Gleason 7 prostate cancer patients diagnosed during the PSA screening era. These apparent differences indicate what would likely to happen to prostate cancer specific mortality if the population scale PSA screening practices will be changed or abandoned.

Collectively, the analysis of timelines of death events in a watchful waiting cohort indicates that a majority of patients with symptomatic Gleason 6 and 7 prostate cancers will eventually develop clinically significant disease during sufficiently long follow-up period which further underscore the critical need to reliably define

Table 4: Classification performance of the 98-genes GES in the TCGA cohort of 550 prostate cancer patients with known therapy outcomes after the initial treatment

Categories	Therapy outcomes after the initial treatment (number of patients with adverse events)		
Patients' sub-group/adverse events	Relapse	Biochemical recurrence	New tumors
Poor prognosis (n = 275)	33	44	60
Good prognosis (n = 275)	10	18	20
Patients' sub-group/adverse events	Therapy outcomes after the initial treatment (percent of patients with adverse events)		
Poor prognosis (top 50% scores)	12.00	16.00	21.82
Good prognosis (bottom 50% scores)	3.64	6.55	7.27
P value*	0.0004	0.0006	< 0.0001

*P values were estimated using 2-tailed Fisher's exact test. At the date of the analyses, the median follow-up time in the prostate cancer TCGA cohort was 2.1 years. GES: gene expression signatures; TCGA: the Cancer Genome Anatomy Project
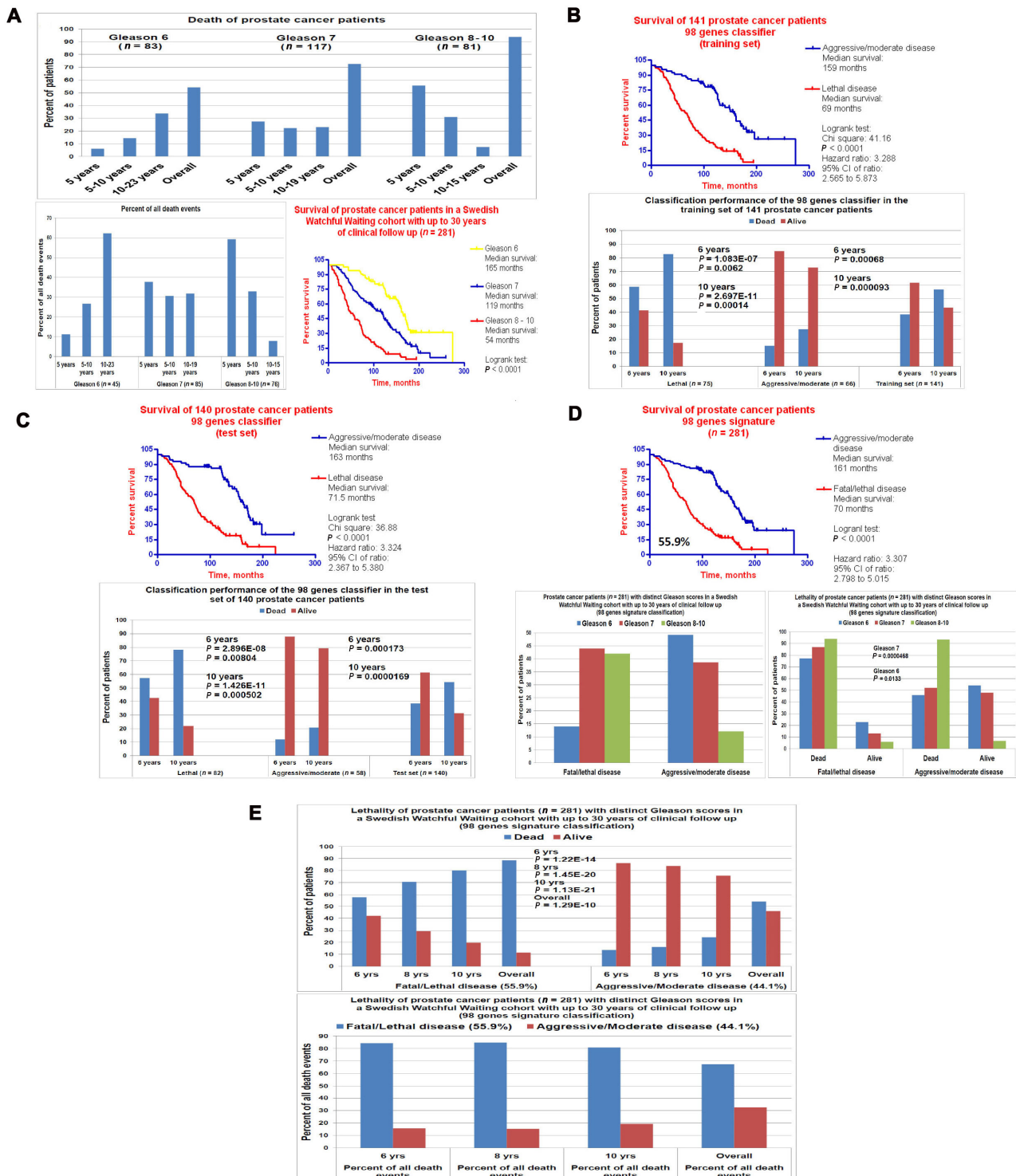
Figure 1: Natural history of prostate cancer progression in patients' population from a Swedish watchful waiting cohort with up to 30 years follow-up (A) and classification performance of the 98 genes signature of lethal disease in prostate cancer patients (B-E). (A) Cancer-specific survival data in the entire watchful waiting cohort are presented to illustrate markedly distinct survival timelines of non-treated prostate cancer patients diagnosed with different Gleason scores prostate cancer. Kaplan-Meier survival analysis of the classification performance of the 98 genes GES in the training set (B), test set (C), and pooled cohort of 281 patients (D, E). Classification threshold 98 genes GES score of 270.43 units was chosen using the training set of 141 prostate cancer patients and consistently applied in all subsequent validation screens using the Kaplan-Meier survival analysis to stratify the patients into lethal disease sub-groups (score ≥ 270.43) and moderate/aggressive disease sub-group (score < 270.43). Percent value indicates the proportion of patients in the lethal disease sub-group. *P* values indicate the significance of the differences in the numbers of death events and surviving patients between the groups which was determined using two-sided Fisher's exact test. GES: gene expression signatures
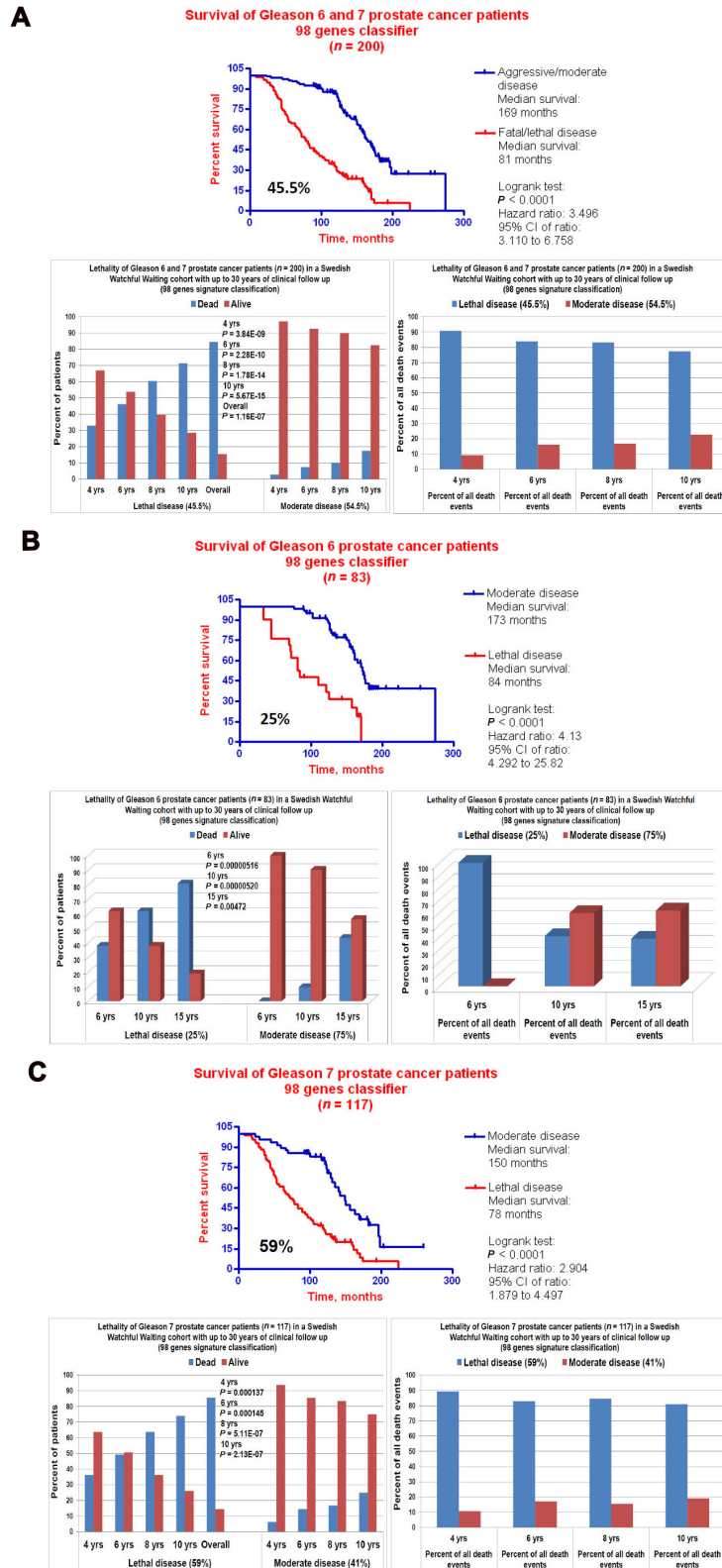
Figure 2: GES-based identification of lethal disease in Gleason 6 and 7 prostate cancer patients. Kaplan-Meier survival analysis of the classification performance of the 98 genes GES in 200 Gleason 6 and 7 prostate cancer patients (A), 83 Gleason 6 patients (B), and 117 Gleason 7 patients (C). Classification threshold 98 genes GES score of 270.43 units was chosen using the training set of 141 prostate cancer patients and consistently applied in all subsequent validation screens using the Kaplan-Meier survival analysis to stratify the patients into lethal disease sub-groups (score $\geq$ 270.43) and moderate/aggressive disease sub-group (score < 270.43). Percent values indicate the proportion of patients in the lethal disease sub-group. *P* values indicate the significance of the differences in the numbers of death events and surviving patients between the groups which was determined using two-sided Fisher's exact test. GES: gene expression signatures

lethal prostate cancer at diagnosis. We applied the univariate Cox regression analysis to the entire cohort of 281 patients to identify several GES with the P value < 0.01 which appear to perform better than the best clinico-pathological co-variate, Gleason score (P = 0.0113; Supplemental Table 1). Most of these GES outperformed the clinico-pathological classification model in multivariate Cox regression analysis as well [Supplemental Table 2].

Separating the cohort of 281 patients into training and test cohorts and using the Kaplan-Meier survival analysis, we identified 98 genes GES that manifest the highly significant classification performance in the training set, retained highly consistent classification performance in the test set, and remained a highly significant classifier in the pooled cohort [Figure 1]. It is important to note that in all secondary validation screens following the training set analysis no further adjustments to the threshold selection and classification protocols were made.

Notably, prostate cancer patients with identical Gleason scores (e.g. Gleason 6 patients and Gleason 7 patients) which were segregated into lethal and moderate disease sub-groups based on 98 genes GES classification had highly significant differences in the survival rates [Figure 1]. These data suggest that 98 genes GES may be useful in identifying lethal disease in patients diagnosed with low grade localized prostate cancer [Supplemental Table 3]. To test this hypothesis, we performed Kaplan-Meier survival analysis based on 98 genes GES classification in the cohort of 200 patients with Gleason 6 and 7 prostate cancer [Figure 2]. We found that 98 genes GES is a highly significant classifier of Gleason 6 and 7 prostate cancer patients into sub-groups with lethal and moderate disease [Figure 2]. Ninety-eight genes GES of lethal prostate cancer performs as a highly significant after segregation of patients into separate Gleason 6 and Gleason 7 sub-groups: 89% and 100% of all death events were identified 4 years after diagnosis in Gleason 7 and Gleason 6 patients, respectively; at 6 years follow-up, 83% and 100% of all deaths events were captured in Gleason 7 and 6 patients, respectively [Figure 2].

Age at diagnosis is considered among very important clinical determinants guiding the decision making process in clinical management of prostate cancer. This is particularly important for relatively younger patients because patients diagnosed with prostate cancer at age < 65 years are more likely to benefit from the immediate curative therapies.[7] We therefore attempted to determine whether 98 genes GES will

identify lethal disease in prostate cancer patients of differing ages. Remarkably, Kaplan-Meier survival analysis has determined that 98 genes GES performed very efficiently in stratification of prostate
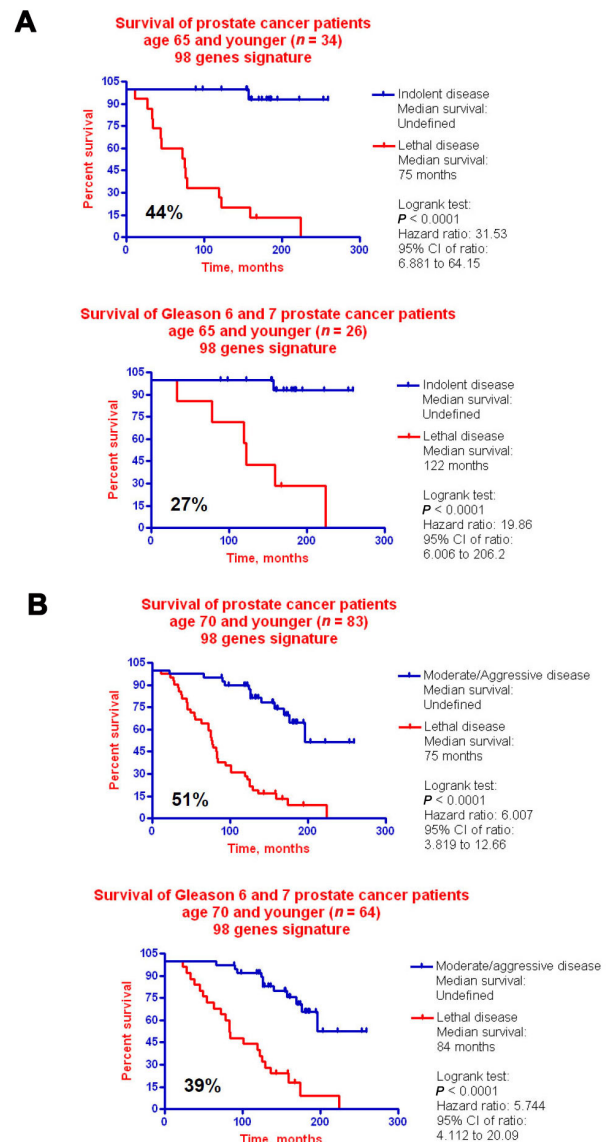

Figure 3: GES-based identification of lethal disease in prostate cancer patients with different age at diagnosis. Kaplan-Meier survival analysis of the classification performance of the 98 genes GES in 34 prostate cancer patients of age 65 or younger (A), 64 prostate cancer patients of age 70 or younger (B). Bottom figures in both A and B panels show the results of Kaplan-Meier survival analysis for Gleason 6 and 7 patients only of corresponding age groups. Classification threshold 98 genes GES score of 270.43 units was chosen using the training set of 141 prostate cancer patients and consistently applied in all subsequent validation screens using the Kaplan-Meier survival analysis to stratify the patients into lethal disease sub-groups (score $\geq$ 270.43) and moderate/aggressive disease sub-group (score < 270.43). Percent values indicate the proportion of patients in the lethal disease sub-group. P values indicate the significance of the differences in the numbers of death events and surviving patients between the groups which was determined using two-sided Fisher's exact test. GES: gene expression signatures

cancer patients of 65 years or younger [Figure 3]: in Gleason 6 and 7 prostate cancer patients of age 65 or younger, GES identifies 86% of all death events during the follow-up. In Gleason 6 and 7 prostate cancer patients of age 70 or younger, GES identifies 90% of all death events 6 years after diagnosis [Figure 3].

Proportion of cancer cells in biopsy samples is highly variable and these variations may have significant impact on performance of gene expression-based classifiers. In biopsy samples from the population-based Swedish Watchful Waiting cohort the reported

percent of cancer cells in a sample varied dramatically from 2% to 90%. We therefore set out to determine whether the number of cancer cells in biopsy samples would have an impact on classification performance of the 98 genes GES of lethal prostate cancer. We applied the 98 genes GES classifier to prostate cancer patients which were segregated into distinct sub-groups based on the percent of cancer cells in a biopsy sample. Kaplan-Meier survival analysis demonstrates that 98 genes GES performs successfully in patients' stratification regardless of the number of cancer cells in biopsy samples [Figures 4 and 5]. Remarkably
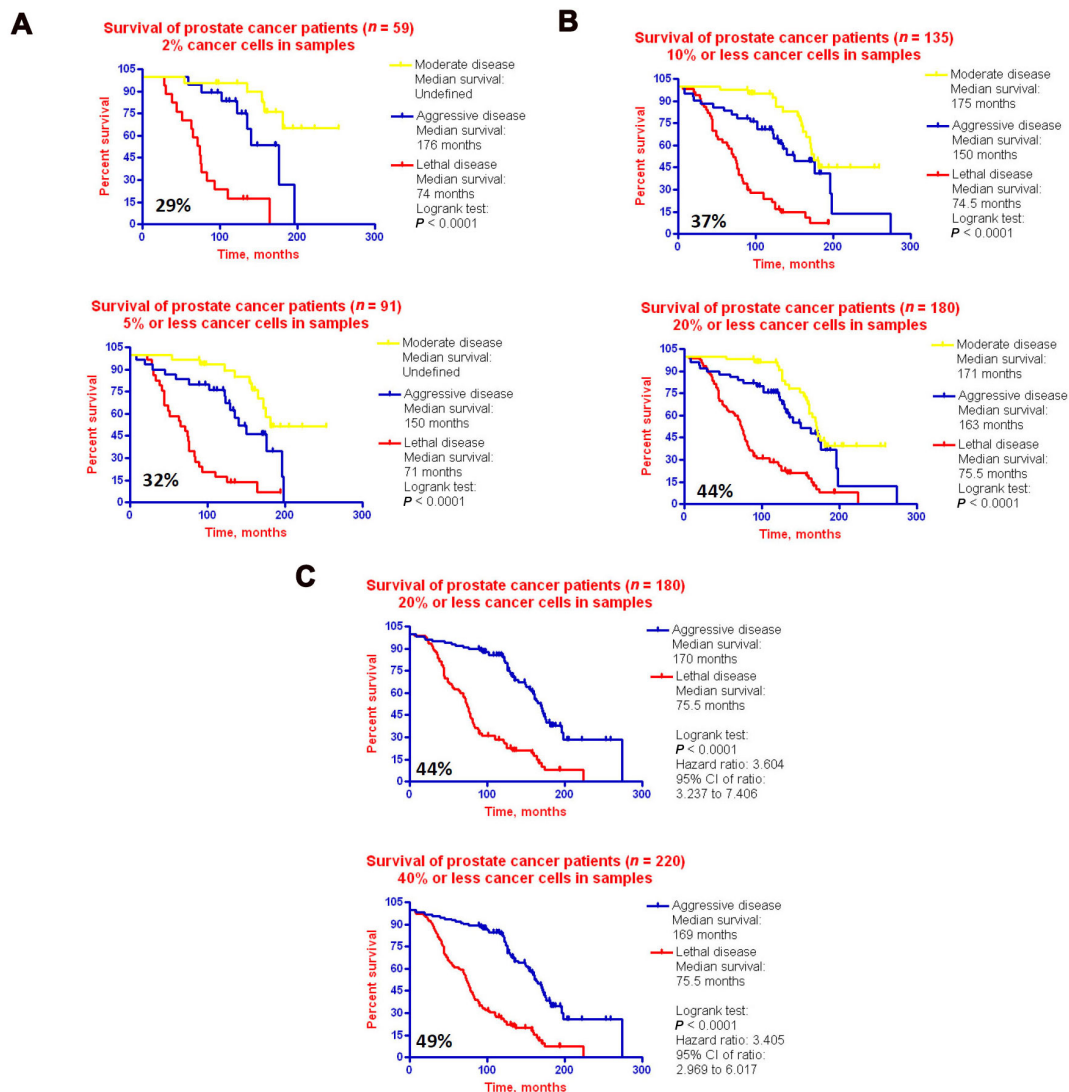

Figure 4: GES-based identification of lethal disease in prostate cancer patients with distinct numbers of cancer cells in biopsy samples. Kaplan-Meier survival analysis of the classification performance of the 98 genes GES in 59 prostate cancer patients having 2% cancer cells in biopsy samples (A, top), 91 patients having 5% or less cancer cells in biopsy samples (A, bottom), 135 patients having 10% or less cancer cells in biopsy samples (B, top), 180 patients having 20% or less cancer cells in biopsy samples (B, bottom; and C, top), 220 patients having 40% or less cancer cells in biopsy samples (C, bottom). Classification threshold 98 genes GES score of 270.43 units was chosen using the training set of 141 prostate cancer patients and consistently applied in all subsequent validation screens using the Kaplan-Meier survival analysis to stratify the patients into lethal disease sub-groups (score ≥ 270.43) and moderate/aggressive disease sub-group (score < 270.43). Percent values indicate the proportion of patients in the lethal disease sub-group. *P* values indicate the significance of the differences in the numbers of death events and surviving patients between the groups which was determined using two-sided Fisher's exact test. GES: gene expression signatures

98 genes GES appear to identify lethal disease in Gleason 6 and 7 prostate cancer patients with as little as 2% of cancer cells in a biopsy specimen [Figure 5]. The conclusions reached based on the Kaplan-Meier survival analyses were confirmed using the receiver operating characteristic area under the curve analysis of the patients' classification based on the 98-genes signature score in training (n = 141) and test (n = 140) groups (A) and different clinically-relevant sub-groups (B-D) of patients [Figure 6; Tables 2 and 3]. Collectively, the results of the present analyses strongly indicate that the 98-genes GES captures a malignant field effect in the human prostates harboring cancer cells

with markedly different clinical aggressiveness.

The most recent release of web-based tools, the UCSC Xena (http://xena.ucsc.edu/), provides powerful resources to explore, analyze, and visualize the comprehensive functional cancer genomics datasets of thousands annotated clinical samples of TCGA (https://xenabrowser.net/datapages/). The classification performance of the 98-genes GES was further validated using TCGA Prostate Cancer cohort of 550 clinical samples with known therapy outcomes after the initial treatment [Table 4]. Notably, the 98-gene GES successfully stratified prostate cancer
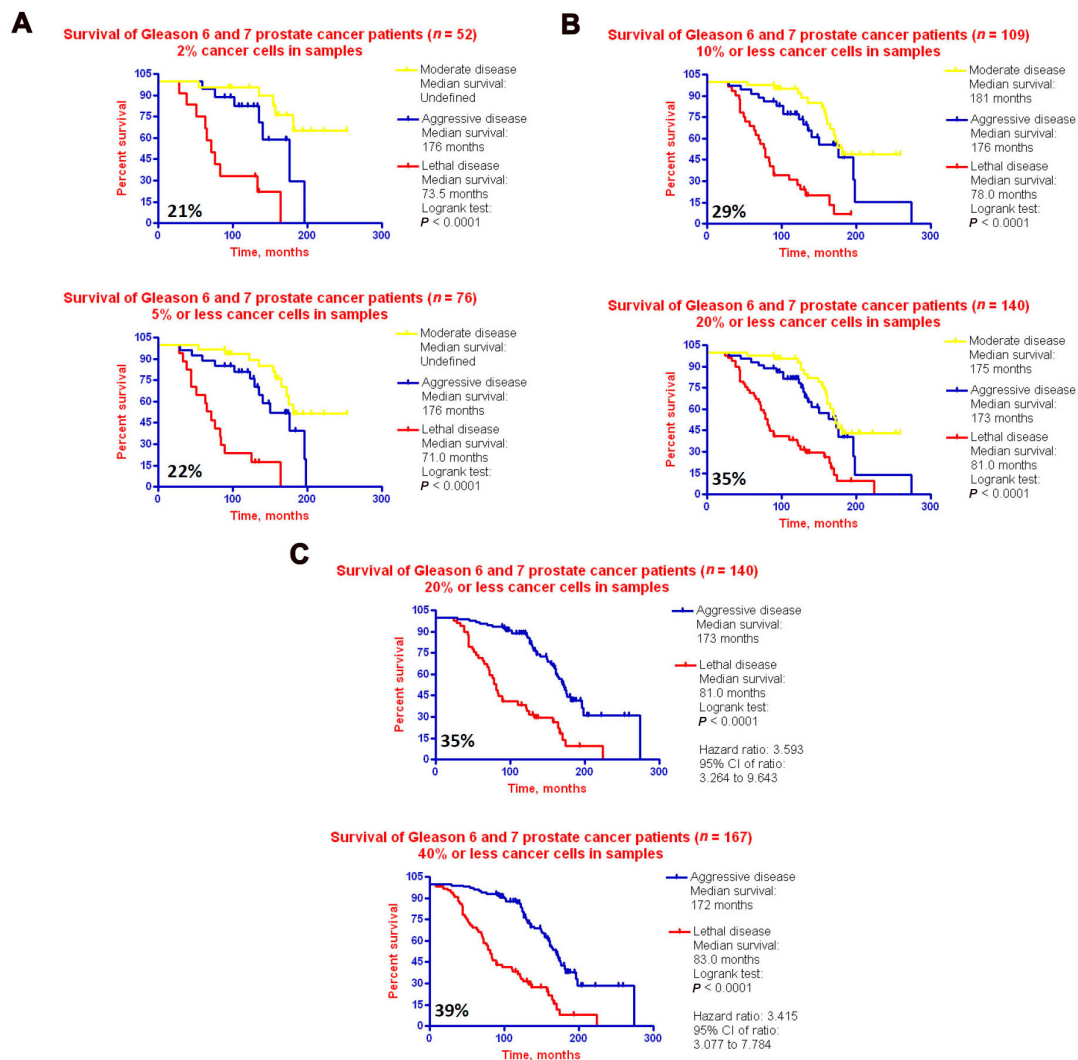

Figure 5: GES-based identification of lethal disease in Gleason 6 and 7 prostate cancer patients with distinct numbers of cancer cells in biopsy samples. Kaplan-Meier survival analysis of the classification performance of the 98 genes GES in 52 prostate cancer patients having 2% cancer cells in biopsy samples (A, top), 76 patients having 5% or less cancer cells in biopsy samples (A, bottom), 109 patients having 10% or less cancer cells in biopsy samples (B, top), 140 patients having 20% or less cancer cells in biopsy samples (B, bottom; and C, top), 167 patients having 40% or less cancer cells in biopsy samples (C, bottom). Classification threshold 98 genes GES score of 270.43 units was chosen using the training set of 141 prostate cancer patients and consistently applied in all subsequent validation screens using the Kaplan-Meier survival analysis to stratify the patients into lethal disease sub-groups (score ≥ 270.43) and moderate/aggressive disease sub-group (score < 270.43). Percent values indicate the proportion of patients in the lethal disease sub-group. *P* values indicate the significance of the differences in the numbers of death events and surviving patients between the groups which was determined using two-sided Fisher's exact test. GES: gene expression signatures

patients into subgroups with markedly distinct therapy outcomes after the initial treatment defined by different indicators of clinical progression such as biochemical recurrence, disease relapse, and appearance of recurrent tumors [Table 4].

DISCUSSION

Decision making process in clinical management of low-risk localized prostate cancer is likely to affect life and death of thousands of patients. The problem is confounded by the fact that statistically significant survival benefits of curative therapy are evident only 10 years after diagnosis of the early-stage prostate cancer. Therefore, any genetic or molecular tests designed to aid physicians and patients in this process would require the regulatory approval following the successful prospective clinical trial. Classification performance of the reported in this study 98 genes GES of lethal prostate cancer appears highly suitable to meet design and feasibility requirements of the prospective 4 to 6 years clinical trial. One of the most remarkable features of the 98-gene signature is that it appears to perform successfully in patients' stratification with as little as 2% of cancer cells in a specimen, indicating that this GES captures a

malignant field effect in human prostates harboring tumors of different degrees of aggressiveness. It will be of interest to investigate the molecular and genetic mechanisms of this phenomenon. Prospectively validated GES of lethal prostate cancer in biopsy specimens of Gleason 6 and 7 tumors will help practicing physicians to identify at the time of diagnosis individual patients who should be considered for exclusion from the active surveillance programs and who would most likely benefit from the immediate curative interventions.

One of the distinguishing features of this unique 281 patients' cohort that will never be replicated for ethical and humanitarian reasons, is that prostate cancer patients were never treated and just subjected to the long-term follow-up observations. In this context, the outcome data on the prostate cancer-specific death of these patients reveal what would happen to prostate cancer patients who will not be treated (i.e. subjected to "watchful waiting"). Importantly, it demonstrates that a majority of prostate cancer patients diagnosed with Gleason 6 and 7 tumors will die from prostate cancer when left untreated. A distinguishing feature of the patients' cohort analyzed in this study is that it represents patients diagnosed with symptomatic early
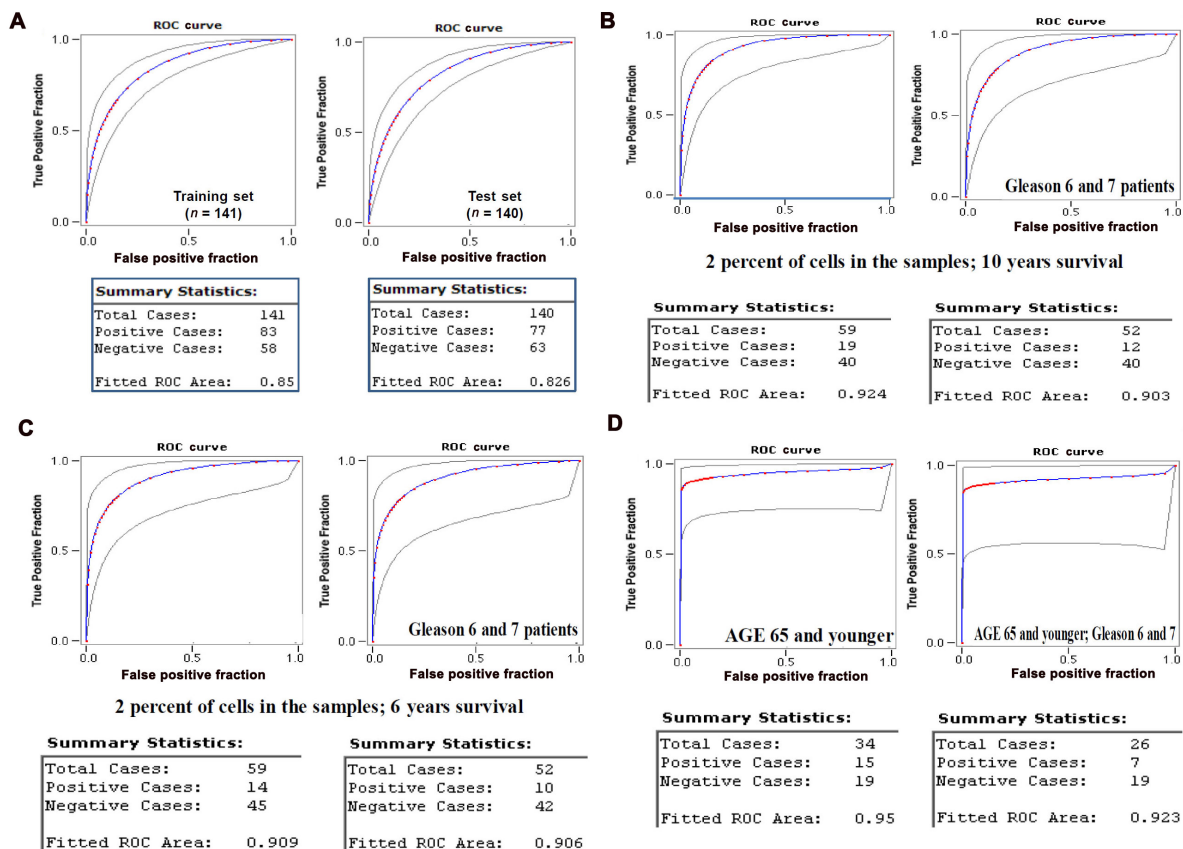

Figure 6: ROC area under the curve analysis of the patients' classification based on the 98-genes signature score in training (*n* = 141) and test (*n* = 140) groups (A) and different clinically-relevant sub-groups (B-D) of patients. ROC: receiver operating characteristic

prostate cancer at the time when no PSA screening programs were in place: these men had symptoms of benign prostatic hyperplasia (lower urinary tract symptoms) and were subsequently diagnosed with prostate cancer. No PSA analyses were performed and these patients did not receive treatment for prostate cancer. Therefore, the prostate cancer-specific death reported in this study represent the realistic natural history of prostate cancer with respect to the assessment of what happen to the prostate cancer patients if they will not be treated and just subjected to observations. In this contribution we evaluated the prostate cancer-specific death as the main outcome.

It has been reported that changes in prostate cancer screening practices in the United States have been associated with the recent decline in the overall incidence of prostate cancer and concomitant significant increase of the annual incidence of metastatic prostate cancer from 2007 to 2013.[23] Statistically significant increase in the annual incidence of metastatic prostate cancer in the United States from 2007 to 2013 has been consistently documented in all age groups with the overall increase of 72% in 2013 compared to 2004. Particularly alarming is the evidence of the greatest increase of the incidence of metastatic prostate cancer in men of the age group of 55 to 69 years who experienced 92% increase in the incidence of metastatic disease from 2004 to 2013.[23] These patients are likely to benefit most from definitive curative treatment of prostate cancer, suggesting that relaxed screening protocols and transition to active surveillance with curative intent strategy as a predominant approach for treatment of early-stage prostate cancer should be considered with extreme caution for this group of men, particularly in the absence of validated genetic tests reliably discriminating indolent prostate cancers from the clinically significant disease.

It is outside of the scope of this contribution to compare the diagnostic, prognostic, therapy-outcome assessments or targeted therapy-selection performances of GES. These questions were extensively explored and debated in the literature. There is no need to attempt a "horse race" comparing the signatures against each other while the state of the art comprehensive microarray and/or RNA-seq platforms enable the analyses of all of the signatures in one run and score all of the signatures simultaneously can be made for the specifically-defined benefits of the patients. It is reasonable to expect that no single signature will fit all clinically-defined disease diagnosis and management criteria and different signatures will address more adequately and resolve more efficiently

the specific needs, which will be ultimately tailored to the need of the particular individual patient.

DECLARATIONS

Authors' contributions
G.V. Glinsky contributed solely to the paper.

Conflicts of interest
There are no conflicts of interest.

Patient consent
Not applicable.

Ethics approval
Not applicable.

REFERENCES

1. American Cancer Society: American Cancer Society Cancer Facts and Figures 2010. Atlanta, GA: American Cancer Society; 2010.
2. Jemal A, Siegel R, Ward E, Hao Y, Xu J, Murray T, Thun MJ. Cancer statistics, 2008. *CA Cancer J Clin* 2008;58:71-96.
3. Johansson JE, Andrén O, Andersson SO, Dickman PW, Holmberg L, Magnuson A, Adami HO. Natural history of early, localized prostate cancer. *JAMA* 2004;291:2713-9.
4. Jemal A, Thun MJ, Ries LA, Howe HL, Weir HK, Center MM, Ward E, Wu XC, Eheman C, Anderson R, Ajani UA, Kohler B, Edwards BK. Annual report to the nation on the status of cancer, 1975-2005, featuring trends in lung cancer, tobacco use, and tobacco control. *J Natl Cancer Inst* 2008;100:1672-94.
5. Bill-Axelson A, Holmberg L, Ruutu M, Häggman M, Andersson SO, Bratell S, Spångberg A, Busch C, Nordling S, Garmo H, Palmgren J, Adami HO, Norlén BJ, Johansson JE; Scandinavian Prostate Cancer Group Study No. 4. Radical prostatectomy versus watchful waiting in early prostate cancer. *N Engl J Med* 2005;352:1977-84.
6. Bill-Axelson A, Holmberg L, Filén F, Ruutu M, Garmo H, Busch C, Nordling S, Häggman M, Andersson SO, Bratell S, Spångberg A, Palmgren J, Adami HO, Johansson JE; Scandinavian Prostate Cancer Group Study Number 4. Radical prostatectomy versus watchful waiting in localized prostate cancer: the Scandinavian prostate cancer group-4 randomized trial. *J Natl Cancer Inst* 2008;100:1144-54.
7. Bill-Axelson A, Holmberg L, Ruutu M, Garmo H, Stark JR, Busch C, Nordling S, Häggman M, Andersson SO, Bratell S, Spångberg A, Palmgren J, Steineck G, Adami HO, Johansson JE; SPCG-4 Investigators. Radical prostatectomy versus watchful waiting in early prostate cancer. *N Engl J Med* 2011;364:1708-17.

8. DeFrances CJ, Lucas CA, Buie VC, Golosinskiy A. 2006 National Hospital Discharge Survey. *Natl Health Stat Report* 2008;(5):1-20.

9. Draisma G, Boer R, Otto SJ, van der Cruijsen IW, Damhuis RA, Schröder FH, de Koning HJ. Lead times and overdetection due to prostate-specific antigen screening: estimates from the European Randomized Study of Screening for Prostate Cancer. *J Natl Cancer Inst* 2003;95:868-78.

10. Schröder FH, Hugosson J, Roobol MJ, Tammela TL, Ciatto S, Nelen V, Kwiatkowski M, Lujan M, Lilja H, Zappa M, Denis LJ, Recker F, Berenguer A, Määttänen L, Bangma CH, Aus G, Villers A, Rebillard X, van der Kwast T, Blijenberg BG, Moss SM, de Koning HJ, Auvinen A; ERSPC Investigators. Screening and prostate-cancer mortality in a randomized European study. *N Engl J Med* 2009;360:1320-8.

11. Hugosson J, Carlsson S, Aus G, Bergdahl S, Khatami A, Lodding P, Pihl CG, Stranne J, Holmberg E, Lilja H. Mortality results from the Göteborg randomised population-based prostate-cancer screening trial. *Lancet Oncol* 2010;11:725-32.

12. Stattin P, Holmberg E, Johansson JE, Holmberg L, Adolfsson J, Hugosson J; National Prostate Cancer Register (NPCR) of Sweden. Outcomes in localized prostate cancer: National Prostate Cancer Register of Sweden follow-up study. *J Natl Cancer Inst* 2010;102:950-8.

13. Tosoian JJ, Trock BJ, Landis P, Feng Z, Epstein JI, Partin AW, Walsh PC, Carter HB. Active surveillance program for prostate cancer: an update of the Johns Hopkins experience. *J Clin Oncol* 2011;29:2185-90.

14. Sboner A, Demichelis F, Calza S, Pawitan Y, Setlur SR, Hoshida Y, Perner S, Adami HO, Fall K, Mucci LA, Kantoff PW, Stampfer M, Andersson SO, Varenhorst E, Johansson JE, Gerstein MB, Golub TR, Rubin MA, Andrén O. Molecular sampling of prostate cancer: a dilemma for predicting disease progression. *BMC Med Genomics* 2010;3:8.

15. Glinskii AB, Ma J, Ma S, Grant D, Lim CU, Sell S, Glinsky GV. Identification of intergenic trans-regulatory RNAs containing a disease-linked SNP sequence and targeting cell cycle progression/differentiation pathways in multiple common human disorders. *Cell Cycle* 2009;8:3925-42.

16. Glinskii AB, Ma S, Ma J, Grant D, Lim CU, Guest I, Sell S, Buttyan R, Glinsky GV. Networks of intergenic long-range enhancers and snpRNAs drive castration-resistant phenotype of prostate cancer and contribute to pathogenesis of multiple common human disorders. *Cell Cycle* 2011;10:3571-97.

17. Glinsky GV, Berezovska O, Glinskii AB. Microarray analysis identifies a death-from-cancer signature predicting therapy failure in patients with multiple types of cancer. *J Clin Invest* 2005;115:1503-21.

18. Glinsky GV, Higashiyama T, Glinskii AB. Classification of human breast cancer using gene expression profiling as a component of the survival predictor algorithm. *Clin Cancer Res* 2004;10:2272-83.

19. Glinsky GV, Glinskii AB, Stephenson AJ, Hoffman RM, Gerald WL. Gene expression profiling predicts clinical outcome of prostate cancer. *J Clin Invest* 2004;113:913-23.

20. Glinsky GV, Krones-Herzig A, Glinskii AB, Gebauer G. Microarray analysis of xenograft-derived cancer cell lines representing multiple experimental models of human prostate cancer. *Mol Carcinog* 2003;37:209-21.

21. Varma S, Simon R. Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics* 2006;7:91.

22. Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM. Systematic determination of genetic network architecture. *Nat Genet* 1999;22:281-5.

23. Weiner AB, Matulewicz RS, Eggener SE, Schaeffer EM. Increasing incidence of metastatic prostate cancer in the United States (2004-2013). *Prostate Cancer Prostatic Dis* 2016;19:395-7.