



US 20100129804A1

(19) **United States**

(12) **Patent Application Publication**
Chinnaiyan et al.

(10) **Pub. No.: US 2010/0129804 A1**

(43) **Pub. Date: May 27, 2010**

(54) **SPINK1 AS A PROSTATE CANCER MARKER AND USES THEREOF**

(86) PCT No.: **PCT/US07/84090**

§ 371 (c)(1),
(2), (4) Date: **Jan. 14, 2010**

(76) Inventors: **Arul M. Chinnaiyan**, Plymouth, MI (US); **Scott A. Tomlins**, Ann Arbor, MI (US); **Daniel R. Rhodes**, Ann Arbor, MI (US); **Rohit Mehra**, Ann Arbor, MI (US)

Related U.S. Application Data

(60) Provisional application No. 60/857,758, filed on Nov. 8, 2006, provisional application No. 60/906,876, filed on Mar. 14, 2007.

Publication Classification

Correspondence Address:
Casimir Jones, S.C.
2275 DEMING WAY, SUITE 310
MIDDLETON, WI 53562 (US)

(51) **Int. Cl.**
C12Q 1/68 (2006.01)
G01N 33/53 (2006.01)

(52) **U.S. Cl.** **435/6; 435/7.1**

(57) **ABSTRACT**

(21) Appl. No.: **12/513,637**

Compositions and methods for cancer research, diagnosis, and treatment, including but not limited to, cancer markers are provided. In particular, SPINK1 and other markers for prostate cancer are provided.

(22) PCT Filed: **Nov. 8, 2007**

Figure 1A

Meta COPA Rank	Gene	# of Studies	Avg Rank
1	<i>ERG</i>	7	19.3
2	<i>SPINK1</i>	5	29.8
3	<i>GPR116</i>	5	46
4	<i>ORM1</i>	4	10
5	<i>ETV1</i>	4	23
6	<i>MYL2</i>	4	26.8
7	<i>NEB</i>	4	27
8	<i>TGM4</i>	4	30.8
9	<i>NELL2</i>	4	33.5
10	<i>KRT13</i>	4	49
11	<i>SLC26A4</i>	4	63.3

Figure 1B

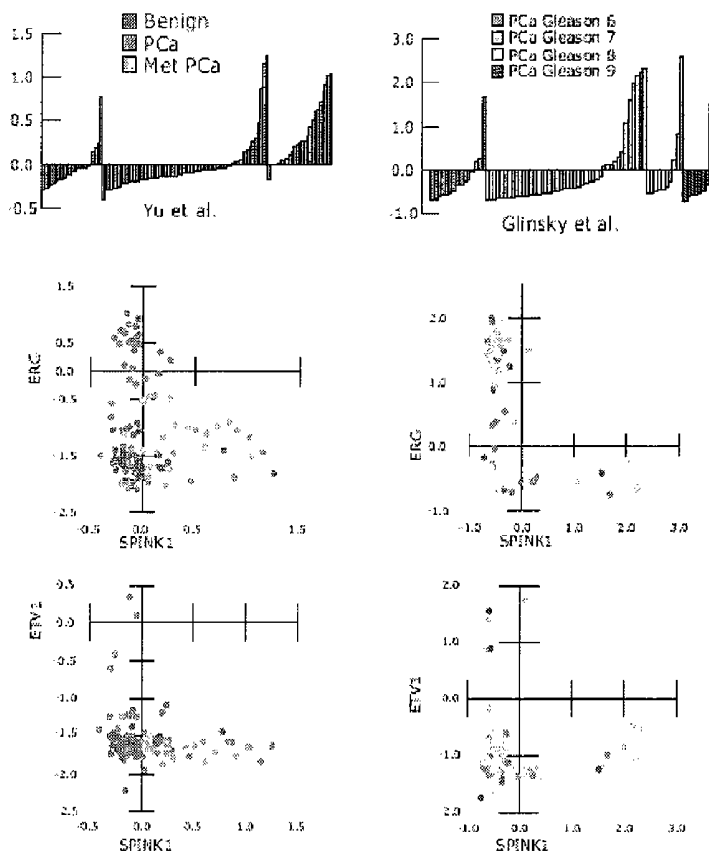


Figure 2

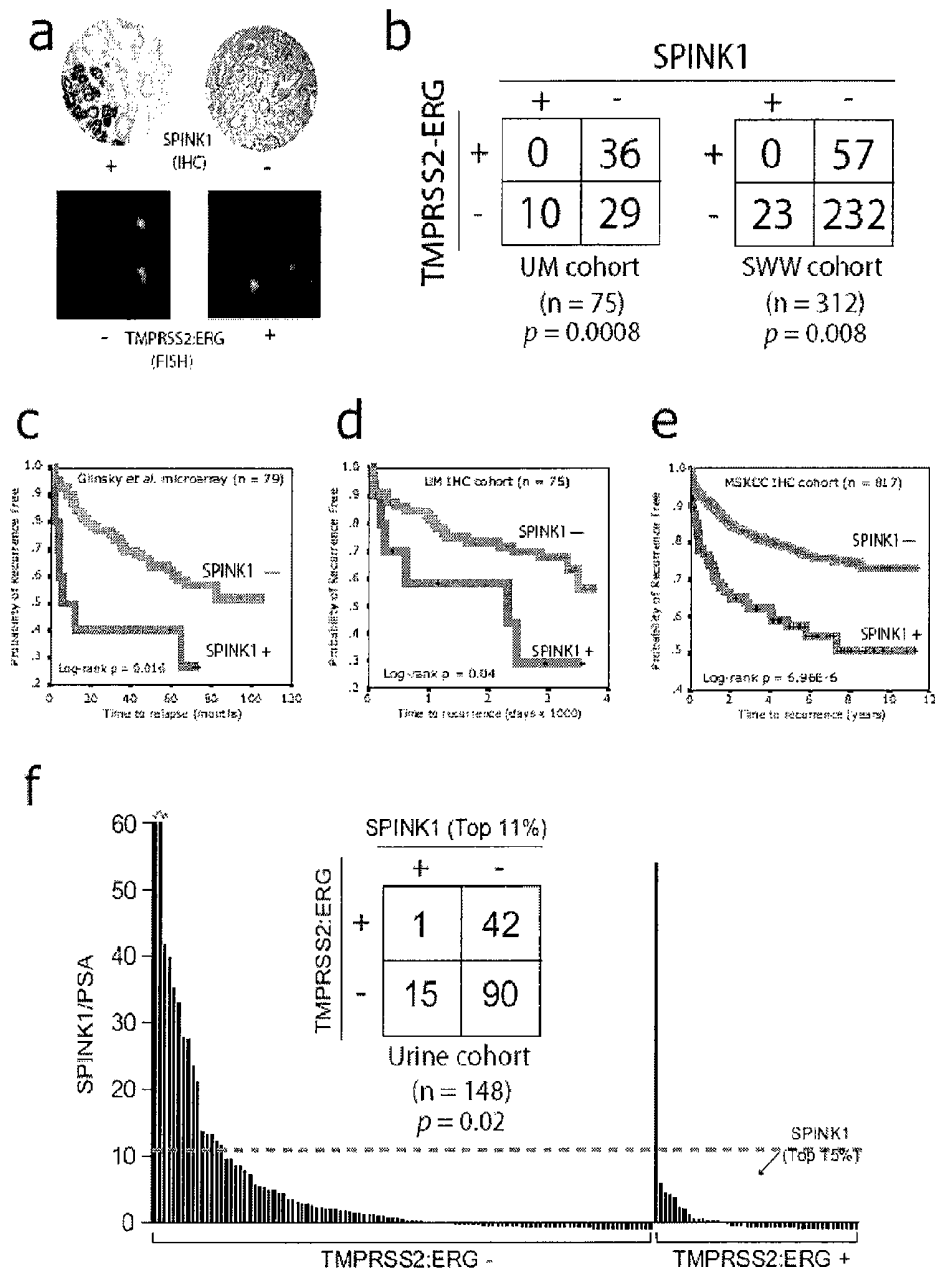


Figure 3

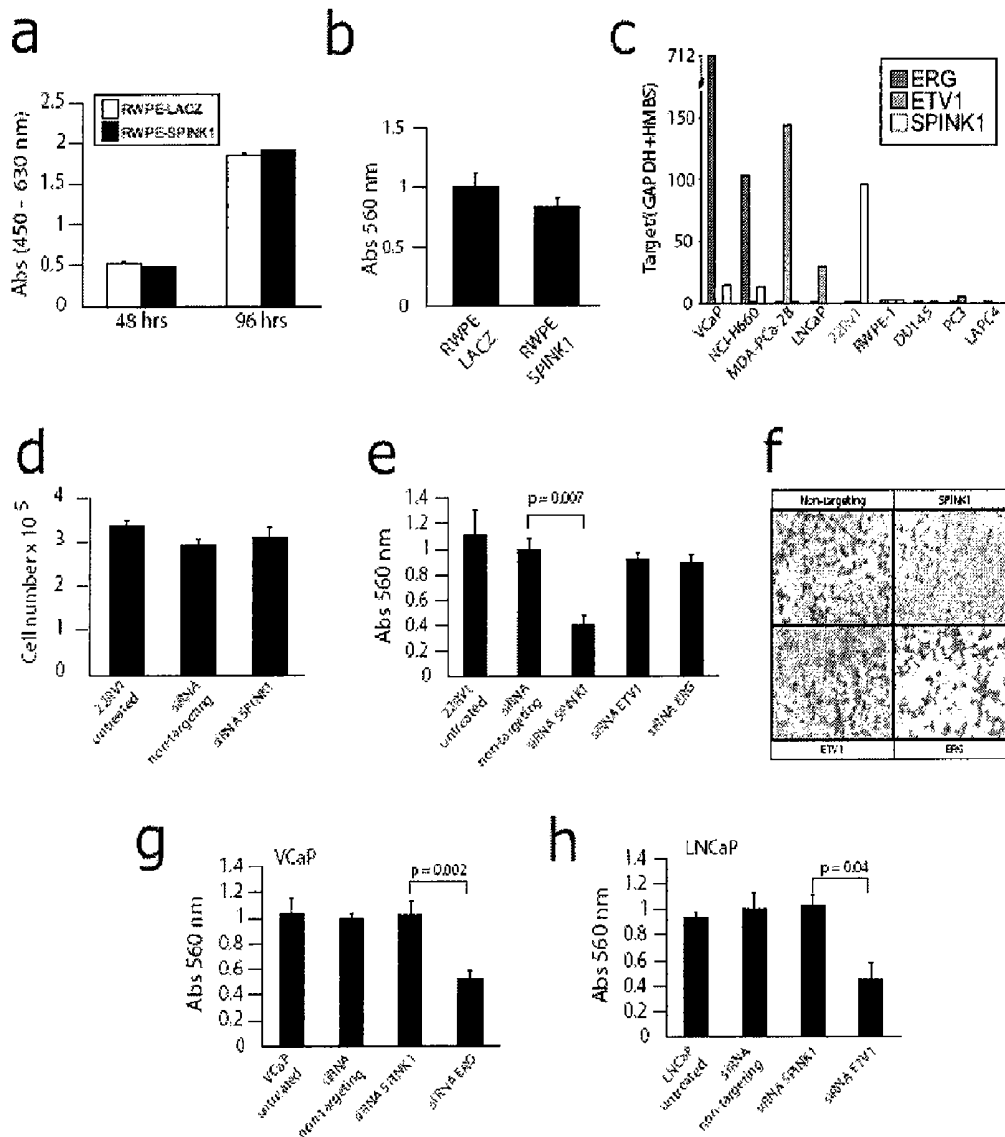
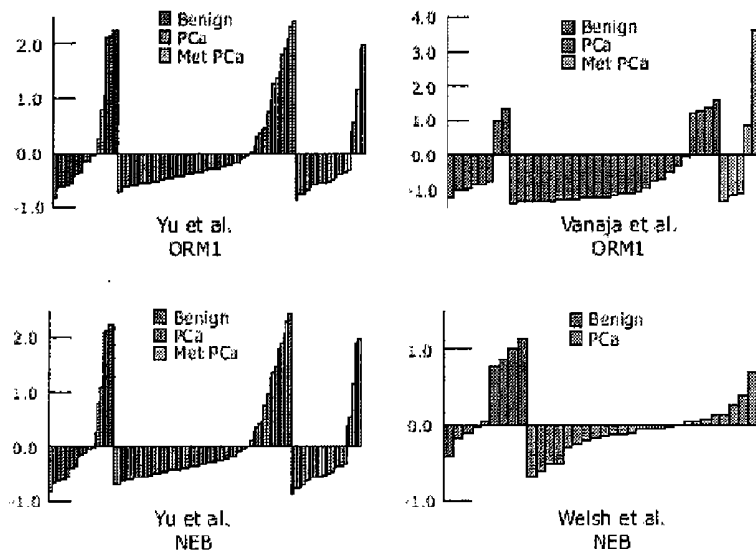


Figure 4

a



b

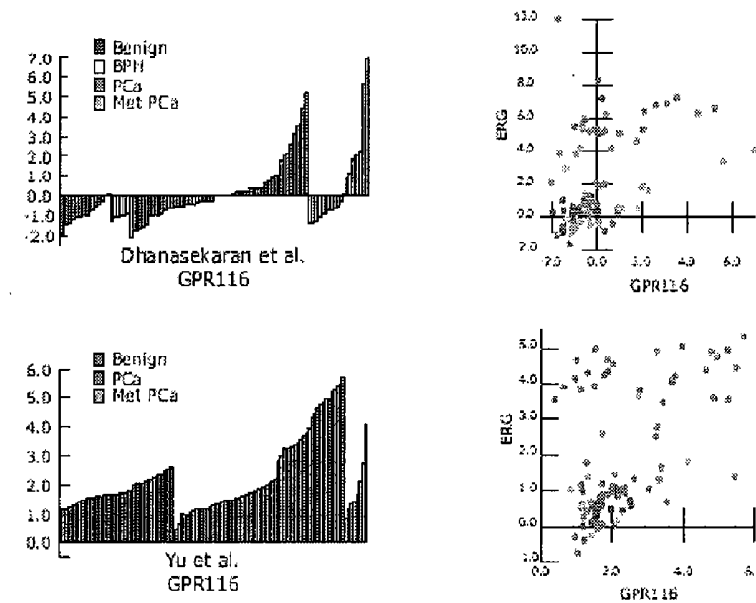


Figure 5

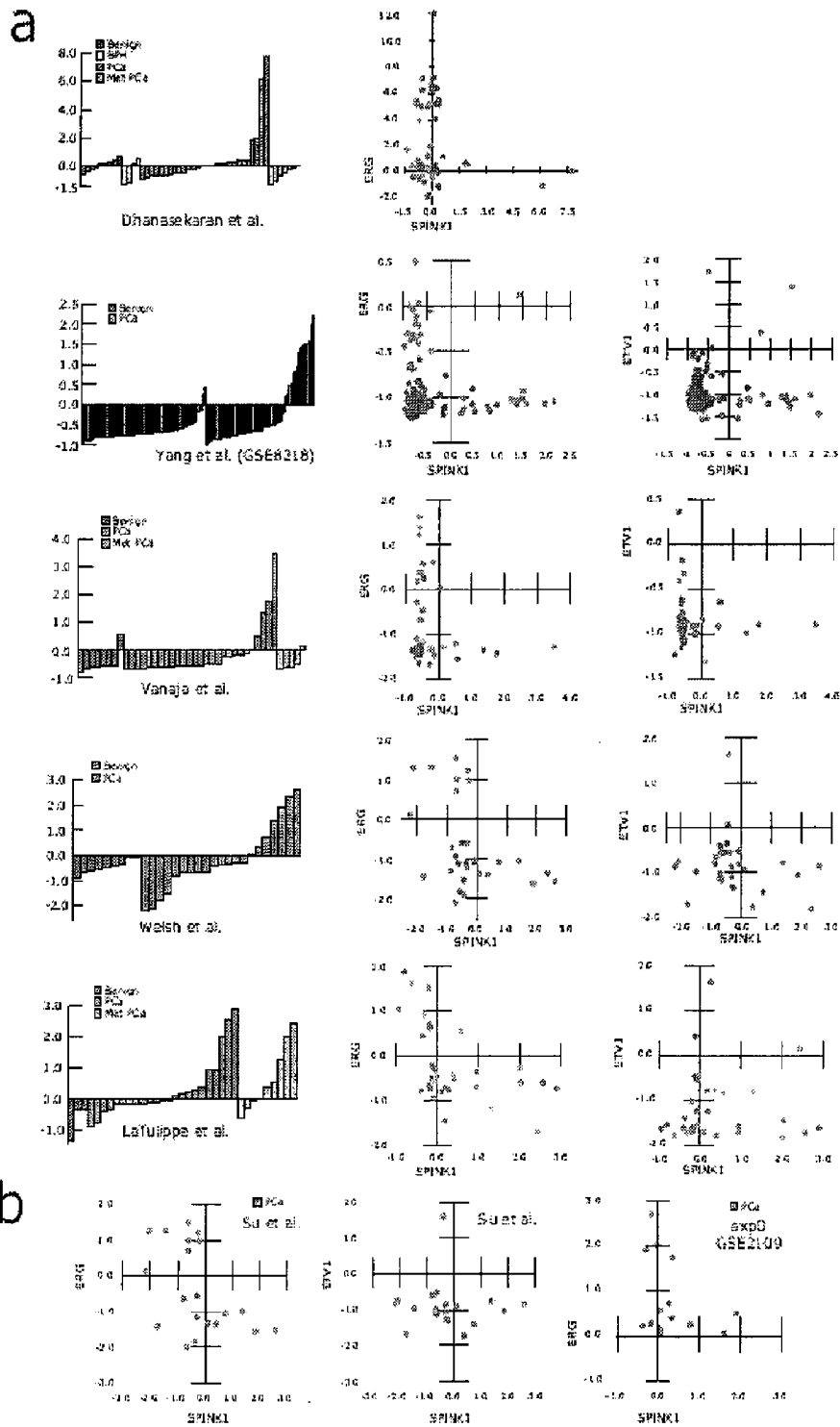


Figure 6

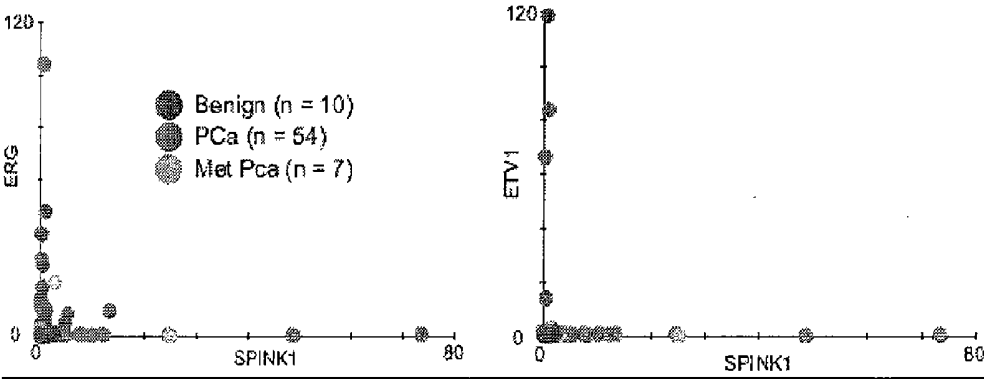


Figure 7

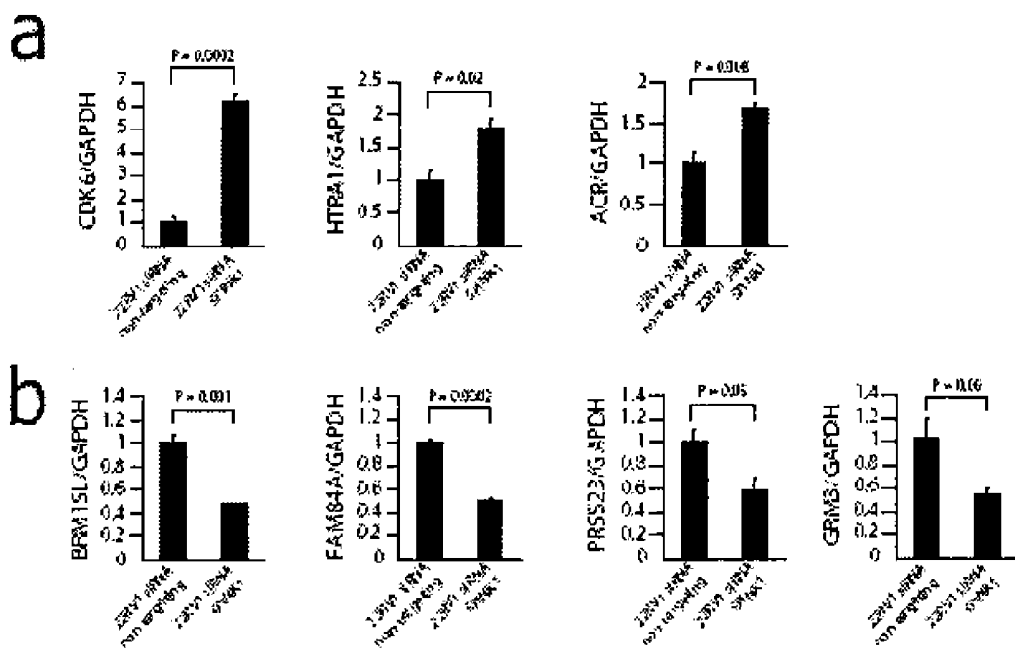


Figure 8

Yu et al.

R Rank	Gene	Reporter ID	R
1	ATM	2000_at	0.51
2	SPINK1	38582_at	0.51
3	CP	39008_at	0.51
4	CASP10	1326_at	0.39
5	PLAT	33452_at	0.39
6	EST	33559_at	0.39
7	DENND4B	35786_at	0.39
8	SERPINH	37259_at	0.39
9	SCNN1B	39682_at	0.39
10	NDS3	576_at	0.39
11	CPB1	41210_at	0.39
12	CASP10	39999_at	0.39
13	TMEM63A	39339_at	0.39
14	PLCH1	30869_at	0.39
15	RIT1	38331_at	0.39
16	RBPS	38789_at	0.39
17	TMEM63A	35442_at	0.39
18	EST	35906_at	0.39
19	EIF2B5	34758_at	0.39
20	HIST1H4B	33049_at	0.38
21	SMAS	41643_at	0.38
22	GUSBP1	41642_at	0.38
23	PTPRT	39227_at	0.38
24	EST	37956_at	0.38
25	IRF7	36412_s_at	0.38

Welsh et al.

R Rank	Gene	Reporter ID	R
1	HPGD	32570_at	0.67
2	HPGD	37322_s_at	0.67
3	SPINK1	38582_at	0.67
4	HPGD	37323_r_at	0.67
5	PRIM2A	122_at	0.50
6	FUT3	38405_s_at	0.50
7	EEF1A2	35174_l_at	0.50
8	PRIM2A	36898_r_at	0.50
9	PAH	33701_at	0.40
10	NPR3	34519_at	0.40
11	UGT2B4	37058_at	0.40

LaTulippe et al.

R Rank	Gene	Reporter ID	R
1	EST	33566_at	0.58
2	SPINK1	38582_at	0.58
3	FABP5	39799_at	0.58
4	HPGD	32570_at	0.42
5	NUP210	41813_at	0.42
6	HPGD	37322_s_at	0.42
7	EST	1525_s_at	0.33
8	EST	32413_at	0.33
9	BCKDHB	41883_l_at	0.33
10	AAK1	40572_at	0.33

Glinsky et al.

R Rank	Gene	Reporter ID	R
1	UGT1A6	204532_x_at	0.72
2	UGT1A6	206094_x_at	0.72
3	SPINK1	206239_s_at	0.72
4	UGT1A6	208596_s_at	0.72
5	UGT1A6	215125_s_at	0.72
6	UGT1A6	207126_x_at	0.72
7	F2RL1	206429_at	0.58
8	EST	213506_at	0.58
9	CS	205500_at	0.42
10	FANCF	205848_at	0.38
11	UGT2B4	206505_at	0.38
12	PPM1E	205938_at	0.38
13	LIMS1	207198_s_at	0.32
14	LIMS1	212687_at	0.32
15	ACSL5	218322_s_at	0.32
16	DNMBP	212838_at	0.32

Vanaja et al.

R Rank	Gene	Reporter ID	R
1	SPINK1	206239_s_at	0.52
2	TBC1D10C	228258_at	0.52
3	LY6K	223887_s_at	0.39
4	ITGB6	228535_at	0.39
5	SH3BGRL	201312_s_at	0.29
6	YIPF5	212341_at	0.29
7	EME1	234464_s_at	0.29
8	SEC22C	235288_at	0.29

Figure 9

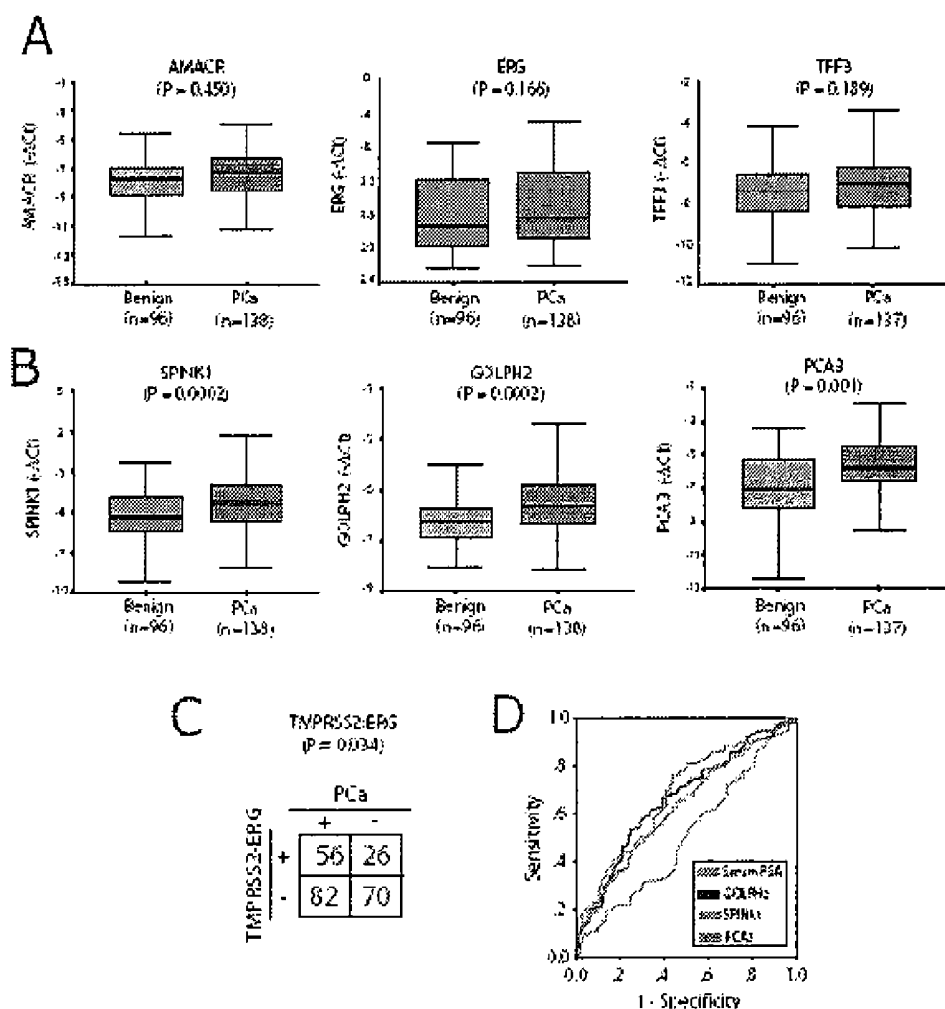
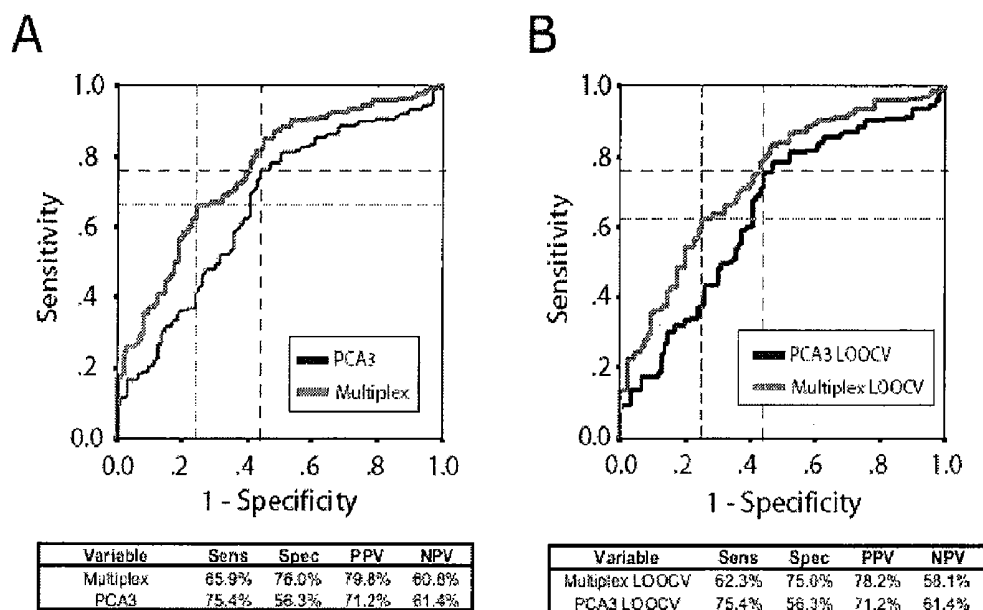


Figure 10



SPINK1 AS A PROSTATE CANCER MARKER AND USES THEREOF

[0001] This invention was made with government support under DA021519, CA046592, CA069568, CA102872 and CA111275 from the National Institutes of Health and DAMD17-03-2-0033 and W81XWH-06-1-0224 from the Army Medical Research and Materiel Command. The government has certain rights in the invention.

FIELD OF THE INVENTION

[0002] Compositions and methods for cancer research, diagnosis, and treatment, including but not limited to, cancer markers are provided. In particular, SPINK1 and other markers for prostate cancer are provided.

BACKGROUND OF THE INVENTION

[0003] Prostate cancer is the most common nondermatologic cancer and the second most common cause of cancer-related deaths in American men. The number of prostate cancers recorded in cancer registries in the United States and the United Kingdom has increased markedly in the past 15 years. This change predominantly represents an increase in the number of cancers diagnosed rather than a real increase in the number of cancers in the population. In 2006, 234,460 new cases and 27,350 deaths were estimated to occur. It was determined that approximately 91% of these new cases would be diagnosed at local or regional stages.

[0004] Prostate cancer (PCa) is typically diagnosed with a digital rectal exam and/or prostate specific antigen (PSA) screening. An elevated serum PSA level can indicate the presence of PCa. PSA is used as a marker for prostate cancer because it is secreted only by prostate cells. A healthy prostate will produce a stable amount—typically below 4 nanograms per milliliter, or a PSA reading of “4” or less—whereas cancer cells produce escalating amounts that correspond with the severity of the cancer. A level between 4 and 10 may raise a doctor’s suspicion that a patient has prostate cancer, while amounts above 50 may show that the tumor has spread elsewhere in the body.

[0005] When PSA or digital tests indicate a strong likelihood that cancer is present, a transrectal ultrasound (TRUS) is used to map the prostate and show any suspicious areas. Biopsies of various sectors of the prostate are used to determine if prostate cancer is present. Treatment options depend on the stage of the cancer. Men with a 10-year life expectancy or less who have a low Gleason number and whose tumor has not spread beyond the prostate are often treated with watchful waiting (no treatment). Treatment options for more aggressive cancers include surgical treatments, such as radical prostatectomy (RP) in which the prostate is completely removed (with or without nerve sparing techniques), and radiation, applied through an external beam that directs the dose to the prostate from outside the body or via low-dose radioactive seeds that are implanted within the prostate to kill cancer cells locally. Anti-androgen hormone therapy is also used, alone or in conjunction with surgery or radiation. Hormone therapy uses luteinizing hormone-releasing hormones (LH-RH) analogs, which block the pituitary from producing hormones that stimulate testosterone production. Patients must have injections of LH-RH analogs for the rest of their lives.

[0006] While surgical and hormonal treatments are often effective for localized PCa, advanced disease remains essentially incurable. Androgen ablation is the most common therapy for advanced PCa, leading to massive apoptosis of androgen-dependent malignant cells and temporary tumor regression. In most cases, however, the tumor reemerges with a vengeance and can proliferate independent of androgen signals.

[0007] The advent of PSA screening has led to earlier detection of PCa and significantly reduced PCa-associated fatalities. PSA screening is currently the single best test for prostate cancer and is widely used in the diagnosis of prostate cancer, but it does not help in determining whether the detected cancer will cause clinically significant disease. Whereas PSA is an excellent marker for the follow-up of patients with established prostate cancer, some men with prostate cancer may have normal PSA levels. A moderate elevation of the PSA level (4-10 ng/mL) has a low specificity for prostate cancer, and an elevated PSA level is not specific for prostate cancer. Elevated serum PSA levels may also be associated with prostatitis, prostate infarction, PIN, prostate biopsy, transurethral resection of the prostate, and urethral catheterization.

[0008] Because of the limitations of PSA screening, there have been efforts to improve its diagnostic specificity through the use of derivative indices such as PSA density, age-related PSA levels, TZ-PSA density, PSA velocity, free PSA levels, complexed PSA (cPSA) measurements, and free-to-total PSA ratios. The free-to-total PSA ratio measures both bound and free PSA as a percentage of total PSA and is a useful additional discriminator between cancer and benign pathology, especially in patients with moderately elevated serum PSA levels (4-10 ng/mL). This ratio is also useful in determining whether a repeat biopsy is appropriate in a patient with a moderately elevated PSA level whose initial systematic biopsy results are negative. The lower the percentage of free PSA, the higher the likelihood of cancer.

[0009] Thus, development of additional serum and tissue biomarkers to supplement PSA screening is needed.

SUMMARY OF THE INVENTION

[0010] A method is provided for identifying prostate cancer in a patient comprising: providing a sample containing prostate cells from the patient; detecting in the sample containing prostate cells overexpression of SPINK1 compared to normal expression of SPINK1; and, detecting in the sample containing prostate cells normal expression of ERG and/or ETV1, wherein detecting in the sample containing prostate cells mutually exclusive overexpression of SPINK1 compared to normal expression of ERG and/or ETV1 identifies prostate cancer in the patient. In some embodiments, the detecting step comprises detecting overexpression of SPINK1 RNA. In other embodiments, the detecting step comprises detecting overexpression of SPINK1 protein. In some embodiments, the sample containing prostate cells is prostate tissue, blood, urine, semen, prostatic secretions or isolated prostate cells. In some embodiments, detecting in the sample containing prostate cells overexpression of SPINK1 identifies invasive prostate cancer in the patient. In some embodiments, the sample containing prostate cells is from a patient following radical prostatectomy and overexpression of SPINK1 identifies recurrence of prostate cancer in the patient following radical prostatectomy.

[0011] A method is further provided for identifying prostate cancer in a patient comprising: providing a sample containing prostate cells from the patient; and, detecting in the sample containing prostate cells: (a) overexpression of SPINK1 compared to normal expression of SPINK1 and overexpression of PCA3 compared to normal expression of PCA3; (b) overexpression of SPINK1 compared to normal expression of SPINK1 and overexpression of GOLPH2 compared to normal expression of GOLPH2; (c) overexpression of SPINK1 compared to normal expression of SPINK1 and presence of TMRSS2:ERG; (d) overexpression of SPINK1 compared to normal expression of SPINK1, overexpression of PCA3 compared to normal expression of PCA3 and overexpression of GOLPH2 compared to normal expression of GOLPH2; (e) overexpression of SPINK1 compared to normal expression of SPINK1, overexpression of PCA3 compared to normal expression of PCA3 and presence of TMRSS2:ERG; (f) overexpression of SPINK1 compared to normal expression of SPINK1, overexpression of GOLPH2 compared to normal expression of GOLPH2 and presence of TMRSS2:ERG; or (g) overexpression of SPINK1 compared to normal expression of SPINK1, overexpression of PCA3 compared to normal expression of PCA3, overexpression of GOLPH2 compared to normal expression of GOLPH2 and presence of TMRSS2:ERG, wherein detecting in the sample containing prostate cells overexpression of SPINK1 identifies prostate cancer in the patient. In some embodiments, the detecting step comprises detecting overexpression of SPINK1 RNA. In other embodiments, the detecting step comprises detecting overexpression of SPINK1 protein. In some embodiments, the sample containing prostate cells is prostate tissue, blood, urine, semen, prostatic secretions or isolated prostate cells.

[0012] A composition is additionally provided comprising at least one of the following: (a) a first oligonucleotide probe comprising a sequence that hybridizes specifically to SPINK1 RNA or cDNA, a second oligonucleotide probe comprising a sequence that hybridizes specifically to ERG RNA or cDNA, and a third oligonucleotide probe comprising a sequence that hybridizes specifically to ETV1 RNA or cDNA; (b) a first pair of amplification oligonucleotides wherein each amplification oligonucleotide in the first pair comprises a sequence that hybridizes specifically to SPINK1 RNA or cDNA, a second pair of amplification oligonucleotides wherein each amplification oligonucleotide in the second pair comprises a sequence that hybridizes specifically to ERG RNA or cDNA, and a third pair of amplification oligonucleotides wherein each amplification oligonucleotide comprises a sequence that hybridizes specifically to ETV1 RNA or cDNA; or (c) a first antibody that binds specifically to SPINK1 protein, a second antibody that binds specifically to ERG protein, and a third antibody that binds specifically to ETV1 protein.

[0013] A composition is further provided comprising at least one of the following: (a) at least two oligonucleotide probes comprising: (i) an oligonucleotide probe comprising a sequence that hybridizes specifically to SPINK1 RNA or cDNA; and (ii) at least one other oligonucleotide probe comprising a sequence that hybridizes specifically to PCA3 RNA or cDNA, GOLPH2 RNA or cDNA, or a junction of a chimeric RNA or cDNA in which a 5' portion of the chimeric RNA is transcribed from a TMRSS2 gene and a 3' portion of the chimeric RNA is from transcribed from an ERG gene; (b) at least two pairs of amplification oligonucleotides compris-

ing: (i) a pair of amplification oligonucleotides wherein each amplification oligonucleotide comprises a sequence that hybridizes specifically to SPINK1 RNA or cDNA; and (ii) at least one other pair of amplification oligonucleotides wherein each amplification oligonucleotide comprises a sequence that hybridizes specifically to PCA3 RNA or cDNA, each amplification oligonucleotide comprises a sequence that hybridizes specifically to GOLPH2 RNA or cDNA, or a first amplification oligonucleotide comprises a sequence that hybridizes specifically to a 5' portion of a chimeric RNA transcribed from a TMRSS2 gene or its corresponding cDNA and a second amplification oligonucleotide comprises a sequence that hybridizes specifically to a 3' portion of the chimeric RNA transcribed from an ERG gene or its corresponding cDNA; or (c) at least two antibodies comprising: (i) an antibody that binds specifically to SPINK1 protein; and (ii) at least one other antibody that binds specifically to GOLPH2 protein, a native ERG protein, an amino-terminally truncated ERG protein encoded by a fusion of a TMRSS2 gene and an ERG gene, or a chimeric protein having an amino-terminal portion encoded by a TMRSS2 gene and a carboxy-terminal portion encoded by an ERG gene.

[0014] Additional embodiments are described herein.

DESCRIPTION OF THE FIGURES

[0015] FIG. 1 shows that meta COPA identified SPINK1 as a mutually exclusive outlier with ERG and ETV1 in prostate cancer. a. Genes were ranked by the number of studies in which they scored in the top 100 outliers (ranked by COPA) at any of the three pre-defined percentile cutoffs (75th, 90th, 95th). Genes were further ranked by their average COPA rank (Avg. Rank) in studies where they ranked in the top 100. b. The expression of SPINK1 and scatter plots of ERG vs. SPINK1 and ETV1 vs. SPINK1 expression are shown from two studies where SPINK1 ranked as a top 100 COPA outlier. Scatter plots are shown for ERG vs. SPINK1 (middle panels) and ETV1 vs. SPINK1 (lower panels) for all samples in both studies.

[0016] FIG. 2 shows that SPINK1 over-expression identifies an aggressive subset of ETS negative prostate cancers and can be detected non-invasively. a-b. SPINK1 expression was evaluated in two cohorts (University of Michigan (UM) and Swedish Watchful Waiting (SWW)) using immunohistochemistry (IHC) on tissue microarrays that have previously been evaluated for TMRSS2:ERG status by fluorescence in situ hybridization (FISH). a. Representative SPINK1 positive and negative cores are shown, along with cells from the same cores negative and positive for TMRSS2:ERG rearrangement by FISH. c. Contingency tables for SPINK1 expression and TMRSS2:ERG status and p-values for Fisher's exact tests for both cohorts are indicated. c-e. Relationship between SPINK1 outlier expression and biochemical recurrence after surgical resection. Kaplan-Meier analyses of outlier SPINK1 expression from the (c) Glinsky et al. DNA microarray dataset and SPINK1 IHC from the (d) UM and (e) Memorial Sloan Kettering Cancer Center (MSKCC) cohorts and biochemical recurrence after surgical resection are shown. f. Non-invasive detection of SPINK1 outlier-expression in men with TMRSS2:ERG negative prostate cancer. Contingency table for SPINK1 outlier expression and TMRSS2:ERG status and the Fisher's exact test p-value is shown.

[0017] FIG. 3 shows that knockdown of SPINK1 in 22RV1 prostate cancer cells attenuates invasiveness. a-b. The benign immortalized prostate cell line RWPE was infected with

SPINK1 or LACZ adenovirus as indicated and assayed for (a) proliferation or (b) invasion through a modified basement membrane. c. qPCR for SPINK1, ERG and ETV1 outlier expression. d-f. SPINK1 mediates invasiveness in 22RV1 cells. Cells were assayed for (d) proliferation and (e) invasion. Photomicrographs of invaded cells treated with the indicated siRNA are shown in f. g. VCaP (TMPRSS2:ERG positive) and g) LNCaP (ETV1 rearrangement positive) prostate cancer cell lines were treated with transfection reagent alone (untreated), or transfected with non-targeting nucleic acid or siRNA against SPINK1, ETV1 or ERG as indicated and assayed for invasion.

[0018] FIG. 4 shows meta-outlier genes showing over-expression in benign prostate tissue and ETS positive prostate cancers. a. The expression of meta-outlier genes ORM (ranked 4th) and NEB (ranked 7th) in normalized expression units are shown from the indicated studies, according to the sample classes described in FIG. 1, revealing outlier expression in multiple benign samples. b. The expression of the 3rd ranked meta-outlier gene GPR116 (left panels) and scatter plots of GPR116 vs. ERG (right panels) for all profiled samples in two studies shows co-outlier expression of GPR116 and ERG in multiple samples.

[0019] FIG. 5 shows over-expression of SPINK1 in prostate cancer compared to benign prostate tissue and mutually exclusive over-expression with ERG and ETV1 in DNA microarray studies. The expression of SPINK1 and scatter plots of SPINK1 vs. ERG and SPINK1 vs. ETV1 (if measured) for five studies profiling distinct classes of prostate tissue (a) and two studies profiling prostate cancers as part of multi-cancer studies (b) are shown as in FIG. 1.

[0020] FIG. 6 shows over-expression of SPINK1 in prostate cancer compared to benign prostate tissue and mutually exclusive over-expression with ERG and ETV1. a. Scatter plots of ERG vs. SPINK1 (left panel) and ETV1 vs. SPINK1 (right panel) by qPCR in 10 benign prostate samples, 54 localized prostate cancers (PCa) and 7 metastatic (Met) PCa samples.

[0021] FIG. 7 shows qPCR confirmation of genes differentially expressed upon SPINK1 knockdown in 22RV1 cells. Selected a) over- and b) under-expressed genes in 22RV1 siSPINK1 cells were assessed by quantitative PCR as shown.

[0022] FIG. 8 shows identification of genes showing co-expression with SPINK1 across multiple prostate cancer profiling studies.

[0023] FIG. 9 shows characterization of candidate urine-based biomarkers of prostate cancer. A-C. Quantitative PCR (qPCR) of whole transcriptome amplified (WTA) cDNA from urine obtained from patients presenting for needle biopsy or prostatectomy. Biomarker expression in patients with negative needle biopsies or patients with prostate cancer are shown. The $-\Delta\text{Ct}$ values of genes that were not significant predictors of prostate cancer by univariate analysis (see Table 5) are shown in A, and the expression of those that were significant predictors are shown in B and C. D. Receiver operator characteristic (ROC) curves for individual variables for the diagnosis of prostate cancer.

[0024] FIG. 10 shows a multiplexed set of urine biomarkers out performs PCA3 alone in the detection of prostate cancer. A. Multivariate regression analysis resulted in a multiplexed model including SPINK1, PCA3, GOLPH2 and TMPRSS2:ERG as significant predictors of prostate cancer (see Table 5). The point on the ROC curve with the maximum sum of sensitivity (Sens) and specificity (Spec) is indicated by the

dashed line, and the positive and negative predictive values (PPV and NPV, respectively) are given. B. As in A, except a leave-one-out cross validation (LOOCV) strategy was used to generate unbiased area under curves.

GENERAL DESCRIPTION

[0025] When applied to the Oncomine database (Rhodes et al., Proc Natl Acad Sci USA 101, 9309 [2004]; Rhodes et al., Neoplasia 6, 1 [2004]), the methodology termed Cancer Outlier Profile Analysis (COPA) correctly identified several known oncogenes, including PBX1 in leukemia and CCND1 in multiple myeloma (Tomlins et al., Science 310, 644 [2005]). In addition, COPA nominated ETS family genes as candidate oncogenes in prostate cancer prompting the discovery of recurrent chromosomal rearrangements involving ERG or ETV1 and the androgen-regulated gene TMPRSS2 (Tomlins et al., [2005], supra).

[0026] As 50-70% of prostate cancers harbor TMPRSS2:ETS gene fusions, experiments were conducted to identify additional candidate oncogenes in prostate cancers. Experiments conducted used a meta-analysis of COPA applied to 7 prostate cancer profiling studies and analyzed candidates for outlier expression in prostate cancer and mutually exclusive over-expression with ERG and ETV1. SPINK1, which was identified as the 2nd ranked meta-outlier, met both criteria across 8 data sets. SPINK1 showed marked overexpression in 50 of 325 (15.4%) profiled prostate cancers, but only 1 of 56 (1.8%) benign prostate tissue samples. In all 325 profiled prostate cancer samples, SPINK1, ERG and ETV1 showed mutually exclusive outlier expression. The over-expression of SPINK1 in a fraction of cancer samples compared to benign prostate tissue and the mutually exclusive over-expression of SPINK1, ERG and ETV1 was confirmed by quantitative PCR. Fluorescence in situ hybridization from tissue from one of the localized prostate cancers over-expressing SPINK1 did not reveal gene rearrangements or amplification, indicating that SPINK1 is up-regulated through increased transcription. Together these results, consistent across different assays, microarray platforms, laboratories and sample cohorts, demonstrate that SPINK1 is exclusively over-expressed in prostate cancers without TMPRSS2:ETS gene fusions (as indicated by ERG or ETV1 over-expression).

[0027] Further experiments demonstrated that SPINK1 outlier expression was correlated with an increase in prostate cancer recurrence (e.g., following surgery). Accordingly, in some embodiments, the methods of screening samples to determine likelihood of prostate cancer recurrence are provided. Subjects at increased risk of recurrence can then be offered more aggressive treatment or additional therapies. Conversely, subjects found to be at decreased risk of recurrence or lacking increased risk of recurrence can be spared the side effects of unnecessary treatments.

DEFINITIONS

[0028] To facilitate an understanding of the disclosure and claims described herein, a number of terms and phrases are defined below:

[0029] As used herein, the term "outlier expression of SPINK1" refers to an altered level of expression of SPINK1 nucleic acid (e.g., mRNA) or protein relative to the level normally found (e.g., the level in a subject not diagnosed with cancer). In some embodiments, normal levels are the average level in a population of one or more individuals not diagnosed

with cancer. In other embodiments, normal levels are determined within other tissues of the individual to be diagnosed. In some embodiments, expression is altered by at least 10%, preferably at least 20%, even more preferably at least 50%, yet more preferably at least 75%, still more preferably at least 90%, and most preferably at least 100% relative to the level of expression normally found (e.g., in non-cancerous tissue). Expression levels may be determined using any suitable method, including, but not limited to, those disclosed herein (e.g., Example 1 below). In some embodiments, samples positive for outlier expression of SPINK1 are those whose expression differs by greater than about 0.1, preferably greater than 0.2, and even more preferably greater than 0.5 normalized expression units. Normalized expression units may be calculated using any suitable method, including, but not limited to, those described in the experimental section below.

[0030] As used herein, the term “overexpression of SPINK1” refers to a higher level of expression of SPINK1 nucleic acid (e.g., mRNA) or protein relative to the level normally found. In some embodiments, expression is increased at least 10%, preferably at least 20%, even more preferably at least 50%, yet more preferably at least 75%, still more preferably at least 90%, and most preferably at least 100% relative to the level of expression normally found. The level of expression normally found may be determined using any number of suitable parameters. Examples include, but are not limited to, the level in non-cancerous prostate (e.g., an average of the level of SPINK1 expression in prostate tissues from multiple subjects not diagnosed with prostate cancer), the level in non-cancerous tissues (e.g., an average of the level of SPINK1 expression in non-prostate tissues from multiple subjects not diagnosed with cancer), the level in non-cancerous prostate cell lines, or a relative level of expression (e.g., the level over time in the same individual). Expression levels may be determined using any suitable method, including, but not limited to, those disclosed herein. In some embodiments, expression levels are compared to the level of expression of a known gene (e.g., the level of expression or the relative expression). In some embodiments, the known gene is PSA.

[0031] As used herein, the term “gene expression associated with prostate cancer recurrence” refers to a gene expression profile (e.g., outlier expression of SPINK1) associated with prostate cancer recurrence (e.g., in the prostate or metastatic) following treatment (e.g., surgery) for a primary tumor. In some embodiments, prostate cancer recurrence is increased at least 10%, preferably at least 20%, even more preferably at least 50%, yet more preferably at least 75%, still more preferably at least 90%, and most preferably at least 100% relative to the level of recurrence in representative subject population (e.g., average of a large population (e.g., one or more, preferably 100 or more, even more preferably 1000 or more and still more preferably 10,000 or more subjects) of subjects lacking “outlier expression of SPINK1”).

[0032] The term “epitope” as used herein refers to that portion of an antigen that makes contact with a particular antibody.

[0033] When a protein or fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to a given region or three-dimensional structure on the protein; these regions or structures are referred to as “antigenic determinants”. An antigenic determinant may compete with the

intact antigen (i.e., the “immunogen” used to elicit the immune response) for binding to an antibody.

[0034] The terms “specific binding” or “specifically binding” when used in reference to the interaction of an antibody and a protein or peptide means that the interaction is dependent upon the presence of a particular structure (i.e., the antigenic determinant or epitope) on the protein; in other words the antibody is recognizing and binding to a specific protein structure rather than to proteins in general. For example, if an antibody is specific for epitope “A,” the presence of a protein containing epitope A (or free, unlabelled A) in a reaction containing labeled “A” and the antibody will reduce the amount of labeled A bound to the antibody.

[0035] As used herein, the terms “non-specific binding” and “background binding” when used in reference to the interaction of an antibody and a protein or peptide refer to an interaction that is not dependent on the presence of a particular structure (i.e., the antibody is binding to proteins in general rather than a particular structure such as an epitope).

[0036] As used herein, the term “subject” refers to any animal (e.g., a mammal), including, but not limited to, humans, non-human primates, rodents, and the like, which is to be the recipient of a particular treatment. Typically, the terms “subject” and “patient” are used interchangeably herein in reference to a human subject.

[0037] As used herein, the term “subject suspected of having cancer” refers to a subject that presents one or more symptoms indicative of a cancer (e.g., a noticeable lump or mass) or is being screened for a cancer (e.g., during a routine physical). A subject suspected of having cancer may also have one or more risk factors. A subject suspected of having cancer has generally not been tested for cancer. However, a “subject suspected of having cancer” encompasses an individual who has received an initial diagnosis (e.g., a CT scan showing a mass or increased PSA level) but for whom the stage of cancer is not known. The term further includes people who once had cancer (e.g., an individual in remission).

[0038] As used herein, the term “characterizing cancer in subject” refers to the identification of one or more properties of a cancer sample in a subject, including but not limited to, the presence of benign, pre-cancerous or cancerous tissue, the stage of the cancer, and the subject’s prognosis. Cancers may be characterized by the identification of the expression of one or more cancer marker genes, including but not limited to, the cancer markers disclosed herein.

[0039] As used herein, the term “characterizing prostate tissue in a subject” refers to the identification of one or more properties of a prostate tissue sample (e.g., including but not limited to, the presence of cancerous tissue, the presence of pre-cancerous tissue that is likely to become cancerous, and the presence of cancerous tissue that is likely to metastasize). In some embodiments, tissues are characterized by the identification of the expression of one or more cancer marker genes, including but not limited to, the cancer markers disclosed herein.

[0040] As used herein, the term “cancer marker genes” refers to a gene whose expression level, alone or in combination with other genes, is correlated with cancer or prognosis of cancer. The correlation may relate to either an increased or decreased expression of the gene. For example, the expression of the gene may be indicative of cancer, or lack of expression of the gene may be correlated with poor prognosis in a cancer patient.

[0041] As used herein, the term “a reagent that specifically detects the presence or absence of a cancer marker” refers to reagents used to detect the expression of one or more cancer markers (e.g., including but not limited to, the cancer markers described herein). Examples of suitable reagents include but are not limited to, nucleic acid probes capable of specifically hybridizing to the gene of interest, PCR primers capable of specifically amplifying the gene of interest, and antibodies capable of specifically binding to proteins expressed by the gene of interest. Other non-limiting examples can be found in the description and examples below.

[0042] As used herein, the term “instructions for using said kit for detecting cancer in said subject” includes instructions for using the reagents contained in the kit for the detection and characterization of cancer in a sample from a subject. In some embodiments, the instructions further comprise the statement of intended use required by the U.S. Food and Drug Administration (FDA) in labeling in vitro diagnostic products.

[0043] As used herein, the term “stage of cancer” refers to a qualitative or quantitative assessment of the level of advancement of a cancer. Criteria used to determine the stage of a cancer include, but are not limited to, the size of the tumor, whether the tumor has spread to other parts of the body and where the cancer has spread (e.g., within the same organ or region of the body or to another organ).

[0044] As used herein, the term “providing a prognosis” refers to providing information regarding the impact of the presence of cancer (e.g., as determined by the diagnostic methods described herein) on a subject’s future health (e.g., expected morbidity or mortality, the likelihood of getting cancer, and the risk of metastasis).

[0045] As used herein, the term “initial diagnosis” refers to results of initial cancer diagnosis (e.g. the presence or absence of cancerous cells). An initial diagnosis does not include information about the stage of the cancer or the risk of prostate specific antigen failure.

[0046] As used herein, the term “biopsy tissue” refers to a sample of tissue (e.g., prostate tissue) that is removed from a subject for the purpose of determining if the sample contains cancerous tissue. In some embodiment, biopsy tissue is obtained because a subject is suspected of having cancer. The biopsy tissue is then examined (e.g., by microscopy) for the presence or absence of cancer.

[0047] As used herein, the term “non-human animals” refers to all non-human animals including, but are not limited to, vertebrates such as rodents, non-human primates, ovines, bovines, ruminants, lagomorphs, porcines, caprines, equines, canines, felines, aves, etc.

[0048] The term “gene” refers to a nucleic acid (e.g., DNA) sequence that comprises coding sequences necessary for the production of a polypeptide, precursor, or RNA (e.g., rRNA, tRNA). The polypeptide can be encoded by a full length coding sequence or by any portion of the coding sequence so long as the desired activity or functional properties (e.g., enzymatic activity, ligand binding, signal transduction, immunogenicity, etc.) of the full-length or fragment are retained. The term also encompasses the coding region of a structural gene and the sequences located adjacent to the coding region on both the 5' and 3' ends for a distance of about 1 kb or more on either end such that the gene corresponds to the length of the full-length mRNA. Sequences located 5' of the coding region and present on the mRNA are referred to as 5' non-translated sequences. Sequences located 3' or downstream of the coding region and present on the mRNA are

referred to as 3' non-translated sequences. The term “gene” encompasses both cDNA and genomic forms of a gene. A genomic form or clone of a gene contains the coding region interrupted with non-coding sequences termed “introns” or “intervening regions” or “intervening sequences.” Introns are segments of a gene that are transcribed into nuclear RNA (hnRNA); introns may contain regulatory elements such as enhancers. Introns are removed or “spliced out” from the nuclear or primary transcript; introns therefore are absent in the messenger RNA (mRNA) transcript. The mRNA functions during translation to specify the sequence or order of amino acids in a nascent polypeptide.

[0049] As used herein, the term “heterologous gene” refers to a gene that is not in its natural environment. For example, a heterologous gene includes a gene from one species introduced into another species. A heterologous gene also includes a gene native to an organism that has been altered in some way (e.g., mutated, added in multiple copies, linked to non-native regulatory sequences, etc). Heterologous genes are distinguished from endogenous genes in that the heterologous gene sequences are typically joined to DNA sequences that are not found naturally associated with the gene sequences in the chromosome or are associated with portions of the chromosome not found in nature (e.g., genes expressed in loci where the gene is not normally expressed).

[0050] As used herein, the term “gene expression” refers to the process of converting genetic information encoded in a gene into RNA (e.g., mRNA, rRNA, tRNA, or snRNA) through “transcription” of the gene (i.e., via the enzymatic action of an RNA polymerase), and for protein encoding genes, into protein through “translation” of mRNA. Gene expression can be regulated at many stages in the process. “Up-regulation” or “activation” refers to regulation that increases the production of gene expression products (i.e., RNA or protein), while “down-regulation” or “repression” refers to regulation that decrease production. Molecules (e.g., transcription factors) that are involved in up-regulation or down-regulation are often called “activators” and “repressors,” respectively.

[0051] In addition to containing introns, genomic forms of a gene may also include sequences located on both the 5' and 3' end of the sequences that are present on the RNA transcript. These sequences are referred to as “flanking” sequences or regions (these flanking sequences are located 5' or 3' to the non-translated sequences present on the mRNA transcript). The 5' flanking region may contain regulatory sequences such as promoters and enhancers that control or influence the transcription of the gene. The 3' flanking region may contain sequences that direct the termination of transcription, post-transcriptional cleavage and polyadenylation.

[0052] As used herein, the term “oligonucleotide,” refers to a short length of single-stranded polynucleotide chain. Oligonucleotides are typically less than 200 residues long (e.g., between 15 and 100), however, as used herein, the term is also intended to encompass longer polynucleotide chains. Oligonucleotides are often referred to by their length. For example a 24 residue oligonucleotide is referred to as a “24-mer”. Oligonucleotides can form secondary and tertiary structures by self-hybridizing or by hybridizing to other polynucleotides. Such structures can include, but are not limited to, duplexes, hairpins, cruciforms, bends, and triplexes.

[0053] As used herein, the terms “complementary” or “complementarity” are used in reference to polynucleotides (i.e., a sequence of nucleotides) related by the base-pairing

rules. For example, the sequence "5'-A-G-T-3'," is complementary to the sequence "3'-T-C-A-5'." Complementarity may be "partial," in which only some of the nucleic acids' bases are matched according to the base pairing rules. Or, there may be "complete" or "total" complementarity between the nucleic acids. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands. This is of particular importance in amplification reactions, as well as detection methods that depend upon binding between nucleic acids.

[0054] The term "homology" refers to a degree of complementarity. There may be partial homology or complete homology (i.e., identity). A partially complementary sequence is a nucleic acid molecule that at least partially inhibits a completely complementary nucleic acid molecule from hybridizing to a target nucleic acid is "substantially homologous." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or Northern blot, solution hybridization and the like) under conditions of low stringency. A substantially homologous sequence or probe will compete for and inhibit the binding (i.e., the hybridization) of a completely homologous nucleic acid molecule to a target under conditions of low stringency. This is not to say that conditions of low stringency are such that non-specific binding is permitted; low stringency conditions require that the binding of two sequences to one another be a specific (i.e., selective) interaction. The absence of non-specific binding may be tested by the use of a second target that is substantially non-complementary (e.g., less than about 30% identity); in the absence of non-specific binding the probe will not hybridize to the second non-complementary target.

[0055] When used in reference to a double-stranded nucleic acid sequence such as a cDNA or genomic clone, the term "substantially homologous" refers to any probe that can hybridize to either or both strands of the double-stranded nucleic acid sequence under conditions of low stringency as described above.

[0056] When used in reference to a single-stranded nucleic acid sequence, the term "substantially homologous" refers to any probe that can hybridize (i.e., it is the complement of) the single-stranded nucleic acid sequence under conditions of low stringency as described above.

[0057] As used herein, the term "hybridization" is used in reference to the pairing of complementary nucleic acids. Hybridization and the strength of hybridization (i.e., the strength of the association between the nucleic acids) is impacted by such factors as the degree of complementarity between the nucleic acids, stringency of the conditions involved, the T_m of the formed hybrid, and the G:C ratio within the nucleic acids. A single molecule that contains pairing of complementary nucleic acids within its structure is said to be "self-hybridized."

[0058] As used herein, the term " T_m " is used in reference to the "melting temperature." The melting temperature is the temperature at which a population of double-stranded nucleic acid molecules becomes half dissociated into single strands. The equation for calculating the T_m of nucleic acids is well known in the art. As indicated by standard references, a simple estimate of the T_m value may be calculated by the equation: $T_m = 81.5 + 0.41(\% G+C)$, when a nucleic acid is in aqueous solution at 1 M NaCl (See e.g., Anderson and Young,

Quantitative Filter Hybridization, in Nucleic Acid Hybridization [1985]). Other references include more sophisticated computations that take structural as well as sequence characteristics into account for the calculation of T_m .

[0059] As used herein the term "stringency" is used in reference to the conditions of temperature, ionic strength, and the presence of other compounds such as organic solvents, under which nucleic acid hybridizations are conducted. Under "low stringency conditions" a nucleic acid sequence of interest will hybridize to its exact complement, sequences with single base mismatches, closely related sequences (e.g., sequences with 90% or greater homology), and sequences having only partial homology (e.g., sequences with 50-90% homology). Under "medium stringency conditions," a nucleic acid sequence of interest will hybridize only to its exact complement, sequences with single base mismatches, and closely related sequences (e.g., 90% or greater homology). Under "high stringency conditions," a nucleic acid sequence of interest will hybridize only to its exact complement, and (depending on conditions such as temperature) sequences with single base mismatches. In other words, under conditions of high stringency the temperature can be raised so as to exclude hybridization to sequences with single base mismatches.

[0060] "High stringency conditions" when used in reference to nucleic acid hybridization comprise conditions equivalent to binding or hybridization at 42° C. in a solution consisting of 5xSSPE (43.8 g/l NaCl, 6.9 g/l $\text{NaH}_2\text{PO}_4\cdot\text{H}_2\text{O}$ and 1.85 g/l EDTA, pH adjusted to 7.4 with NaOH), 0.5% SDS, 5xDenhardt's reagent and 100 µg/ml denatured salmon sperm DNA followed by washing in a solution comprising 0.1xSSPE, 1.0% SDS at 42° C. when a probe of about 500 nucleotides in length is employed.

[0061] "Medium stringency conditions" when used in reference to nucleic acid hybridization comprise conditions equivalent to binding or hybridization at 42° C. in a solution consisting of 5xSSPE (43.8 g/l NaCl, 6.9 g/l $\text{NaH}_2\text{PO}_4\cdot\text{H}_2\text{O}$ and 1.85 g/l EDTA, pH adjusted to 7.4 with NaOH), 0.5% SDS, 5xDenhardt's reagent and 100 µg/ml denatured salmon sperm DNA followed by washing in a solution comprising 1.0xSSPE, 1.0% SDS at 42° C. when a probe of about 500 nucleotides in length is employed.

[0062] "Low stringency conditions" comprise conditions equivalent to binding or hybridization at 42° C. in a solution consisting of 5xSSPE (43.8 g/l NaCl, 6.9 g/l $\text{NaH}_2\text{PO}_4\cdot\text{H}_2\text{O}$ and 1.85 g/l EDTA, pH adjusted to 7.4 with NaOH), 0.1% SDS, 5xDenhardt's reagent [50xDenhardt's contains per 500 ml: 5 g Ficoll (Type 400, Pharmacia), 5 g BSA (Fraction V; Sigma)] and 100 µg/ml denatured salmon sperm DNA followed by washing in a solution comprising 5xSSPE, 0.1% SDS at 42° C. when a probe of about 500 nucleotides in length is employed.

[0063] The art knows well that numerous equivalent conditions may be employed to comprise low stringency conditions; factors such as the length and nature (DNA, RNA, base composition) of the probe and nature of the target (DNA, RNA, base composition, present in solution or immobilized, etc.) and the concentration of the salts and other components (e.g., the presence or absence of formamide, dextran sulfate, polyethylene glycol) are considered and the hybridization solution may be varied to generate conditions of low stringency hybridization different from, but equivalent to, the above listed conditions. In addition, the art knows conditions that promote hybridization under conditions of high stringency (e.g., increasing the temperature of the hybridization

and/or wash steps, the use of formamide in the hybridization solution, etc.) (see definition above for “stringency”).

[0064] As used herein, the term “probe” refers to an oligonucleotide (i.e., a sequence of nucleotides), whether occurring naturally as in a purified restriction digest or produced synthetically, recombinantly or by PCR amplification, that is capable of hybridizing to at least a portion of another oligonucleotide of interest. A probe may be single-stranded or double-stranded. Probes are useful in the detection, identification and isolation of particular gene sequences. It is contemplated that any probe used herein will be labeled with any “reporter molecule,” so that is detectable in any detection system, including, but not limited to enzyme (e.g., ELISA, as well as enzyme-based histochemical assays), fluorescent, radioactive, and luminescent systems. It is not intended that the described methods be limited to any particular detection system or label.

[0065] As used herein the term “portion” when in reference to a nucleotide sequence (as in “a portion of a given nucleotide sequence”) refers to fragments of that sequence. The fragments may range in size from four nucleotides to the entire nucleotide sequence minus one nucleotide (10 nucleotides, 20, 30, 40, 50, 100, 200, etc.).

[0066] “Amino acid sequence” and terms such as “polypeptide” or “protein” are not meant to limit the amino acid sequence to the complete, native amino acid sequence associated with the recited protein molecule.

[0067] The term “native protein” as used herein to indicate that a protein does not contain amino acid residues encoded by vector sequences; that is, the native protein contains only those amino acids found in the protein as it occurs in nature. A native protein may be produced by recombinant means or may be isolated from a naturally occurring source.

[0068] As used herein the term “portion” when in reference to a protein (as in “a portion of a given protein”) refers to fragments of that protein. The fragments may range in size from four amino acid residues to the entire amino acid sequence minus one amino acid.

[0069] The terms “overexpression” and “overexpressing” and grammatical equivalents, are used in reference to levels of mRNA to indicate a level of expression approximately 3-fold higher (or greater) than that observed in a given tissue in a control or non-transgenic animal. Levels of mRNA are measured using any of a number of techniques known to those skilled in the art including, but not limited to Northern blot analysis. Appropriate controls are included on the Northern blot to control for differences in the amount of RNA loaded from each tissue analyzed (e.g., the amount of 28S rRNA, an abundant RNA transcript present at essentially the same amount in all tissues, present in each sample can be used as a means of normalizing or standardizing the mRNA-specific signal observed on Northern blots). The amount of mRNA present in the band corresponding in size to the correctly spliced transgene RNA is quantified; other minor species of RNA which hybridize to the transgene probe are not considered in the quantification of the expression of the transgenic mRNA.

[0070] As used herein, the term “in vitro” refers to an artificial environment and to processes or reactions that occur within an artificial environment. In vitro environments can consist of, but are not limited to, test tubes and cell culture. The term “in vivo” refers to the natural environment (e.g., an animal or a cell) and to processes or reaction that occur within a natural environment.

[0071] The terms “test compound” and “candidate compound” refer to any chemical entity, pharmaceutical, drug, and the like that is a candidate for use to treat or prevent a disease, illness, sickness, or disorder of bodily function (e.g., cancer). Test compounds comprise both known and potential therapeutic compounds. A test compound can be determined to be therapeutic by screening using the screening methods described herein. In some embodiments, test compounds include antisense compounds.

[0072] As used herein, the term “sample” is used in its broadest sense. In one sense, it is meant to include a specimen or culture obtained from any source, as well as biological and environmental samples. Biological samples may be obtained from animals (including humans) and encompass fluids, solids, tissues, and gases. Biological samples include blood products, such as plasma, serum and the like. Environmental samples include environmental material such as surface matter, soil, water, crystals and industrial samples. Such examples are not however to be construed as limiting the sample types applicable to the described compositions and methods.

[0073] As used herein, the term “prostate sample” refers to any sample containing prostate cells or secretions. Example of prostate samples include, but are not limited to, a prostate tissue sample (e.g., a biopsy sample) or a urine sample.

[0074] As used herein, the terms “detect”, “detecting”, or “detection” may describe either the general act of discovering or discerning or the specific observation of a detectably labeled composition.

[0075] As used herein, the term “siRNAs” refers to small interfering RNAs. In some embodiments, siRNAs comprise a duplex, or double-stranded region, of about 18-25 nucleotides long; often siRNAs contain from about two to four unpaired nucleotides at the 3' end of each strand. At least one strand of the duplex or double-stranded region of a siRNA is substantially homologous to, or substantially complementary to, a target RNA molecule. The strand complementary to a target RNA molecule is the “antisense strand;” the strand homologous to the target RNA molecule is the “sense strand,” and is also complementary to the siRNA antisense strand. siRNAs may also contain additional sequences; non-limiting examples of such sequences include linking sequences, or loops, as well as stem and other folded structures. siRNAs appear to function as key intermediaries in triggering RNA interference in invertebrates and in vertebrates, and in triggering sequence-specific RNA degradation during posttranscriptional gene silencing in plants.

[0076] The term “RNA interference” or “RNAi” refers to the silencing or decreasing of gene expression by siRNAs. It is the process of sequence-specific, post-transcriptional gene silencing in animals and plants, initiated by siRNA that is homologous in its duplex region to the sequence of the silenced gene. The gene may be endogenous or exogenous to the organism, present integrated into a chromosome or present in a transfection vector that is not integrated into the genome. The expression of the gene is either completely or partially inhibited. RNAi may also be considered to inhibit the function of a target RNA; the function of the target RNA may be complete or partial.

DETAILED DESCRIPTION OF THE INVENTION

[0077] Compositions and methods for cancer research, diagnosis, and treatment, including but not limited to, cancer markers are described. In particular, SPINK1 and other mark-

ers for prostate cancer are provided. Accordingly, methods and kits for the detection of markers, as well as drug screening and therapeutic applications are provided.

I. Prostate Cancer Markers

[0078] Markers whose expression is specifically altered in cancerous prostate tissues are provided. Such markers find use in the diagnosis and characterization of prostate cancer. For example, experiments described herein identified SPINK1 as being overexpressed in prostate cancer. In addition, SPINK1, ERG and ETV1 showed mutually exclusive outlier expression. Further experiments identified a multiplex panel assay for detection of gene expression related to prostate cancer including SPINK1, PCA3, GOLPH2 and TMPRSS2:ERG fusions.

[0079] A. SPINK1

[0080] The cDNA sequence of SPINK1 (serine peptidase inhibitor, Kazal type 1) is provided in Genbank accession number NM_003122.2. The peptide encoded by SPINK1, also known as PSTI or TATI, was originally isolated from bovine pancreas and human pancreatic juice; its normal function is thought to be the inhibition of trypsin in the pancreas (Haverback et al., *Am J Med* 29, 421-33 (1960); Kazal et al., *Journal of the American Chemical Society* 70, 3034-3040 (1948); Paju et al., *Crit. Rev Clin Lab Sci* 43, 103-42 (2006); Greene et al., *Methods Enzymol* 45, 813-25 (1976)). SPINK1 mRNA and protein have been detected in a variety of benign and cancerous tissues, however its expression in prostate has not been described (reviewed in Paju and Stenman, *Crit. Rev Clin Lab Sci* 43, 103-42 (2006), Stenman, *Clin Chem* 48, 1206-9 (2002)). SPINK1 encodes a 79 amino acid peptide with a 23 amino acid signal peptide and is detectable in the urine and serum of healthy individuals (Paju and Stenman, supra). In addition to being strongly elevated during severe inflammation and pancreatitis, serum levels of SPINK1 may be dysregulated in numerous cancers, including pancreatic, gastric, liver, lung, breast, bladder, renal, head and neck, colorectal, kidney and ovarian cancer (reviewed in Paju and Stenman, supra, Stenman, supra). In conclusion, using a bioinformatics based approach, this study identified marked over-expression of SPINK1 in a subset of TMPRSS2:ETS negative prostate cancers, and confirmed these results using qPCR. Together, these results indicate that SPINK1 over-expression plays a role in prostate cancer development as well as serving as a biomarker for a molecular subtype of prostate cancer.

[0081] B. ERG

[0082] ERG (NM_004449), has been demonstrated to be highly expressed in prostate epithelium relative to other normal human tissues. The ERG gene is located on chromosome 21. The gene is located at 38,675,671-38,955,488 base pairs from the pter. The ERG gene is 279,817 total bp; minus strand orientation. The corresponding ERG cDNA and protein sequences are given at GenBank accession no. M17254 and GenBank accession no. NP04440 (Swiss Protein acc. no. P11308), respectively.

[0083] C. ETV1

[0084] The ETV1 gene is located on chromosome 7 (GenBank accession nos. NC_000007.11; NC_086703.11; and NT_007819.15). The gene is located at 13,708330-13,803,555 base pairs from the pter. The ETV1 gene is 95,225 bp total, minus strand orientation. The corresponding ETV1

cDNA and protein sequences are given at GenBank accession nos. NM_004956 and NP_004947 (Swiss protein acc. no. P50549), respectively.

[0085] D. PCA3

[0086] The gene for PCA3, also known as DD3 (Bussemakers, PCT Publication No. WO 98/45420, Schalken, *Eur. Urol.* 34 (suppl. 3):3-6 (1998), Bussemakers et al., *Cancer Res.* 59:5975-5979 (1999) and Bussemakers, *Eur. Urol.* 35:408-412 (1999)) is located on chromosome 9 and more precisely to region 9q21-22. It consists of four exons, which give rise, by both alternative splicing and alternative polyadenylation, to differently sized transcripts. By RT-PCR, PCA3^{dd3} expression was found to be limited to the prostate tissue and absent in all other tissues tested, including testis, seminal vesicle, ovary, placenta and bladder. In addition Northern blot analysis showed that PCA3^{dd3} is highly expressed in the vast majority of prostate cancer examined (47 out of 50) whereas no or very low expression is detected in BPH or normal prostate cells from the same patients. There is at least a 20-fold over-expression of PCA3^{dd3} in prostate carcinoma in comparison to normal or BPH tissues. PCA3^{dd3} expression seems to increase with tumor grade and is detected in metastatic lesions.

[0087] PCA3 is a gene wherein significant alternative splicing (as well as alternative polyadenylation) occurs, as evidenced by the differently sized transcripts observed on Northern blots and the different types of clones identified. Virtually every combination of exons is possible.

[0088] From the 80 analyzed cDNA clones, at least four different transcripts were shown to be present due to alternative splicing or alternative polyadenylation (See e.g., U.S. Pat. No. 7,008,765, PCT Publication No. WO05/003387A2, U.S. Pat. No. 7,138,235 and de Kok et al., *Cancer Res* 2002; 62:2695-8; each of which is herein incorporated by reference). Sequence analysis of the genomic clones as compared to the cDNA clones revealed the genomic structure of the PCA3 gene. Three introns and 4 exons are present. The first intron is approximately 20 kb in length.

[0089] The first cDNA species is found in approximately 5% of the cDNA clones and contains exons 1, 2, 3, 4a and 4b (poly-adenylation after 4b is preceded by a real consensus poly-A-addition signal)

[0090] The second cDNA species, found in approximately 15% of the cDNA cloned, contains exons 1, 3, 4a, 4b and 4c, arises by alternative splicing of the second exon (not present in this cDNA) and terminates at a different (real consensus) poly-A-addition signal.

[0091] The third cDNA species contains exons 1, 3, 4a, and 4b and is the most common one found (approximately 65% of 80 clones) (FIG. 1). This cDNA is most likely responsible for the most prominent transcript seen by Northern blot analysis.

[0092] The fourth cDNA species detected contains exons 1, 3, and 4a representing about 15% of clones, and terminates after 4a, which is the original DD3 clone stop site. The poly-A-addition signal present here is close to the consensus sequence.

[0093] E. GOLPH2

[0094] GOLPH2 (GP73) is a glycoprotein marker of hepatitis B associated with liver cancer (Block et al., (2005) *Proc. Natl. Acad. Sci. USA*, 102, 779-784). GP73 is a type II Golgi transmembrane protein that is expressed at high level in the hepatocytes of patients with viral hepatitis (Kladney, et al., 2000, *Gene* 249, 53-65). GP73 is constitutively expressed in biliary epithelial cells, and minimally expressed in normal

hepatocytes. In contrast, livers of patients with giant-cell hepatitis display strong immunoreactivity to GP73 in multinucleated hepatocytes. GP73 mRNA and protein are expressed in highly differentiated HepG2 hepatoma cells after infection with viruses, including adenoviruses. Because GP73 is a Golgi transmembrane protein, it is not expected to exist in significant amounts in the serum, even in subjects with damaged or diseased livers.

[0095] Significant increases in whole-organ levels of GP73 have been found in liver disease due to viral causes (HBV, HCV) or nonviral causes (alcohol-induced liver disease, autoimmune hepatitis) (Kladney, et al., 2002, *Hepatology* 35 (6):1431-40). Hepatocyte expression of GP73 is unregulated in diseased livers, regardless of etiology, whereas biliary epithelial cell expression does not change appreciably. Gp73 has been used as a marker for hepatocellular cancer (US 20050112711, herein incorporated by reference in its entirety).

[0096] F. TMRSS2:ERG

[0097] TMRSS2:ERG and other gene fusion markers of prostate cancer are described in US Publication No. US 20070212702 A1, herein incorporated by reference in its entirety. TMRSS2 (NM_005656), has been demonstrated to be highly expressed in prostate epithelium relative to other normal human tissues (Lin et al., *Cancer Research* 59: 4180 (1999)). The TMRSS2 gene is located on chromosome 21. This gene is located at 41,750,797-41,801,948 by from the pter (51,151 total bp; minus strand orientation). The human TMRSS2 protein sequence may be found at GenBank accession no. AAC51784 (Swiss Protein accession no. O15393) and the corresponding cDNA at GenBank accession no. U75329 (see also, Paoloni-Giacobino, et al., *Genomics* 44: 309 (1997)).

II. Antibodies

[0098] SPINK1 or other cancer marker polypeptides, which include fragments, derivatives and analogs thereof may be used as immunogens to produce antibodies useful for diagnostic and therapeutic applications. Such antibodies may be polyclonal or monoclonal, chimeric, humanized, single chain or Fab fragments, which may be labeled or unlabeled, all of which may be produced by using well known procedures and standard laboratory practices. See, e.g., Burns, ed., *Immunochemical Protocols*, 3rd ed., Humana Press (2005); Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory (1988); Kozbor et al., *Immunology Today* 4: 72 (1983); Köhler and Milstein, *Nature* 256: 495 (1975).

[0099] In some embodiments, compositions and methods utilize commercially available antibodies. Antibodies to SPINK1 are available, for example, from GeneTex, San Antonio, Tex.; Novus Biologicals, Littleton, Colo.; Santa Cruz Biotechnology, Santa Cruz, Calif.; and Abnova Corporation, Taipei City, Taiwan. Antibodies to SLC22A3 are available, for example, from Abnova Corporation, Taipei City, Taiwan.

III. Diagnostic Applications

[0100] In some embodiments, methods for detection of expression of cancer markers (e.g., SPINK1) are provided. In some embodiments, detection methods measure the level of cancer marker detected (e.g., in comparison to the level of the cancer marker in normal prostate tissue).

[0101] In some embodiments, expression is measured directly (e.g., at the RNA or protein level). In some embodiments, expression is detected in tissue samples (e.g., biopsy tissue). In other embodiments, expression is detected in bodily fluids (e.g., including but not limited to, plasma, serum, whole blood, mucus, and urine). In certain embodiments, the presence of a cancer marker is used to provide a prognosis to a subject. For example, the detection of overexpression of SPINK1 is indicative of gene expression associated with prostate cancer, outlier expression of SPINK1 is associated with prostate cancer recurrence in a subject. The information provided is also used to select a treatment course of action. For example, if a subject is found to have a marker indicative of a highly metastasizing tumor, additional therapies (e.g., hormonal or radiation therapies) can be started at an earlier point when they are more likely to be effective (e.g., before metastasis). In addition, if a subject is found to have a tumor that is not responsive to hormonal therapy, the expense and inconvenience of such therapies can be avoided.

[0102] The cancer markers (e.g., SPINK1) described herein may be detected along with other markers in a multiplex or panel format. Markers are selected for their predictive value alone or in combination with SPINK1. In one exemplary embodiment, two or more of SPINK1, PCA3, GOLPH2 and TMRSS2:ERG are detected in a multiplex assay. Other exemplary prostate cancer markers include, but are not limited to: AMACR/P504S (U.S. Pat. No. 6,262,245); PCA3 (U.S. Pat. No. 7,008,765); PCGEM1 (U.S. Pat. No. 6,828,429); prostein/P501S, P503S, P504S, P509S, P510S, prostate/P703P, P710P (U.S. Publication No. 20030185830); and, those disclosed in U.S. Pat. Nos. 5,854,206 and 6,034,218, and U.S. Publication No. 20030175736, each of which is herein incorporated by reference in its entirety. Markers for other cancers, diseases, infections, and metabolic conditions are also contemplated for inclusion in a multiplex of panel format.

[0103] The diagnostic methods described herein may also be modified with reference to data correlating particular cancer marker(s) with the stage, aggressiveness or progression of the disease or the presence or risk of metastasis. Ultimately, the information provided by the methods described herein will assist a physician in choosing the best course of treatment for a particular patient.

[0104] A. Sample

[0105] Any patient sample suspected of containing the cancer marker may be tested according to the methods described herein. By way of non-limiting examples, the sample may be tissue (e.g., a prostate biopsy sample or a tissue sample obtained by prostatectomy), blood, urine, semen, prostatic secretions or a fraction thereof (e.g., plasma, serum, urine supernatant, urine cell pellet or prostate cells). A urine sample is preferably collected immediately following an attentive digital rectal examination (DRE), which causes prostate cells from the prostate gland to shed into the urinary tract.

[0106] The patient sample typically requires preliminary processing designed to isolate or enrich the sample for the gene fusions or cells that contain the gene fusions. A variety of techniques known to those of ordinary skill in the art may be used for this purpose, including but not limited: centrifugation; immunocapture; cell lysis; and, nucleic acid target capture (See, e.g., EP Pat. No. 1 409 727, herein incorporated by reference in its entirety).

[0107] B. DNA and RNA Detection

[0108] The cancer markers described herein may be detected as RNA using a variety of nucleic acid techniques known to those of ordinary skill in the art, including but not limited to: nucleic acid sequencing; nucleic acid hybridization; and, nucleic acid amplification.

[0109] 1. Sequencing

[0110] Illustrative non-limiting examples of nucleic acid sequencing techniques include, but are not limited to, chain terminator (Sanger) sequencing and dye terminator sequencing. Those of ordinary skill in the art will recognize that because RNA is less stable in the cell and more prone to nuclease attack experimentally RNA is usually reverse transcribed to DNA before sequencing.

[0111] Chain terminator sequencing uses sequence-specific termination of a DNA synthesis reaction using modified nucleotide substrates. Extension is initiated at a specific site on the template DNA by using a short radioactive, or other labeled, oligonucleotide primer complementary to the template at that region. The oligonucleotide primer is extended using a DNA polymerase, standard four deoxynucleotide bases, and a low concentration of one chain terminating nucleotide, most commonly a di-deoxynucleotide. This reaction is repeated in four separate tubes with each of the bases taking turns as the di-deoxynucleotide. Limited incorporation of the chain terminating nucleotide by the DNA polymerase results in a series of related DNA fragments that are terminated only at positions where that particular di-deoxynucleotide is used. For each reaction tube, the fragments are size-separated by electrophoresis in a slab polyacrylamide gel or a capillary tube filled with a viscous polymer. The sequence is determined by reading which lane produces a visualized mark from the labeled primer as you scan from the top of the gel to the bottom.

[0112] Dye terminator sequencing alternatively labels the terminators. Complete sequencing can be performed in a single reaction by labeling each of the di-deoxynucleotide chain-terminators with a separate fluorescent dye, which fluoresces at a different wavelength.

[0113] 2. Hybridization

[0114] Illustrative non-limiting examples of nucleic acid hybridization techniques include, but are not limited to, in situ hybridization (ISH), microarray, and Southern or Northern blot.

[0115] In situ hybridization (ISH) is a type of hybridization that uses a labeled complementary DNA or RNA strand as a probe to localize a specific DNA or RNA sequence in a portion or section of tissue (in situ), or, if the tissue is small enough, the entire tissue (whole mount ISH). DNA ISH can be used to determine the structure of chromosomes. RNA ISH is used to measure and localize mRNAs and other transcripts within tissue sections or whole mounts. Sample cells and tissues are usually treated to fix the target transcripts in place and to increase access of the probe. The probe hybridizes to the target sequence at elevated temperature, and then the excess probe is washed away. The probe that was labeled with either radio-, fluorescent- or antigen-labeled bases is localized and quantitated in the tissue using either autoradiography, fluorescence microscopy or immunohistochemistry, respectively. ISH can also use two or more probes, labeled with radioactivity or the other non-radioactive labels, to simultaneously detect two or more transcripts.

[0116] 2.1 FISH

[0117] In some embodiments, cancer marker sequences are detected using fluorescence in situ hybridization (FISH). The preferred FISH assays utilize bacterial artificial chromosomes (BACs). These have been used extensively in the human genome sequencing project (see *Nature* 409: 953-958 (2001)) and clones containing specific BACs are available through distributors that can be located through many sources, e.g., NCBI. Each BAC clone from the human genome has been given a reference name that unambiguously identifies it. These names can be used to find a corresponding GenBank sequence and to order copies of the clone from a distributor.

[0118] A method of performing a FISH assay on human prostate cells, human prostate tissue or on the fluid surrounding said human prostate cells or human prostate tissue is further provided.

[0119] Specific protocols are well known in the art and can be readily adapted for use. Guidance regarding methodology may be obtained from many references including: *In situ Hybridization: Medical Applications* (eds. G. R. Coulton and J. de Belleruche), Kluwer Academic Publishers, Boston (1992); *In situ Hybridization: In Neurobiology; Advances in Methodology* (eds. J. H. Eberwine, K. L. Valentino, and J. D. Barchas), Oxford University Press Inc., England (1994); *In situ Hybridization: A Practical Approach* (ed. D. G. Wilkinson), Oxford University Press Inc., England (1992); Kuo, et al., *Am. J. Hum. Genet.* 49:112-119 (1991); Klinger, et al., *Am. J. Hum. Genet.* 51:55-65 (1992); and Ward, et al., *Am. J. Hum. Genet.* 52:854-865 (1993)). There are also kits that are commercially available and that provide protocols for performing FISH assays (available from e.g., Oncor, Inc., Gaithersburg, Md.). Patents providing guidance on methodology include U.S. Pat. Nos. 5,225,326; 5,545,524; 6,121,489 and 6,573,043. All of these references are hereby incorporated by reference in their entirety and may be used along with similar references in the art and with the information provided in the Examples section herein to establish procedural steps convenient for a particular laboratory.

[0120] 2.2 Microarrays

[0121] Different kinds of biological assays are called microarrays including, but not limited to: DNA microarrays (e.g., cDNA microarrays and oligonucleotide microarrays); protein microarrays; tissue microarrays; transfection or cell microarrays; chemical compound microarrays; and, antibody microarrays. A DNA microarray, commonly known as gene chip, DNA chip, or biochip, is a collection of microscopic DNA spots attached to a solid surface (e.g., glass, plastic or silicon chip) forming an array for the purpose of expression profiling or monitoring expression levels for thousands of genes simultaneously. The affixed DNA segments are known as probes, thousands of which can be used in a single DNA microarray. Microarrays can be used to identify disease genes by comparing gene expression in disease and normal cells. Microarrays can be fabricated using a variety of technologies, including but not limiting: printing with fine-pointed pins onto glass slides; photolithography using pre-made masks; photolithography using dynamic micromirror devices; ink jet printing; or, electrochemistry on microelectrode arrays.

[0122] Southern and Northern blotting is used to detect specific DNA or RNA sequences, respectively. DNA or RNA extracted from a sample is fragmented, electrophoretically separated on a matrix gel, and transferred to a membrane filter. The filter bound DNA or RNA is subject to hybridiza-

tion with a labeled probe complementary to the sequence of interest. Hybridized probe bound to the filter is detected. A variant of the procedure is the reverse Northern blot, in which the substrate nucleic acid that is affixed to the membrane is a collection of isolated DNA fragments and the probe is RNA extracted from a tissue and labeled.

[0123] 3. Amplification

[0124] Cancer marker genomic DNA and mRNA may be amplified prior to or simultaneous with detection. Illustrative non-limiting examples of nucleic acid amplification techniques include, but are not limited to, polymerase chain reaction (PCR), reverse transcription polymerase chain reaction (RT-PCR), transcription-mediated amplification (TMA), ligase chain reaction (LCR), strand displacement amplification (SDA), and nucleic acid sequence based amplification (NASBA). Those of ordinary skill in the art will recognize that certain amplification techniques (e.g., PCR) require that RNA be reverse transcribed to DNA prior to amplification (e.g., RT-PCR), whereas other amplification techniques directly amplify RNA (e.g., TMA and NASBA).

[0125] The polymerase chain reaction (U.S. Pat. Nos. 4,683,195, 4,683,202, 4,800,159 and 4,965,188, each of which is herein incorporated by reference in its entirety), commonly referred to as PCR, uses multiple cycles of denaturation, annealing of primer pairs to opposite strands, and primer extension to exponentially increase copy numbers of a target nucleic acid sequence. In a variation called RT-PCR, reverse transcriptase (RT) is used to make a complementary DNA (cDNA) from mRNA, and the cDNA is then amplified by PCR to produce multiple copies of DNA. For other various permutations of PCR see, e.g., U.S. Pat. Nos. 4,683,195, 4,683,202 and 4,800,159; Mullis et al., *Meth. Enzymol.* 155: 335 (1987); and, Murakawa et al., *DNA* 7: 287 (1988), each of which is herein incorporated by reference in its entirety.

[0126] Transcription mediated amplification (U.S. Pat. Nos. 5,480,784 and 5,399,491, each of which is herein incorporated by reference in its entirety), commonly referred to as TMA, synthesizes multiple copies of a target nucleic acid sequence autocatalytically under conditions of substantially constant temperature, ionic strength, and pH in which multiple RNA copies of the target sequence autocatalytically generate additional copies. See, e.g., U.S. Pat. Nos. 5,399,491 and 5,824,518, each of which is herein incorporated by reference in its entirety. In a variation described in U.S. Publ. No. 20060046265 (herein incorporated by reference in its entirety), TMA optionally incorporates the use of blocking moieties, terminating moieties, and other modifying moieties to improve TMA process sensitivity and accuracy.

[0127] The ligase chain reaction (Weiss, R., *Science* 254: 1292 (1991), herein incorporated by reference in its entirety), commonly referred to as LCR, uses two sets of complementary DNA oligonucleotides that hybridize to adjacent regions of the target nucleic acid. The DNA oligonucleotides are covalently linked by a DNA ligase in repeated cycles of thermal denaturation, hybridization and ligation to produce a detectable double-stranded ligated oligonucleotide product.

[0128] Strand displacement amplification (Walker, G. et al., *Proc. Natl. Acad. Sci. USA* 89: 392-396 (1992); U.S. Pat. Nos. 5,270,184 and 5,455,166, each of which is herein incorporated by reference in its entirety), commonly referred to as SDA, uses cycles of annealing pairs of primer sequences to opposite strands of a target sequence, primer extension in the presence of a dNTP α S to produce a duplex hemiphosphorothioated primer extension product, endonuclease-mediated

nicking of a hemimodified restriction endonuclease recognition site, and polymerase-mediated primer extension from the 3' end of the nick to displace an existing strand and produce a strand for the next round of primer annealing, nicking and strand displacement, resulting in geometric amplification of product. Thermophilic SDA (tSDA) uses thermophilic endonucleases and polymerases at higher temperatures in essentially the same method (EP Pat. No. 0 684 315).

[0129] Other amplification methods include, for example: nucleic acid sequence based amplification (U.S. Pat. No. 5,130,238, herein incorporated by reference in its entirety), commonly referred to as NASBA; one that uses an RNA replicase to amplify the probe molecule itself (Lizardi et al., *BioTechnol.* 6: 1197 (1988), herein incorporated by reference in its entirety), commonly referred to as Q β replicase; a transcription based amplification method (Kwoh et al., *Proc. Natl. Acad. Sci. USA* 86:1173 (1989)); and, self-sustained sequence replication (Guatelli et al., *Proc. Natl. Acad. Sci. USA* 87: 1874 (1990), each of which is herein incorporated by reference in its entirety). For further discussion of known amplification methods see Persing, David H., "In Vitro Nucleic Acid Amplification Techniques" in *Diagnostic Medical Microbiology: Principles and Applications* (Persing et al., Eds.), pp. 51-87 (American Society for Microbiology, Washington, D.C. (1993)).

[0130] 4. Detection Methods

[0131] Non-amplified or amplified nucleic acids can be detected by any conventional means. For example, the cancer marker nucleic acid can be detected by hybridization with a detectably labeled probe and measurement of the resulting hybrids. Illustrative non-limiting examples of detection methods are described below.

[0132] One illustrative detection method, the Hybridization Protection Assay (HPA) involves hybridizing a chemiluminescent oligonucleotide probe (e.g., an acridinium ester-labeled (AE) probe) to the target sequence, selectively hydrolyzing the chemiluminescent label present on unhybridized probe, and measuring the chemiluminescence produced from the remaining probe in a luminometer. See, e.g., U.S. Pat. No. 5,283,174 and Norman C. Nelson et al., *Nonisotopic Probing, Blotting, and Sequencing*, ch. 17 (Larry J. Kricka ed., 2d ed. 1995, each of which is herein incorporated by reference in its entirety).

[0133] Another illustrative detection method provides for quantitative evaluation of the amplification process in real-time. Evaluation of an amplification process in "real-time" involves determining the amount of amplicon in the reaction mixture either continuously or periodically during the amplification reaction, and using the determined values to calculate the amount of target sequence initially present in the sample. A variety of methods for determining the amount of initial target sequence present in a sample based on real-time amplification are well known in the art. These include methods disclosed in U.S. Pat. Nos. 6,303,305 and 6,541,205, each of which is herein incorporated by reference in its entirety. Another method for determining the quantity of target sequence initially present in a sample, but which is not based on a real-time amplification, is disclosed in U.S. Pat. No. 5,710,029, herein incorporated by reference in its entirety.

[0134] Amplification products may be detected in real-time through the use of various self-hybridizing probes, most of which have a stem-loop structure. Such self-hybridizing probes are labeled so that they emit differently detectable signals, depending on whether the probes are in a self-hybrid-

ized state or an altered state through hybridization to a target sequence. By way of non-limiting example, “molecular torches” are a type of self-hybridizing probe that includes distinct regions of self-complementarity (referred to as “the target binding domain” and “the target closing domain”) which are connected by a joining region (e.g., non-nucleotide linker) and which hybridize to each other under predetermined hybridization assay conditions. In a preferred embodiment, molecular torches contain single-stranded base regions in the target binding domain that are from 1 to about 20 bases in length and are accessible for hybridization to a target sequence present in an amplification reaction under strand displacement conditions. Under strand displacement conditions, hybridization of the two complementary regions, which may be fully or partially complementary, of the molecular torch is favored, except in the presence of the target sequence, which will bind to the single-stranded region present in the target binding domain and displace all or a portion of the target closing domain. The target binding domain and the target closing domain of a molecular torch include a detectable label or a pair of interacting labels (e.g., luminescent/quencher) positioned so that a different signal is produced when the molecular torch is self-hybridized than when the molecular torch is hybridized to the target sequence, thereby permitting detection of probe:target duplexes in a test sample in the presence of unhybridized molecular torches. Molecular torches and a variety of types of interacting label pairs are disclosed in U.S. Pat. No. 6,534,274, herein incorporated by reference in its entirety.

[0135] Another example of a detection probe having self-complementarity is a “molecular beacon.” Molecular beacons include nucleic acid molecules having a target complementary sequence, an affinity pair (or nucleic acid arms) holding the probe in a closed conformation in the absence of a target sequence present in an amplification reaction, and a label pair that interacts when the probe is in a closed conformation. Hybridization of the target sequence and the target complementary sequence separates the members of the affinity pair, thereby shifting the probe to an open conformation. The shift to the open conformation is detectable due to reduced interaction of the label pair, which may be, for example, a fluorophore and a quencher (e.g., DABCYL and EDANS). Molecular beacons are disclosed in U.S. Pat. Nos. 5,925,517 and 6,150,097, herein incorporated by reference in its entirety.

[0136] Other self-hybridizing probes are well known to those of ordinary skill in the art. By way of non-limiting example, probe binding pairs having interacting labels, such as those disclosed in U.S. Pat. No. 5,928,862 (herein incorporated by reference in its entirety) might be adapted for use. Probe systems used to detect single nucleotide polymorphisms (SNPs) might also be utilized in. Additional detection systems include “molecular switches,” as disclosed in U.S. Publ. No. 20050042638, herein incorporated by reference in its entirety. Other probes, such as those comprising intercalating dyes and/or fluorochromes, are also useful for detection of amplification products. See, e.g., U.S. Pat. No. 5,814,447 (herein incorporated by reference in its entirety).

[0137] C. Protein Detection

[0138] The cancer markers may be detected as proteins using a variety of protein techniques known to those of ordinary skill in the art, including but not limited to: protein sequencing; and, immunoassays.

[0139] 1. Sequencing

[0140] Illustrative non-limiting examples of protein sequencing techniques include, but are not limited to, mass spectrometry and Edman degradation.

[0141] Mass spectrometry can, in principle, sequence any size protein but becomes computationally more difficult as size increases. A protein is digested by an endoprotease, and the resulting solution is passed through a high pressure liquid chromatography column. At the end of this column, the solution is sprayed out of a narrow nozzle charged to a high positive potential into the mass spectrometer. The charge on the droplets causes them to fragment until only single ions remain. The peptides are then fragmented and the mass-charge ratios of the fragments measured. The mass spectrum is analyzed by computer and often compared against a database of previously sequenced proteins in order to determine the sequences of the fragments. The process is then repeated with a different digestion enzyme, and the overlaps in sequences are used to construct a sequence for the protein.

[0142] In the Edman degradation reaction, the peptide to be sequenced is adsorbed onto a solid surface (e.g., a glass fiber coated with polybrene). The Edman reagent, phenylisothiocyanate (PTC), is added to the adsorbed peptide, together with a mildly basic buffer solution of 12% trimethylamine, and reacts with the amine group of the N-terminal amino acid. The terminal amino acid derivative can then be selectively detached by the addition of anhydrous acid. The derivative isomerizes to give a substituted phenylthiohydantoin, which can be washed off and identified by chromatography, and the cycle can be repeated. The efficiency of each step is about 98%, which allows about 50 amino acids to be reliably determined.

[0143] 2. Immunoassays

[0144] Illustrative non-limiting examples of immunoassays include, but are not limited to: immunoprecipitation; Western blot; ELISA; immunohistochemistry; immunocytochemistry; flow cytometry; and, immuno-PCR. Polyclonal or monoclonal antibodies detectably labeled using various techniques known to those of ordinary skill in the art (e.g., colorimetric, fluorescent, chemiluminescent or radioactive) are suitable for use in the immunoassays.

[0145] Immunoprecipitation is the technique of precipitating an antigen out of solution using an antibody specific to that antigen. The process can be used to identify protein complexes present in cell extracts by targeting a protein believed to be in the complex. The complexes are brought out of solution by insoluble antibody-binding proteins isolated initially from bacteria, such as Protein A and Protein G. The antibodies can also be coupled to sepharose beads that can easily be isolated out of solution. After washing, the precipitate can be analyzed using mass spectrometry, Western blotting, or any number of other methods for identifying constituents in the complex.

[0146] A Western blot, or immunoblot, is a method to detect protein in a given sample of tissue homogenate or extract. It uses gel electrophoresis to separate denatured proteins by mass. The proteins are then transferred out of the gel and onto a membrane, typically polyvinylidene difluoride or nitrocellulose, where they are probed using antibodies specific to the protein of interest. As a result, researchers can examine the amount of protein in a given sample and compare levels between several groups.

[0147] An ELISA, short for Enzyme-Linked Immunosorbent Assay, is a biochemical technique to detect the presence

of an antibody or an antigen in a sample. It utilizes a minimum of two antibodies, one of which is specific to the antigen and the other of which is coupled to an enzyme. The second antibody will cause a chromogenic or fluorogenic substrate to produce a signal. Variations of ELISA include sandwich ELISA, competitive ELISA, and ELISPOT. Because the ELISA can be performed to evaluate either the presence of antigen or the presence of antibody in a sample, it is a useful tool both for determining serum antibody concentrations and also for detecting the presence of antigen.

[0148] Immunohistochemistry and immunocytochemistry refer to the process of localizing proteins in a tissue section or cell, respectively, via the principle of antigens in tissue or cells binding to their respective antibodies. Visualization is enabled by tagging the antibody with color producing or fluorescent tags. Typical examples of color tags include, but are not limited to, horseradish peroxidase and alkaline phosphatase. Typical examples of fluorophore tags include, but are not limited to, fluorescein isothiocyanate (FITC) or phycoerythrin (PE).

[0149] Flow cytometry is a technique for counting, examining and sorting microscopic particles suspended in a stream of fluid. It allows simultaneous multiparametric analysis of the physical and/or chemical characteristics of single cells flowing through an optical/electronic detection apparatus. A beam of light (e.g., a laser) of a single frequency or color is directed onto a hydrodynamically focused stream of fluid. A number of detectors are aimed at the point where the stream passes through the light beam; one in line with the light beam (Forward Scatter or FSC) and several perpendicular to it (Side Scatter (SSC) and one or more fluorescent detectors). Each suspended particle passing through the beam scatters the light in some way, and fluorescent chemicals in the particle may be excited into emitting light at a lower frequency than the light source. The combination of scattered and fluorescent light is picked up by the detectors, and by analyzing fluctuations in brightness at each detector, one for each fluorescent emission peak, it is possible to deduce various facts about the physical and chemical structure of each individual particle. FSC correlates with the cell volume and SSC correlates with the density or inner complexity of the particle (e.g., shape of the nucleus, the amount and type of cytoplasmic granules or the membrane roughness).

[0150] Immuno-polymerase chain reaction (IPCR) utilizes nucleic acid amplification techniques to increase signal generation in antibody-based immunoassays. Because no protein equivalence of PCR exists, that is, proteins cannot be replicated in the same manner that nucleic acid is replicated during PCR, the only way to increase detection sensitivity is by signal amplification. The target proteins are bound to antibodies which are directly or indirectly conjugated to oligonucleotides. Unbound antibodies are washed away and the remaining bound antibodies have their oligonucleotides amplified. Protein detection occurs via detection of amplified oligonucleotides using standard nucleic acid detection methods, including real-time methods.

[0151] D. Data Analysis

[0152] In some embodiments, a computer-based analysis program is used to translate the raw data generated by the detection assay (e.g., the presence, absence, or amount of a given marker or markers) into data of predictive value for a clinician. The clinician can access the predictive data using any suitable means. Thus, in some preferred embodiments, the clinician, who is not likely to be trained in genetics or

molecular biology, need not understand the raw data. The data is presented directly to the clinician in its most useful form. The clinician is then able to immediately utilize the information in order to optimize the care of the subject.

[0153] Any method capable of receiving, processing, and transmitting the information to and from laboratories conducting the assays, information provides, medical personal, and subjects may be utilized. For example, in some embodiments, a sample (e.g., a biopsy or a serum or urine sample) is obtained from a subject and submitted to a profiling service (e.g., clinical lab at a medical facility, genomic profiling business, etc.), located in any part of the world (e.g., in a country different than the country where the subject resides or where the information is ultimately used) to generate raw data. Where the sample comprises a tissue or other biological sample, the subject may visit a medical center to have the sample obtained and sent to the profiling center, or subjects may collect the sample themselves (e.g., a urine sample) and directly send it to a profiling center. Where the sample comprises previously determined biological information, the information may be directly sent to the profiling service by the subject (e.g., an information card containing the information may be scanned by a computer and the data transmitted to a computer of the profiling center using an electronic communication systems). Once received by the profiling service, the sample is processed and a profile is produced (i.e., expression data), specific for the diagnostic or prognostic information desired for the subject.

[0154] The profile data is then prepared in a format suitable for interpretation by a treating clinician. For example, rather than providing raw expression data, the prepared format may represent a diagnosis or risk assessment (e.g., likelihood of cancer being present or the subtype of cancer) for the subject, along with recommendations for particular treatment options. The data may be displayed to the clinician by any suitable method. For example, in some embodiments, the profiling service generates a report that can be printed for the clinician (e.g., at the point of care) or displayed to the clinician on a computer monitor.

[0155] In some embodiments, the information is first analyzed at the point of care or at a regional facility. The raw data is then sent to a central processing facility for further analysis and/or to convert the raw data to information useful for a clinician or patient. The central processing facility provides the advantage of privacy (all data is stored in a central facility with uniform security protocols), speed, and uniformity of data analysis. The central processing facility can then control the fate of the data following treatment of the subject. For example, using an electronic communication system, the central facility can provide data to the clinician, the subject, or researchers.

[0156] In some embodiments, the subject is able to directly access the data using the electronic communication system. The subject may chose further intervention or counseling based on the results. In some embodiments, the data is used for research use. For example, the data may be used to further optimize the inclusion or elimination of markers as useful indicators of a particular condition or stage of disease.

[0157] In some embodiments, the results are used in a clinical setting to determine a further diagnostic (e.g., the additional of further screening (e.g., PSA or other markers) or diagnostic (e.g., biopsy) course of action. In other embodiments, the results are used to determine a treatment course of action (e.g., choice of therapies or watchful waiting).

[0158] E. In Vivo Imaging

[0159] In some embodiments, in vivo imaging techniques are used to visualize the expression of cancer markers in an animal (e.g., a human or non-human mammal). For example, in some embodiments, cancer marker mRNA or protein is labeled using a labeled antibody specific for the cancer marker. A specifically bound and labeled antibody can be detected in an individual using an in vivo imaging method, including, but not limited to, radionuclide imaging, positron emission tomography, computerized axial tomography, X-ray or magnetic resonance imaging method, fluorescence detection, and chemiluminescent detection. Methods for generating antibodies to the cancer markers of are described above.

[0160] In vivo imaging methods are useful in the diagnosis of cancers that express the cancer markers described herein (e.g., SPINK1). In vivo imaging is used to visualize the presence of a marker indicative of the cancer. Such techniques allow for diagnosis without the use of an unpleasant biopsy. The in vivo imaging methods are also useful for providing prognoses to cancer patients. For example, the presence of a marker indicative of cancers likely to metastasize can be detected. The in vivo imaging methods can further be used to detect metastatic cancers in other parts of the body.

[0161] In some embodiments, reagents (e.g., antibodies) specific for the cancer markers are fluorescently labeled. The labeled antibodies are introduced into a subject (e.g., orally or parenterally). Fluorescently labeled antibodies are detected using any suitable method (e.g., using the apparatus described in U.S. Pat. No. 6,198,107, herein incorporated by reference).

[0162] In other embodiments, antibodies are radioactively labeled. The use of antibodies for in vivo diagnosis is well known in the art. Sumerdon et al., (Nucl. Med. Biol 17:247-254 [1990] have described an optimized antibody-chelator for the radioimmunosintographic imaging of tumors using Indium-111 as the label. Griffin et al., (J Clin Onc 9:631-640 [1991]) have described the use of this agent in detecting tumors in patients suspected of having recurrent colorectal cancer. The use of similar agents with paramagnetic ions as labels for magnetic resonance imaging is known in the art (Lauffer, Magnetic Resonance in Medicine 22:339-342 [1991]). The label used will depend on the imaging modality chosen. Radioactive labels such as Indium-111, Technetium-99m, or Iodine-131 can be used for planar scans or single photon emission computed tomography (SPECT). Positron emitting labels such as Fluorine-19 can also be used for positron emission tomography (PET). For MRI, paramagnetic ions such as Gadolinium (III) or Manganese (II) can be used.

[0163] Radioactive metals with half-lives ranging from 1 hour to 3.5 days are available for conjugation to antibodies, such as scandium-47 (3.5 days) gallium-67 (2.8 days), gallium-68 (68 minutes), technetium-99m (6 hours), and indium-111 (3.2 days), of which gallium-67, technetium-99m, and indium-111 are preferable for gamma camera imaging, gallium-68 is preferable for positron emission tomography.

[0164] A useful method of labeling antibodies with such radiometals is by means of a bifunctional chelating agent, such as diethylenetriaminepentaacetic acid (DTPA), as described, for example, by Khaw et al. (Science 209:295 [1980]) for In-111 and Tc-99m, and by Scheinberg et al. (Science 215:1511 [1982]). Other chelating agents may also be used, but the 1-(p-carboxymethoxybenzyl)EDTA and the

carboxycarbonic anhydride of DTPA are advantageous because their use permits conjugation without affecting the antibody's immunoreactivity substantially.

[0165] Another method for coupling DTPA to proteins is by use of the cyclic anhydride of DTPA, as described by Hnatowich et al. (Int. J. Appl. Radiat. Isot. 33:327 [1982]) for labeling of albumin with In-111, but which can be adapted for labeling of antibodies. A suitable method of labeling antibodies with Tc-99m which does not use chelation with DTPA is the pretinning method of Crockford et al., (U.S. Pat. No. 4,323,546, herein incorporated by reference).

[0166] A preferred method of labeling immunoglobulins with Tc-99m is that described by Wong et al. (Int. J. Appl. Radiat. Isot., 29:251 [1978]) for plasma protein, and recently applied successfully by Wong et al. (J. Nucl. Med., 23:229 [1981]) for labeling antibodies.

[0167] In the case of the radiometals conjugated to the specific antibody, it is likewise desirable to introduce as high a proportion of the radiolabel as possible into the antibody molecule without destroying its immunospecificity.

[0168] A further improvement may be achieved by effecting radiolabeling in the presence of the specific cancer marker, to insure that the antigen binding site on the antibody will be protected. The antigen is separated after labeling.

[0169] In still further embodiments, in vivo biophotonic imaging (Xenogen, Alameda, Calif.) is utilized for in vivo imaging. This real-time in vivo imaging utilizes luciferase. The luciferase gene is incorporated into cells, microorganisms, and animals (e.g., as a fusion protein with a cancer marker). When active, it leads to a reaction that emits light. A CCD camera and software is used to capture the image and analyze it.

[0170] F. Compositions and Kits

[0171] In yet other embodiments, kits for the detection and characterization of prostate cancer are provided. In some embodiments, the kits contain antibodies specific for a cancer marker (e.g., SPINK1), in addition to detection reagents and buffers. In other embodiments, the kits contain reagents specific for the detection of the level of mRNA or cDNA or the presence or absence of chromosomal deletions (e.g., oligonucleotide probes or primers). In preferred embodiments, the kits contain all of the components necessary or sufficient to perform a detection assay, including all controls, directions for performing assays, and any necessary software for analysis and presentation of results.

IV. Drug Screening

[0172] In some embodiments, drug screening assays (e.g., to screen for anticancer drugs) are provided. The screening methods utilize cancer markers identified herein (e.g., including but not limited to, SPINK1). For example, in some embodiments, methods of screening for compounds that alter (e.g., decrease or increase) the expression of cancer marker genes or alter the health of individuals expression, not expressing, or possessing a particular cancer marker are provided. In some embodiments, candidate compounds are antisense agents (e.g., oligonucleotides) directed against cancer markers. See Section IV below for a discussion of antisense therapy. In other embodiments, candidate compounds are antibodies or small molecules that specifically bind to a cancer marker and inhibit its biological function.

[0173] In one screening method, candidate compounds are evaluated for their ability to alter cancer marker expression by contacting a compound with a cell expressing a cancer marker

and then assaying for the effect of the candidate compounds on expression. In some embodiments, the effect of candidate compounds on expression of a cancer marker gene is assayed for by detecting the level of cancer marker mRNA expressed by the cell. mRNA expression can be detected by any suitable method. In other embodiments, the effect of candidate compounds on expression of cancer marker genes is assayed by measuring the level of polypeptide encoded by the cancer markers. The level of polypeptide expressed can be measured using any suitable method, including but not limited to, those disclosed herein.

[0174] Specifically, screening methods for identifying modulators, i.e., candidate or test compounds or agents (e.g., proteins, peptides, peptidomimetics, peptoids, small molecules or other drugs) which bind to cancer markers described herein, have an inhibitory (or stimulatory) effect on, for example, cancer marker expression or cancer marker activity, or have a stimulatory or inhibitory effect on, for example, the expression or activity of a cancer marker substrate are provided. Compounds thus identified can be used to modulate the activity of target gene products (e.g., cancer marker genes) either directly or indirectly in a therapeutic protocol, to elaborate the biological function of the target gene product, or to identify compounds that disrupt normal target gene interactions. Compounds that inhibit the activity or expression of cancer markers are useful in the treatment of proliferative disorders, e.g., cancer, particularly prostate cancer.

[0175] In one embodiment, the assays for screening candidate or test compounds that are substrates of a cancer marker protein or polypeptide or a biologically active portion thereof are provided. In another embodiment, the assays for screening candidate or test compounds that bind to or modulate the activity of a cancer marker protein or polypeptide or a biologically active portion thereof are provided.

[0176] The test compounds can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including biological libraries; peptoid libraries (libraries of molecules having the functionalities of peptides, but with a novel, non-peptide backbone, which are resistant to enzymatic degradation but which nevertheless remain bioactive; see, e.g., Zuckermann et al., *J. Med. Chem.* 37: 2678-85 [1994]); spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the 'one-bead one-compound' library method; and synthetic library methods using affinity chromatography selection. The biological library and peptoid library approaches are preferred for use with peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam (1997) *Anticancer Drug Des.* 12:145).

[0177] Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt et al., *Proc. Natl. Acad. Sci. U.S.A.* 90:6909 [1993]; Erb et al., *Proc. Natl. Acad. Sci. USA* 91:11422 [1994]; Zuckermann et al., *J. Med. Chem.* 37:2678 [1994]; Cho et al., *Science* 261:1303 [1993]; Carrell et al., *Angew. Chem. Int. Ed. Engl.* 33:2059 [1994]; Carrell et al., *Angew. Chem. Int. Ed. Engl.* 33:2061 [1994]; and Gallop et al., *J. Med. Chem.* 37:1233 [1994].

[0178] Libraries of compounds may be presented in solution (e.g., Houghten, *Biotechniques* 13:412-421 [1992]), or on beads (Lam, *Nature* 354:82-84 [1991]), chips (Fodor, *Nature* 364:555-556 [1993]), bacteria or spores (U.S. Pat. No. 5,223,409; herein incorporated by reference), plasmids (Cull et al., *Proc. Natl. Acad. Sci. USA* 89:18651869 [1992]) or on

phage (Scott and Smith, *Science* 249:386-390 [1990]; Devlin *Science* 249:404-406 [1990]; Cwirla et al., *Proc. Natl. Acad. Sci.* 87:6378-6382 [1990]; Felici, *J. Mol. Biol.* 222:301 [1991]).

V. Cancer Therapies

[0179] In some embodiments, therapies for cancer (e.g., prostate cancer) are provided. In some embodiments, therapies target cancer markers (e.g., including but not limited to, SPINK1). For example, in some embodiments, antisense or RNAi therapies that target the expression of SPINK1 are utilized. Additional therapeutic agents are identified, for example, using the drug screening methods described herein.

EXPERIMENTAL

[0180] The following examples are provided in order to demonstrate and further illustrate certain preferred embodiments and aspects and are not to be construed as limiting the scope thereof.

Example 1

Outlier Expression of SPINK1 in Prostate Cancer

[0181] This Example describes the characterization of SPINK1 expression in prostate cancer.

A. Materials and Methods

Cancer Outlier Profile Analysis (COPA)

[0182] COPA analysis was performed on 7 prostate cancer gene expression data sets (Yu et al., *J Clin Oncol* 22, 2790-9 (2004); Lapointe et al., *Proc Natl Acad Sci USA* 101, 811-6 (2004); Glinsky et al., *J Clin Invest* 113, 913-23 (2004); Vanaja et al., *Cancer Res* 63, 3877-82 (2003), Dhanasekaran et al., *Nature* 412, 822-6 (2001), LaTulippe et al., *Cancer Res* 62, 4499-506 (2002); Welsh et al., *Cancer Res* 61, 5974-8 (2001)) in Oncomine 3.0 as described (Tomlins et al. *Science* 310, 644-8 (2005)). COPA has three steps. First, gene expression values are median-centered, setting each gene's median expression value to zero. Second, the median absolute deviation (MAD) is calculated and scaled to 1 by dividing each gene expression value by its MAD. Of note, median and MAD were used for transformation as opposed to mean and standard deviation so that outlier expression values do not unduly influence the distribution estimates, and are thus preserved post normalization. Third, the 75th, 90th, and 95th percentiles of the transformed expression values are tabulated for each gene and then genes are rank-ordered by their percentile scores, providing a prioritized list of outlier profiles. Genes scoring in the top 100 outliers at any of the three percentile cutoffs were called outliers. Genes identified as outliers in the same number of studies were further ranked by their average outlier rank in those outlier studies. SPINK1 expression was also interrogated in prostate cancer specimens from two multi-cancer profiling studies (Bittner et al., *Nat Biotechnol* 23, 183-4 (2005), Su et al., *Cancer Res* 61, 7388-93 (2001)).

Samples

[0183] Tissues used for quantitative PCR were from the radical prostatectomy series at the University of Michigan and from the Rapid Autopsy Program, which are both part of University of Michigan Prostate Cancer Specialized Program

of Research Excellence (S.P.O.R.E.) Tissue Core. For combined fluorescence in situ hybridization (FISH) and immunohistochemistry (IHC) evaluation, the University of Michigan (UM) cohort consisted of samples from the radical prostatectomy series. The Swedish Watchful Waiting (SWW) cohort consisted of samples from a Swedish population-based cohort of men with localized prostate cancer diagnosed incidentally by trans-urethral resection of the prostate for symptomatic benign prostatic hyperplasia as described (Andren et al., *J Urol* 175, 1337-40 (2006); Johansson et al., *JAMA* 291, 2713-9 (2004)). The Memorial Sloan Kettering Cancer Center (MSKCC) cohort consisted of patients with localized or locally advanced prostate cancer that were treated by radical prostatectomy at MSKCC between 1985 and 2003. All samples were obtained with IRB approval from the respective institution. The prostate cancer cell line 22RV1 was provided by Jill Macoska (University of Michigan).

Quantitative PCR (qPCR) from Tissue Samples

[0184] Quantitative PCR (qPCR) was performed using SYBR Green dye on an Applied Biosystems 7300 Real Time PCR system (Applied Biosystems, Foster City, Calif.) essentially as described (Tomlins et al., *Science* 310, 644-8 (2005); Tomlins et al., *Cancer Res* 66, 3396-400 (2006)). Briefly, total RNA was isolated from tissues using Trizol (Invitrogen, Carlsbad, Calif.). RNA was quantified using a ND-1000 spectrophotometer (Nanodrop Technologies, Wilmington, Del.) and 3-5 μ g of total RNA was reverse transcribed into cDNA using SuperScript III (Invitrogen) in the presence of random primers. All qPCR reactions were performed with Power SYBR Green Master Mix (Applied Biosystems) and 25 ng of both the forward and reverse primer using the manufacturer's recommended thermocycling conditions. For each experiment, threshold levels were set during the exponential phase of the qPCR reaction using Sequence Detection Software version 1.2.2 (Applied Biosystems). The amount of ERG, ETV1 and SPINK1 relative to the average of the housekeeping genes GAPDH and HMBS for each sample was determined using the comparative threshold cycle (C_t) method (according to the Applied Biosystems User Bulletin #2). The relative amount of ERG, ETV1 and SPINK1 for each sample was calibrated to the median amount from all samples for each gene. All oligonucleotide primers were synthesized by Integrated DNA Technologies (Coralville, Iowa). GAPDH and HMBS, and ERG (exon5_6) and ETV1 (exon6_7) primers were as described (Tomlins et al., 2005, supra). Sequences for SPINK1 are as follows:

SPINK1_f- CAAAAATCTGGGCCTTGCTGAGAAC (SEQ ID NO: 1)

SPINK1_r- AGGCCTCGCGGTGACCTGAT (SEQ ID NO: 2)

[0185] Approximately equal efficiencies of the primers were confirmed using serial dilutions of pooled prostate cancer cDNA in order to use the comparative C_t method. All reactions were subjected to melt curve analysis.

Immunohistochemistry (IHC) and Fluorescence In Situ Hybridization (FISH)

[0186] IHC for the University of Michigan (UM) and Swedish Watchful Waiting (SWW) cohorts was performed using a mouse monoclonal antibody against SPINK1 (H00006690-M01, Abnova, Taipei City, Taiwan) on tissue microarrays containing cores from 75 (UM) and 312 (SWW) evaluable cases of localized prostate cancer. Staining in

greater than 1% of cancerous epithelial cells was deemed positive. Previously, we have evaluated cases on these tissue microarrays for TMRPSS2:ERG fusion status by FISH using break apart ERG assays as previously described (Tomlins et al., 2005, supra). A one-sided Fisher's exact test was used to evaluate the relationship between SPINK1 and fusion status, as these studies were performed with the prior hypothesis that there was an inverse correlation between SPINK1 expression and fusion status.

MSKCC Immunohistochemistry

[0187] IHC for the MSKCC cohort was performed using an in-house mouse monoclonal antibody against SPINK1 (code 6E832) on tissue microarrays containing triplicate cores from 817 evaluable cases of localized prostate cancer. The percentage of positive tumor cells in each core was estimated and assigned values of 0%, 5%, or multiples of 10%. The intensity of the expression was assigned a value of 0, 1, 2, or 3. Triplicate cores from each specimen were scored separately and the presence of tumorous tissue in at least two interpretable cores was required to include a case for analysis. Cases were marked as SPINK1 positive if any of the three cores showed >80% of cancerous cells showing positive SPINK1 immunoreactivity (intensity 1-3).

Outcome Analyses

[0188] For Kaplan-Meier analysis of the Glinsky et al. and UM datasets, biochemical recurrence was defined as a 0.2 ng/ml increase in PSA or recurrence of disease after prostatectomy, such as development of metastatic cancer, if biochemical recurrence information was not available. For the MSKCC cohort, only biochemical recurrence, defined as PSA >0.2 ng/ml after surgical resection with a second confirmatory PSA-measurement >0.2 ng/ml, was considered, as all patients with a clinical failure had previously had a biochemical recurrence. For outcome analysis from the Glinsky et al. dataset, samples positive for outlier expression of SPINK1 were defined as those with greater than 0.5 normalized expression units (as shown in FIG. 1A). For the IHC analysis of the UM and MSKCC cohorts, positive cases were defined as described above. Kaplan-Meier analysis and multivariate Cox proportional-hazards regression were then used to examine the association of SPINK1 with biochemical PSA recurrence. To predict the probability of disease recurrence, the Kattan 7-year post-operative nomogram (Kattan et al., *J Clin Oncol* 17, 1499-507 (1999)) was used, and the concordance index of the nomogram and the nomogram plus SPINK1 status was evaluated using 1000 times bootstrapping as described (Kattan et al., *J Clin Oncol* 21, 3573-9 (2003)).

Urine Based Detection of SPINK1 Expression

[0189] The collection of urine, isolation of RNA, RNA amplification and qPCR for TMRPSS2:ERG from men with prostate cancer was as described (Laxman, B. et al. Noninvasive detection of TMRPSS2:ERG fusion transcripts in the urine of men with prostate cancer. *Neoplasia* 8, 885-8 (2006)). Briefly, 25 ng of isolated RNA was amplified using TransPlex Whole Transcriptome Amplification (WTA) kit (Rubicon Genomics, Ann Arbor, Mich.) according to the manufacturer's instructions. For each qPCR reaction, 10 ng of WTA amplified cDNA was used as template. 2 \times Power SYBR Green Master Mix (Applied Biosystems, Foster City, Calif.) and 25 ng of both the forward and reverse primers were used

for SPINK1, ERG (primers as described above) and PSA (Laxman et al., 2006, supra). For all experiments, the same threshold and baseline was set using Sequence Detection Software version 1.2.2 (Applied Biosystems). All samples with a threshold cycle (C_t) value greater than 26 for PSA were excluded to remove samples with insufficient prostate cell recovery. Samples were considered TMRPSS2:ERG positive if both ERG and TMRPSS2:ERG assays showed C_t values less than 37. The amount of SPINK1 relative to PSA was determined for each sample using the comparative C_t method and the relative amount in each sample was normalized to the median of all samples. The top 11% of SPINK1 over-expressing samples (the average percentage of SPINK1 positive samples across the other five cohorts evaluated in this study (See Table 3)) were identified as SPINK1 positive. A one-sided Fisher's exact test was used to evaluate the relationship between SPINK1 and TMRPSS2:ERG status, as this study was performed with the prior hypothesis that there was an inverse correlation between SPINK1 expression and fusion status.

In Vitro Over-Expression of SPINK1

[0190] cDNA of SPINK1 (NM_003122.2), as present in a clinical prostate cancer specimen over-expressing SPINK1, was amplified by RT-PCR using the following primers, with the forward primer including a consensus Kozak sequence (start and stop codons underlined):

```

                                     (SEQ ID NO: 3)
SPINK1_full-f:  ACCACCATGAAGGTAACAGGCATCTTCTT
                                     (SEQ ID NO: 4)
SPINK1_full-r:  TCAGCAAGGCCAGATTTTGA

```

The cDNA product was TOPO cloned into the Gateway entry vector pCR8/GW/TOPO (Invitrogen), yielding pCR8-SPINK1. To generate adenoviral constructs, pCR8-SPINK1 was recombined with pAD/CMV/V5 (Invitrogen) using LR Clonase II (Invitrogen). Control pAD/CMV/LACZ clones were obtained from Invitrogen. Adenoviruses were generated by the University of Michigan Vector Core. The benign immortalized prostate cell line RWPE was infected with SPINK1 or LACZ adenoviruses, generating RWPE-SPINK1 and RWPE-LACZ for transient over-expression.

Proliferation Assay

[0191] Proliferation for RWPE-LACZ and RWPE-SPINK1 cells was measured by a colorimetric assay based on the cleavage of the tetrazolium salt WST-1 by mitochondrial dehydrogenases (cell proliferation reagent WST1, Roche Diagnostics, Mannheim, Germany) at the indicated time points in triplicate. Cell counts for 22RV1 cells were estimated by trypsinizing cells and analysis by Coulter counter (Beckman Coulter, Fullerton, Calif.) at 72 hours in triplicate.

Invasion Assays

[0192] For invasion assays, RWPE-SPINK1 and RWPE-LACZ cells (48 hours after infection with adenoviruses), or 22RV1 cells were used. Equal numbers of the indicated cells were seeded onto the basement membrane matrix (EC matrix, Chemicon, Temecula, Calif.) present in the insert of a 24 well culture plate, with fetal bovine serum added to the lower chamber as a chemoattractant. After 48 hours, non-invading

cells and EC matrix were removed by a cotton swab. Invaded cells were stained with crystal violet and photographed. The inserts were treated with 10% acetic acid and absorbance was measured at 560 nm.

SPINK1 Knockdown

[0193] For siRNA knockdown of SPINK1 in 22RV1 cells, the individual siRNAs composing the Dharmacon SMART-pool against SPINK1 (LQ-019724-00, Chicago, Ill.) were tested for SPINK1 knockdown by qPCR, and the most effective single siRNA (J-019724-07) was used for further experiments. siCONTROL Non-Targeting siRNA #1 (D-001210-01) or siRNA against SPINK1 was transfected into 22RV1 cells using Oligofectamine (Invitrogen). After 24 hours a second identical transfection was carried out and cells were harvested 24 hours later for RNA isolation, invasion assays or proliferation assays as described above.

Expression Profiling

[0194] Expression profiling was performed using the Agilent Whole Human Genome Oligo Microarray (Santa Clara, Calif.). Total RNA isolated using Trizol was purified using the Qiagen RNeasy Micro kit (Valencia, Calif.). One μ g of total RNA was converted to cRNA and labeled according to the manufacturer's protocol (Agilent). Hybridizations were performed for 16 hrs at 65° C., and arrays were scanned on an Agilent DNA microarray scanner Images were analyzed and data extracted using Agilent Feature Extraction Software 9.1.3.1, with linear and lowess normalization performed for each array. For 22RV1-siSPINK1 hybridizations, the reference was 22RV1 cells infected with non-targeting siRNA. Duplicate hybridizations were performed with duplicate dye flips, for a total of four arrays. Over and under-expressed signatures were generated by filtering to include only features with significant differential expression (PValueLogRatio<0.01) in all hybridizations and Cy5/Cy3 ratios > or <1 in all hybridizations.

B. Results

[0195] As described above, SPINK1 (serine peptidase inhibitor, Kazal type 1), the 2nd ranked meta-outlier, was identified as showing over-expression in prostate cancer compared to benign prostate tissue and mutually exclusive over-expression with ERG and ETV1 across multiple studies. The profile of SPINK1 expression and scatter plots with ERG and ETV1 for two studies (Glinsky et al., J Clin Invest 113, 913-23 (2004); Yu et al., J Clin Oncol 22, 2790-9 (2004)) where SPINK1 was identified as a top 100 outlier are shown in FIG. 1b, with plots from 7 additional studies (Dhanasekaran et al., Nature 412, 822-6 (2001); LaTulippe et al., Cancer Res 62, 4499-506 (2002); Vanaja et al., Cancer Res 63, 3877-82 (2003); Welsh et al., Cancer Res 61, 5974-8 (2001); Su et al., Cancer Res 61, 7388-93 (2001), GSE2109 and GSE8218) measuring SPINK1 expression shown in FIG. 5. In total, from these studies, SPINK1 showed outlier expression in only 2 of 127 (1.6%) benign prostate tissue samples and 64 of 387 (16.5%) prostate cancers (two sided Fisher's exact test, p=9.5E-7). Three hundred eighty four of 387 profiled prostate cancers (99.2%) showed mutually exclusive over-expression of SPINK1, ERG, and ETV1, as shown in FIGS. 1b and 5.

[0196] To confirm the outlier expression of SPINK1 exclusively in ETS negative prostate cancers, SPINK1, ERG and ETV1 expression was measured by quantitative PCR (qPCR)

in an independent cohort of 10 benign prostate tissues and 61 prostate cancers (54 clinically localized and 7 metastatic samples). While ERG, ETV1, or SPINK1 was markedly over-expressed in 25 (41%), 4 (6.5%), and 4 (6.5%) of 61 prostate cancers, respectively, no benign prostate tissue samples demonstrated over-expression of these genes. Confirming the profiling studies described above, ERG, ETV1 and SPINK1 were over-expressed in distinct samples (FIG. 6).

[0197] After demonstrating that outlier expression of SPINK1 defines a subset of ETS rearrangement negative prostate cancers at the transcript level, the expression of SPINK1 protein in prostate cancers was evaluated. By immunohistochemical (IHC) analysis on tissue microarrays, SPINK1 expression was evaluated in two independent cohorts, (University of Michigan (UM) and Swedish Watchful Waiting (SWW)) representing a total of 392 cases of clinically localized prostate cancers. Both cohorts have been previously evaluated for TMRPSS2:ERG fusion status by fluorescence in situ hybridization (FISH). In both cohorts, prostate cancer epithelia exhibited either strong or no expression of SPINK1, without intermediate staining as observed for many prostate cancer markers. As shown in FIGS. 2a-b, in the UM cohort, 10 and 36 of 75 cases were positive for SPINK1 expression (13.3%) and TMRPSS2:ERG fusions (48%), respectively, with all SPINK1 positive cases being TMRPSS2:ERG negative (one sided Fisher's exact test, $p=0.0008$). In the SWW cohort, 23 and 57 of 312 cases were positive for SPINK1 expression (7.4%) and TMRPSS2:ERG fusions (18.3%), respectively, again with all SPINK1 positive cases being TMRPSS2:ERG negative (one sided Fisher's exact test, $p=0.008$).

[0198] Approximately 25-40% of patients treated by radical prostatectomy for clinically localized prostate cancer will experience disease recurrence, initially indicated by an increase in the serum level of PSA (biochemical recurrence) (Han et al., *Urol Clin North Am* 28, 555-65 (2001); Hull et al., G. W. et al. Cancer control with radical prostatectomy alone in 1,000 consecutive patients. *J Urol* 167, 528-34 (2002)). Thus, it was next determined if SPINK1 outlier status was associated with biochemical recurrence after surgical resection. Two datasets were identified from the evaluated cohorts for which there was follow-up biochemical recurrence information and a sufficient number of SPINK1 positive cases (>5). The Glinsky gene expression dataset, which contained tumors from 79 patients (with 37 recurrences), 10 of which showed outlier mRNA transcript expression of SPINK1 (>0.5 normalized expression units), was examined. These patients had a significantly higher risk of recurrence than patients without outlier SPINK1 expression (hazard ratio: 2.65, 95% CI: 1.16-6.07, log rank $p=0.016$) by Kaplan-Meier analysis (FIG. 2c). Multivariate Cox proportional-hazards regression analysis also revealed that SPINK1 outlier status, independent of Gleason score, lymph node status, surgical margin status, age and pre-operative PSA, was a significant predictor of clinical recurrence of prostate cancer (hazard ratio: 2.5; 95% CI: 1.1-6.0; $p=0.035$, Table 2).

[0199] The same analysis was next performed on the UM cohort (75 cases, 28 recurrences) evaluated for SPINK1 status by IHC. By Kaplan-Meier analysis, SPINK1 positive staining was significantly associated with biochemical recurrence (hazard ratio: 2.49, 95% CI: 1.01-6.18, $p=0.04$, FIG. 2d). Multivariate Cox proportional-hazards regression analysis again confirmed that SPINK1 status predicted recurrence independently of other clinical parameters (Table 2). With an

adjusted hazard ratio of 4.1 (95% CI: 1.4-11.7, $p=0.009$), it was the strongest predictor in this model.

[0200] As a final validation, IHC for SPINK1 status was performed on an independent cohort of 817 evaluable prostate cancers (200 recurrences) from the Memorial Sloan Kettering Cancer Center (MSKCC). In this MSKCC cohort, with IHC performed independently from the UM and SWW cohorts using a different SPINK1 antibody, 297 of the 817 cases (36%) of cases showed positive SPINK1 immunoreactivity in at least one of three triplicate cores. In addition, staining intensity was more variable than that observed in the UM and SWW cohorts. As the percentage of cases in this cohort with SPINK1 staining (36%) is far greater than the other IHC cohorts (13% and 7%) or the percentage of SPINK1 outlier samples from DNA microarray and qPCR studies (17% and 7%, see Table 3), SPINK1 positive cases in the MSKCC cohort were defined as those with at least one core showing greater than 80% of cells showing positive SPINK1 immunoreactivity, resulting in 75 SPINK1 positive cases (9%), consistent with the other studies. By Kaplan-Meier analysis, SPINK1 positive cases in the MSKCC cohort showed significantly shorter time to biochemical recurrence (hazard ratio: 2.32, 95% CI: 1.59-3.39, $P=6.96E-06$, FIG. 2e). Multivariate Cox proportional-hazards regression analysis again confirmed that SPINK1 outlier status, independent of Gleason score, lymph node status, surgical margin status, seminal vesicle invasion, extracapsular extension and pre-operative PSA, was a significant predictor of clinical recurrence (hazard ratio: 2.02; 95% CI: 1.37-2.99; $p=0.0004$, Table 2). Clinically, nomograms are commonly used to predict the likelihood of biochemical recurrence after surgical resection by optimally incorporating clinical and pathological parameters. To determine whether the addition of SPINK1 improves a validated nomogram for predicting the 7-year post-prostatectomy probability of biochemical recurrence (Kattan et al., *J Clin Oncol* 17, 1499-507 (1999)), the concordance index (Kattan et al., *J Clin Oncol* 21, 3573-9 (2003)) (the probability that given two randomly selected patients, the patient with the worse outcome is indeed predicted to have a worse outcome) of the nomogram and the nomogram plus SPINK1 status were assessed. The bootstrap-corrected concordance index was minimally improved in all three datasets by the addition of SPINK1 status to the nomogram (Glinsky et al: 0.772 vs 0.762, UM IHC: 0.698 vs 0.676 and MSKCC: 0.775 vs. 0.765). Thus, while SPINK1 does not dramatically add to the predictive ability of an optimized multivariate model, it was demonstrated by analyzing 971 cancers from three cohorts that SPINK1 outlier status identifies an aggressive subset of prostate cancers.

[0201] It was next determined if outlier expression of SPINK1 could be detected non-invasively. As serum levels of SPINK1 may be dysregulated as the result of a number of malignancies (Paju and Stenman, *Crit Rev Clin Lab Sci* 43, 103-42 (2006); Stenman et al., *Clin Chem* 48, 1206-9 (2002) and 44% of patients with prostate cancer are reported to have elevated serum levels of SPINK1 (Paju et al., *Eur Urol* (2007)), a specific assay to identify the ~1% of patients with outlier expression (See Table 3) was developed. The detection of TMRPSS2:ERG fusion transcripts in the urine of men with prostate cancer was recently described (Laxman et al., *Neoplasia* 8, 885-8 (2006)), and this assay allows one to more directly assess transcripts contributed by prostatic cells. Thus SPINK1 expression was assessed from a cohort of 148 urine samples collected from men with prostate cancer that have

been characterized as TMRPSS2:ERG positive (43) or negative (105). As shown in FIG. 2f, while 1 of the 43 TMRPSS2:ERG positive samples (2.3%) showed marked SPINK1 over-expression (top 11% of normalized SPINK1 expression), 15 of the 105 TMRPSS2:ERG negative samples (14%) showed marked SPINK1 over-expression (Fisher's exact test, $p=0.02$).

[0202] To investigate a functional role for SPINK1 in prostate cancer, adenoviruses expressing SPINK1 were generated and benign immortalized prostate epithelial cell line RWPE was infected with the adenovirus to generate RWPE-SPINK1 cells. Over-expression of SPINK1 had no significant effect on the proliferation or invasion of RWPE cells (FIG. 3a-b). As SPINK1 over-expression had no effect on benign prostate cells, we hypothesized that SPINK1 over-expression may occur later in prostate cancer progression in the presence of co-existing genetic lesions, consistent with its association with aggressive prostate cancer. Thus, a panel of prostate cancer cell lines was analyzed for SPINK1 outlier-expression. Outlier expression of SPINK1 was identified exclusively in the 22RV1 cell line (FIG. 3c), consistent with previous work (Paju et al., supra). The aggressive 22RV1 prostate cancer cell line was derived from a human prostate carcinoma xenograft that was serially propagated in nude mice after castration-induced regression and relapse of the parental, androgen-dependent CWR22 xenograft (Sramkowski, R. M. et al. A new human prostate carcinoma cell line, 22RV1. *In Vitro Cell Dev Biol Anim* 35, 403-9 (1999)). 22RV1 does not over-express ERG or ETV1 (FIG. 3c), similar to clinical SPINK1 outlier cases, supporting its use as a cell line model of SPINK1 outlier expression.

[0203] To assess the function of SPINK1 in 22RV1, siRNA knockdown was utilized. While SPINK1 knockdown had no effect on 22RV1 proliferation (FIG. 3d), SPINK1 knockdown markedly attenuated the invasiveness of 22RV1 cells through a modified basement membrane (FIGS. 3e-f). Consistent with the mutually exclusive over-expression of ERG, ETV1 and SPINK1, siRNA knockdown of ERG or ETV1 in 22RV1 had no effect on invasion, while SPINK1 knockdown had no effect on the invasiveness of VCaP (TMRPSS2:ERG positive) or LNCaP (ETV1 rearrangement positive) (FIGS. 3g-h). siRNA knockdown of ERG in VCaP and ETV1 in LNCaP similarly attenuated invasion (FIGS. 3g-h) without affecting proliferation (Tomlins et al., 2007, supra). Additionally, microarray analysis of 22RV1-siSPINK1 cells revealed only limited transcriptional effects (76 features over-expressed, 14 features under-expressed, Table 4 and FIG. 7), indicating that SPINK1 knockdown directly affects cellular invasiveness. Together, these results provide a role for SPINK1 in prostate cancer invasion, consistent with its over-expression in aggressive prostate cancers.

[0204] The outlier expression of SPINK1 in a subset of prostate cancers suggests that SPINK1 may be activated by a unique molecular event, similar to TMRPSS2:ETS positive prostate cancers. However, FISH studies using locus/control and 5'/3' split probes demonstrated no evidence of amplification or gross rearrangements, respectively, in samples with SPINK1 over-expression. Additionally, sequencing of the SPINK1 coding region identified no mutations in samples with SPINK1 outlier-expression. The present disclosure is not limited to a particular mechanism. Indeed, an understanding of the mechanism is not necessary to understand the invention. Nonetheless, it is contemplated that SPINK1 is activated by increased transcription, for example through pro-

motor mutations affecting regulatory elements. Alternatively, SPINK1 may be activated by a unique upstream genetic event. However, few genes show consistent correlation with SPINK1 across data sets (FIG. 8), indicating that SPINK1 is an exclusive downstream target. It is also possible that SPINK1 is down-regulated in TMRPSS2:ETS positive cancers, however one would expect high SPINK1 expression in benign prostatic epithelium and the in vitro data described above supports a role for SPINK1 over-expression in prostate cancer progression.

[0205] In conclusion, using a combination of in silico bio-informatics analysis coupled with independent experimental validation, data on 1,800 prostate cancers was analyzed, demonstrating the consistent outlier expression of SPINK1 in TMRPSS2:ETS negative prostate cancers (Table 3). Evidence is provided that SPINK1 over-expression defines an aggressive molecular sub-type of prostate cancer (~11% of cases) not attributable to known gene fusion events. It was demonstrated that SPINK1 can be monitored non-invasively in urine and thus serves to complement gene fusion based urine testing for prostate cancer. Additionally, the utility of 22RV1 as a cell line model for SPINK1 outlier expression was demonstrated. In further experiments, a meta-COPA strategy was used to nominate candidate oncogenes in specific cancer types.

TABLE 1

Meta-COPA analysis of 7 prostate cancer gene expression profiling datasets in Oncomine.			
Meta COPA Rank	Gene	# of Studies	Avg. Rank
1	ERG	7	19.3
2	SPINK1	5	29.8
3	GPR116	5	46
4	ORM1	4	10
5	ETV1	4	23
6	MYL2	4	26.8
7	NEB	4	27
8	TGM4	4	30.8
9	NELL2	4	33.5
10	KRT13	4	49
11	SLC26A4	4	53.3
12	MYL1	3	8.7
13	CXCL13	3	11
14	HCG3	3	12
15	HPGD	3	20.7
16	MPEP	3	27.3
17	PLAT	3	31.7
18	CST1	3	32.3
19	COL2A1	3	35.3
20	CPS1	3	36.3
21	CRISP3	3	38.7
22	CTAG1A	3	42.3
23	FGF	3	42.3
24	PPFIA2	3	42.7
25	CDH2	3	44
26	PTPRM	3	45.3
27	CVP3A43	3	51.7
28	PROM1	3	53
29	HDAC9	3	51.7

Genes were ranked by the number of studies in which they scored in the top 100 outliers (ranked by COPA) at any of the three pre-defined percentile cutoffs (75th, 90th, 95th). Genes were further ranked by their average COPA rank (Avg. Rank) in studies where they ranked in the top 100.

TABLE 2

Multivariate Cox Proportional-Hazards Regression Analysis for the association of SPINK1 with Biochemical Recurrence.				
Glinsky et al. microarray dataset	Recurrence			P
	Ratio	95% CI		
SPINK1 (Positive vs. Negative)	2.522	1.067	5.950	0.035
Gleason (>=7 vs. <7)	2.768	0.606	9.500	0.106
Lymph node (Positive vs. Negative)	2.537	0.722	8.916	0.147
Surgical Margin (Positive vs. Negative)	2.284	0.975	5.353	0.057
Pre-operative PSA	1.040	1.001	1.081	0.047
Age	1.042	0.979	1.108	0.194

Note:

Sample size is 79 with 37 recurrences. Tumor size was missing the clinical table and thus not included in the model.

University of Michigan IHC cohort	Recurrence			P
	Ratio	95% CI		
SPINK1 (Positive vs. Negative)	4.599	1.431	11.741	0.009
Gleason (>=7 vs. <7)	1.957	0.758	5.371	0.187
Tumor size (>=2 cm vs. <2 cm)	1.365	0.620	3.005	0.439
Surgical Margin (Positive vs. Negative)	1.342	0.588	3.180	0.504
Pre-operative PSA	1.050	0.988	1.118	0.127
Age	1.061	1.006	1.118	0.028

Note:

Sample size is 75 with 28 recurrences.

MSKCC IHC cohort	Recurrence			P
	Ratio	95% CI		
SPINK1 (Positive vs. Negative)	2.02	1.37	2.99	0.004
Pre-operative PSA	1.02	1.01	1.03	0.0002
Seminal Vesical Invasion (Positive vs. Negative)	3.07	2.14	4.40	1.1E-09
Surgical Margin (Positive vs. Negative)	1.57	1.15	2.14	0.004

TABLE 2-continued

Multivariate Cox Proportional-Hazards Regression Analysis for the association of SPINK1 with Biochemical Recurrence.				
Lymph Node Involvement (Pos vs. Neg)	2.89	1.78	4.88	1.5E-05
Extracapsular Extension (Yes vs. No)	1.96	1.42	2.71	4.6E-05
Pathology Gleason (<=7 vs. <7)	1.88	1.18	2.40	0.004

Note:

Sample size is 817 with 200 recurrences.

TABLE 3

SPINK1 outlier status across cohorts.

Cohort	Total	SPINK1+	% SPINK1+	% Exclusive
Microarray	387	64	17%	99.2%
qPCR tissue	61	4	7%	100%
UM IHC/FISH	75	10	13%	100%
SWW IHC/FISH	312	23	7%	100%
MSKCC IHC	517	75	9%	NA
qPCR urine	148	16	11%*	NA
	1800	192	11%*	99.8%*

For the six cohorts evaluated in this study, the total number of prostate cancer samples analyzed, the number and percentage of SPINK1 positive samples (as defined in the Methods for each assay), and the percentage of samples showing mutually exclusive overexpression of SPINK1 and ERG, ETV1 or TMPRSS2:ERG (if measured). The six cohorts are: the in silico microarray data, quantitative PCR (qPCR) on tissue samples, University of Michigan (UM) immunohistochemistry (IHC)/fluorescence in situ hybridization (FISH), Swedish Watchful Waiting (SWW) IHC/FISH, Memorial Sloan Kettering Cancer Center (MSKCC) IHC and qPCR on urine samples. *Individual percentages from the non-urine based cohorts were averaged. To define SPINK1 positive samples from the qPCR urine cohort, the 11% average SPINK1 positivity from the other cohorts was used.

TABLE 4

Differentially expressed genes upon siRNA knockdown of SPINK1 in 22RV1 cells.

Feature ID	ProbeName	Gene	Sequence	siRNA SPINK1 NT siRNA
35382	A_24_P119201	MBD2	NM_015832	-5.97
25530	A_23_P214079	SPINK1	NM_003122	-4.60
31329	A_23_P216053	KLRC4	NM_013431	-2.28
31695	A_23_P20058	GRM3	NM_000840	-2.17
9794	A_24_P56484	BRMS1L	NM_032352	-2.08
531	A_23_P77043	C14orf161	NM_024764	-2.07
41521	A_23_P308211	FAM84A	NM_145175	-1.95
17730	A_24_P937405	PRSS23	NM_007173	-1.90
44387	A_23_P150788	PRSS23	NM_007173	-1.65
13006	A_23_P251647	LOC317671	NM_173362	-1.78
20835	A_24_P413920	FAM84A	NM_145175	-1.67
4415	A_23_P145761	ARL4A	NM_005738	-1.62
17815	A_23_P250164	HGD	NM_000187	-1.62
12965	A_24_P220947	AKR1C1	NM_001353	-1.57
36945	A_32_P75581	BHLHB5	NM_152414	3.33
25438	A_24_P166883	COK6	NM_001259	3.30
4139	A_32_P194563	THC2281660	THC2281660	3.21
5000	A_23_P201623	LAMC1	NM_002293	3.07
2464	A_23_P116414	HRASLS3	NM_007069	2.96
8126	A_23_P116235	MDK	NM_001012334	2.64
11958	A_23_P409336	SLC25A22	NM_024596	2.63
11318	A_32_P125338	FAM43B	NM_207334	2.60
27988	A_32_P218707	THC2314643	THC2314643	2.55
25309	A_23_P416608	LAMP2	NM_013995	2.53

TABLE 4-continued

Differentially expressed genes upon siRNA knockdown of SPINK1 in 22RV1 cells.				
Feature ID	ProbeName	Gene	Sequence	siRNA SPINK1 NT siRNA
44595	A_23_P141394	WIP1I	NM_017983	2.51
42459	A_24_P393311	TMCO1	NM_019026	2.44
5490	A_23_P209625	CYP11B1	NM_000104	2.42
37302	A_24_P276583	TMCO1	NM_019026	2.37
23875	A_23_P371266	DNM3	NM_015569	2.31
43554	A_23_P52793	PPP2R1B	NM_002716	2.31
15132	A_23_P130974	KIAA1683	NM_025249	2.30
5390	A_23_P355067	TMCO1	NM_019026	2.29
35903	A_24_P323084	C17orf55	NM_176519	2.27
10939	A_23_P150609	IGF2	NM_001007139	2.23
6181	A_23_P97990	HTRA1	NM_002775	2.22
11433	A_23_P133725	MARVELD1	NM_001464	2.16
17892	A_23_P380318	EGR4	NM_001965	2.14
28152	A_23_P253561	C20orf121	NM_024331	2.13
9644	A_24_P270333	ENST00000278849	ENST00000278849	2.12
28209	A_32_P191262	ACR	NM_001097	2.09
43210	A_24_P390583	THC2336533	THC2336533	2.06
42543	A_23_P164089	RFFL	NM_057178	2.07
37592	A_24_P80204	MALL	NM_005434	2.05
43445	A_23_P112452	BC045756	BC045756	2.01
14227	A_24_P521994	KLHL24	NM_017644	1.97
15160	A_23_P11915	GDAP2	NM_017686	1.96
9597	A_23_P353717	C16orf75	NM_152306	1.94
6602	A_32_P919718	TMEM105	NM_176520	1.94
38052	A_24_P211565	C10TNF6	NM_031910	1.94
39302	A_23_P168651	CDK6	NM_001259	1.93
13321	A_23_P502797	WDFY1	NM_020830	1.92
33557	A_23_P147135	CANX	NM_001746	1.91
38826	A_23_P100711	PMP22	NM_000304	1.90
15873	A_23_P99540	ZFP36L1	NM_004926	1.90
44652	A_23_P6362	DERL3	NM_196440	1.87
11337	A_23_P107421	TK1	NM_003258	1.86
5137	A_23_P110837	IRX4	NM_016358	1.85
34281	A_23_P90273	CHST3	NM_022467	1.85
2773	A_23_P360804	CPNE5	NM_020939	1.82
34557	A_23_P206724	MT1E	NM_175617	1.82
20703	A_24_P199555	VANGL1	NM_138859	1.81
21796	A_24_P379512	PIGK	NM_005482	1.81
34581	A_23_P434900	C16orf34	NM_144570	1.81
43891	A_32_P216520	WIF1	NM_007191	1.81
6627	A_23_P207131	MAP3K3	NM_203351	1.80
23135	A_23_P111995	LOX12	NM_002318	1.80
17812	A_23_P211631	FBLN1	NM_006486	1.79
18787	A_23_P52837	CD248	NM_020404	1.76
8319	A_24_P649634	ENST00000377093	ENST00000377093	1.77
21139	A_23_P2554	KIAA0152	NM_014730	1.76
19596	A_24_P592544	BC107566	BC107566	1.76
20512	A_23_P14734	RPS27L	NM_015920	1.76
26298	A_23_P5640	GPR30	NM_001505	1.75
708	A_23_P336513	GEMIN5	NM_015465	1.74
560	A_24_P921260	AK022030	AK022030	1.74
9436	A_23_P10335	OTL	NM_016448	1.70
24715	A_32_P206541	AK128714	AK128714	1.70
7603	A_23_P215431	POM121	AB014516	1.69
26525	A_23_P38694	FL11286	NM_016361	1.69
37428	A_23_P201521	TMEM97	NM_014575	1.69
9507	A_23_P162846	LAMP1	NM_005551	1.69
20121	A_23_P10385	DTL	NM_016448	1.69
10990	A_23_P10385	DTL	NM_016448	1.67
13501	A_23_P10385	DTL	NM_016448	1.67
25496	A_32_P104334	AW972815	AW972815	1.65
9342	A_24_P56270	CR812228	CR812228	1.64
20136	A_23_P395637	RG9MTD2	NM_152282	1.64

TABLE 4-continued

Differentially expressed genes upon siRNA knockdown of SPINK1 in 22RV1 cells.				
Feature ID	ProbeName	Gene	Sequence	siRNA SPINK1 NT siRNA
34560	A_24_P399812	ENST00000377275	ENST00000377275	1.63
32009	A_23_P109122	RP5-860F19.3	BC054347	1.61
23338	A_23_P23924	CAFN2	NM_001745	1.60

22RV1 cells were transfected with siRNA against SPINK1 (siRNA SPINK1) or non-targeting control siRNA (NT siRNA). Total RNA was isolated and expression profiling was performed using the Agilent Whole Human Genome Oligo Microarray (GPL4133). Hybridizations (siRNA SPINK1/NT siRNA) were performed in duplicate with duplicate dye flips. Differentially expressed features (see Methods) are indicated, including the probe name, gene name, representative sequence, and average fold change across the four hybridizations, corrected for the dye flip (siRNA SPINK1/NT siRNA).

Example 2

Multiplex Urine Assay

[0206] This example describes a multiplex urine assay for use in the assessment of samples for markers associated with prostate cancer.

A. Methods

Urine Collection, RNA Isolation, Amplification and Quantitative PCR

[0207] This study was approved by the Institutional Review Board (IRB) of the University of Michigan Medical School and samples were obtained from 276 patients with informed consent following a digital rectal exam before either needle biopsy (n=216) or radical prostatectomy (n=60) at the University of Michigan Health System (UMHS). Urine was voided into urine collection cups containing DNA/RNA preservative (Sierra Diagnostics LLC, Sonora, Calif.). Isolation of RNA from urine and whole transcriptome amplification (WTA) were as described in (Laxman et al., Neoplasia 2006; 8:885-8). Quantitative PCR (qPCR) was used to detect seven prostate cancer biomarkers (AMACR, ERG, GOLPH2, PCA3, SPINK1, TFF3, and TMPRSS2:ERG fusions) and the control transcripts PSA and GAPDH from WTA amplified cDNA essentially as described (Laxman et al., 2006, supra; Tomlins et al., Neoplasia 2006; 8:153-62). The primer sequences for ERG (exon5_6) (Tomlins et al., Science 2005; 310:644-8), GAPDH (Vandesompele et al., Genome Biol 2002; 3:RESEARCH0034), AMACR (Kumar-Sinha et al., Am J Pathol 2004; 164:787-93), and PSA (Specht et al., Am J Pathol 2001; 158:419-29) were previously described and for other biomarkers were as follows:

GOLPH2 - f :	CTGGTGGCCTGCATCATCGTCTTG,	(SEQ ID NO: 5)
GOLPH2 - r :	GCTGCTCCCGCTGCTTCTCCA,	(SEQ ID NO: 6)
PCA3 - f :	CATGGTGGGAAGGACCTGATGATAC,	(SEQ ID NO: 7)
PCA3 - r :	GATGTGTGGCCTCAGATGGTAAAGTC,	(SEQ ID NO: 8)
SPINK1 - f :	CAAAAATCTGGGCCTTGCTGAGAAC,	(SEQ ID NO: 9)

-continued

SPINK1 - r :	AGGCCTCGCGGTGACCTGAT,	(SEQ ID NO: 10)
TFF3 - f :	AACCGGGGCTGCTGCTTTGACTC,	(SEQ ID NO: 11)
TFF3 - r :	TCCTGCAGGGGCTTGAACACCA.	(SEQ ID NO: 12)

TMPS2:ERG fusions were detected using Taqman primers/probe, with the following sequences:

TM-ERGA3 - f :	CTGGAGCGCGGAGGAA,	(SEQ ID NO: 13)
TM-ERGA3 - r :	CCGTAGGCACACTCAAACAACA,	(SEQ ID NO: 14)
TM-ERGA3_MGB-probe:	5'-MGB-TTATCAGTTGTGAGTGAGGAC-3'.	(SEQ ID NO: 15)

Threshold levels were set during the exponential phase of the qPCR reaction using Sequence Detection Software version 1.2.2 (Applied Biosystems, Foster City, Calif.), with the same baseline and threshold set for each plate, to generate threshold cycle (C_t) values for all genes for each sample.

Analysis

[0208] qPCR was performed on WTA cDNA from urine collected from 111 biopsy-negative patients and 165 patients with prostate cancer (105 biopsy positive patients and 60 prostatectomy patients). Samples that had PSA C_t values greater than 27 were excluded to ensure sufficiency of the amount of prostate cells in the samples, leading to 105 biopsy-negative and 152 samples from patients with prostate cancer in the analysis. For qPCR analysis, raw $-\Delta C_t$ was used (to stabilize the variance of testing variables) as opposed to testing markers against control ($2^{-\Delta C_t}$). TMPRSS2:ERG was dichotomized as a binary variable to reflect the fusion positive or negative status observed in tissue samples (Tomlins et al., 2005, supra, Perner et al., Am J Surg Pathol 2007; 31:882-8), with positive samples defined as those with C_t values less than 37. As PCA3 has been reported to be a prostate tissue-specific marker (de Kok et al., Cancer Res 2002; 62:2695-8), it was normalized against urine PSA ($C_{tPSA} - C_{tPCA3}$). All other testing variables were adjusted against their mean urine PSA and GAPDH values ($(C_{tPSA} + C_{tGAPDH})/2 - C_{tvariable}$). Twenty two

samples showing outlier values were excluded, as at least one testing variable in those samples showed an adjusted value below 3 standard deviations from its sample mean, indicating qPCR failure. This resulted in a final data set of samples from 138 patients with prostate cancer (86 positive needle biopsy and 52 radical prostatectomy patients) and 96 biopsy-negative patients.

Statistical Analysis

[0209] Univariate and multivariate logistic regressions were used to examine associations between prostate cancer diagnostic status and testing variables. For multivariate logistic regression, the Akaike Information Criterion (AIC)-based backward selection was used to drop insignificant terms (Venables WNaR, B. D Modern Applied Statistics with S, 4th edition: New York: Springer, 2002). All testing markers were included in the initial regression model which was further refined by the AIC-based backward selection. After the final model was determined, the predicted probability for each sample was used as input to generate the receiver operating characteristic (ROC) curve and the area under the curve (AUC) was calculated. As all samples were used for regression model generation, the estimated AUC may be over-optimized. To correct this bias, a leave-one-out cross validation was performed. Briefly, one sample was omitted while the regression model was trained on the remaining samples to select optimal markers and estimate their coefficients. The prediction probability is then calculated based on the model prediction for the left-out sample. This was repeated until every sample was left out once and the generated prediction probability values were then used for ROC analysis. Similarly, PCA3 was fitted in a logistic regression model to generate an AUC. The difference of AUCs was examined as described previously (DeLong et al., Biometrics 1988; 44:837-45). All analyses were performed in R and ROC curves were plotted in SPSS 11.5 (SPSS Inc., Chicago, Ill., USA).

Risk Stratification

[0210] Clinical information was identified from the medical record to determine association with clinical factors such as clinical stage, Gleason score, and risk categories based on biopsy results and pathologic data (D'Amico et al., JAMA 1998; 280:969-74). Clinical nomograms were used to calculate risk of progression free survival and pathologic staging (Kattan et al., J Natl Cancer Inst 1998; 90:766-71; Kattan et al., Cancer 1997; 79:528-37). A novel risk stratification grouping (High risk vs. low risk) was also evaluated by grouping all Gleason 6 tumors, any Gleason 7 tumors with clinical T1c stage or Gleason 3 +4 tumors as a low risk tumors and Gleason 8+, stage T2 Gleason 7 tumors or tumors with Gleason 4 +3 grading as high risk tumors. All variables were tested for univariate association with each clinical risk group.

B. Results

[0211] To develop a multiplexed qPCR based test for prostate cancer, we assessed seven putative prostate cancer biomarkers based on previously published reports and analysis from our group in a final cohort of 138 patients with prostate cancer (86 positive needle biopsy and 52 radical prostatectomy patients) and 96 patients with negative needle biopsies. Biomarkers included those generally over-expressed in prostate cancer, such as PCA3, AMACR and GOLPH2 (Rubin et

al., JAMA 2002; 287:1662-70; de Kok et al., supra), as well as those over-expressed in subsets of prostate cancers, such as ERG and TMPRSS2:ERG, and TFF3 and SPINK1 (Tomlins et al, 2005, supra; Faith et al., Prostate 2004; 61:215-27; Garraway et al., Prostate 2004; 61:209-14).

[0212] All genes were first tested by univariate analysis, with GOLPH2 (P=0.0002), SPINK1 (P=0.0002), PCA3 (P=0.001) and TMPRSS2:ERG fusion (P=0.034) showing significant association for discriminating patients with prostate cancer from patients with negative needle biopsies (FIG. 9 and Table 5). Both AMACR, which has previously been shown to be a sensitive and specific biomarker for prostate cancer in tissues (Rubin et al., 2002, supra) and TFF3, which shows high expression in a subset of prostate cancers (Faith et al., Prostate 2004; 61:215-27; Garraway et al., Prostate 2004; 61:209-14), were not statistically significant predictors of prostate cancer using urine samples (P=0.450 and 0.189, respectively). The lack of specificity of AMACR and TFF3 in urine may be due to expression of these transcripts in urothelial or kidney derived cellular material which can also be shed in the urine. While TMPRSS2:ERG fusion was significantly associated with the presence of prostate cancer (FIG. 9 and Table 5), ERG overexpression was not associated with cancer presence on univariate analysis (P=0.166), indicating that cells from other tissues may be contributing ERG transcripts to the urine. Additionally, serum PSA levels prior to biopsy or prostatectomy were also not associated with cancer presence in this cohort (P=0.376). When tested as individual variables for the ability to detect prostate cancer based on the receiver-operating-characteristic curves (ROC), GOLPH2 (area under the curve (AUC)=0.664, P=2.01E-5), PCA3 (AUC=0.661, P=2.84E-5), and SPINK1 (AUC=0.642, P=0.0002) outperformed serum PSA (AUC=0.508, p=0.837) (FIG. 9). Thus, this study identified a number of novel biomarkers for the non-invasive detection of prostate cancer using patient urine instead of biopsy samples.

[0213] To determine if a multiplex model could improve on the performance of these single biomarkers, the analyzed prostate cancer biomarkers were next tested in a multivariate regression analysis using Akaike Information Criterion (AIC)-based backward selection (Venables, B. D Modern Applied Statistics with S, 4th edition: New York: Springer, 2002) to drop insignificant terms from the model. This analysis resulted in a final model that included SPINK1 (P=7.41E-5), PCA3 (P=0.003), GOLPH2 (P=0.004) and TMPRSS2:ERG (P=0.006) (Table 5). To evaluate the performance of this model for diagnosing prostate cancer, ROC analysis was performed based on the predicted probabilities derived from the final model. The ROC curves from the multiplexed model and PCA3 alone were compared, as urine based detection of PCA3 has previously been evaluated in similar cohorts as a single biomarker using alternative detection technologies (Venables WNaR, B. D Modern Applied Statistics with S, 4th edition: New York: Springer, 2002; Hessels et al., Eur Urol 2003; 44:8-15; discussion-6; Fradet et al., Urology 2004; 64:311-5; discussion 5-6; Groskopf et al., Clin Chem 2006; 52:1089-95; Marks et al., Urology 2007; 69:532-5). For example, van Gils et al. demonstrated that in a cohort of 534 men presenting for prostate biopsy with serum PSA between 3-15 ng/mL, urinary PCA3 detection expression had an AUC of 0.66 compared to 0.57 for serum PSA (van Gils et al., 2007, supra). As shown in FIG. 10A, in the cohort described herein, the AUC for the multiplexed model (0.758, P=1.91E-11) was significantly improved (P=0.003 (DeLong et al., Biometrics

1988; 44:837-45)) compared to the AUC for PCA3 alone (0.662, $P=2.58E-5$). At the point on the multiplex model ROC with the maximum sum of sensitivity and specificity (65.9% and 76.0%, respectively), the positive predictive value was 79.8% and the negative predictive value was 60.8% (FIG. 10A). It was demonstrated that PCA3 showed improved AUC compared to serum PSA. It was also demonstrated that a multiplex model including PCA3 significantly improves the predictive ability of PCA3 alone, demonstrating the ability to improve PCA3 and other single-gene based diagnostic tests.

[0214] As all samples were used to select the best subset of variables for regression analysis, this has the potential to over-optimize the reported AUC. Thus, a leave one-out-cross validation (LOOCV) strategy was used to generate an unbiased AUC. As shown in FIG. 10B, the AUC for the LOOCV multiplex model (0.736) is again significantly better ($P=0.006$) than that for LOOCV PCA3 alone (0.645). At the point on the LOOCV multiplex model ROC with the maximum sum of sensitivity and specificity (62.3% and 75.0%, respectively), the positive predictive value was 78.2% and the negative predictive value was 58.1% (FIG. 10B).

[0215] The ability of these genetic markers to predict clinical risk groups based on patient parameters was next assayed. Clinical risk groups were determined by clinical patient data that direct the decision to pursue biopsy, to determine treatment, or to stratify patients for surveillance regimens (see Methods). Only limited association between these prostate cancer biomarkers and clinical risk groups, with GOLPH2, SPINK1 and TMPRSS2:ERG status showing association with risk groups was observed. Similar to the previously described PCR based test for breast cancer recurrence risk, a prostate cancer risk test can be used to drive high risk patients to therapies more suited for their disease course (van Gils et al., supra).

[0216] In summary, this example describes a multiplexed qPCR based assay on sedimented urine collected from patients presenting for prostate biopsy or prostatectomy that

exhibits superior performance relative to serum PSA or PCA3 alone. The multiplex urine test, which is a combination of PCA3, SPINK1, GOLPH2 and TMPRSS2:ERG status achieves a specificity and positive predictive value of >75%, making it a useful test to complement serum PSA, which has poor specificity in detecting prostate cancer.

TABLE 5

Variable	Coefficient	P-value
<u>Univariate Logistic Regression Analysis</u>		
GOLPH2	0.4444	0.0002
SPINK1	0.25	0.0002
PCA3	0.187	0.001
TMPRSS2:ERG	0.609	0.034
ERG	0.043	0.186
TFF3	0.11	0.189
PSA (serum)	0.0151	0.376
AMACR	0.049	0.45
<u>Multivariate Logistic Regression Analysis</u>		
SPINK1	0.308	7.41E-05
PCA3	0.191	0.003
GOLPH2	0.372	0.004
TMPRSS2:ERG	0.924	0.008

[0217] All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention that are obvious to those skilled in the relevant fields are intended to be within the scope of the following claims.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 15

<210> SEQ ID NO 1
 <211> LENGTH: 25
 <212> TYPE: DNA
 <213> ORGANISM: Artificial
 <220> FEATURE:
 <223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 1

caaaaatctg ggccttgctg agaac

25

<210> SEQ ID NO 2
 <211> LENGTH: 20
 <212> TYPE: DNA
 <213> ORGANISM: Artificial
 <220> FEATURE:
 <223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 2

aggcctcgcg gtgacctgat

20

-continued

<210> SEQ ID NO 3
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 3

accacatga aggtaacagg catctttctt 30

<210> SEQ ID NO 4
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 4

tcagcaaggc ccagatTTTT ga 22

<210> SEQ ID NO 5
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 5

ctggtggcct gcatcatcgt cttg 24

<210> SEQ ID NO 6
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 6

gctgctcccg ctgcttctcc a 21

<210> SEQ ID NO 7
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 7

catggtggga aggacctgat gatac 25

<210> SEQ ID NO 8
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 8

gatgtgtggc ctcagatggt aaagtc 26

<210> SEQ ID NO 9
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial

-continued

<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 9

caaaaatctg ggccttgctg agaac 25

<210> SEQ ID NO 10
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 10

aggcctcgcg gtgacctgat 20

<210> SEQ ID NO 11
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 11

aaccggggct gctgctttga etc 23

<210> SEQ ID NO 12
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 12

tcctgcaggg gcttgaaca cca 23

<210> SEQ ID NO 13
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 13

ctggagcgcg gcaggaa 17

<210> SEQ ID NO 14
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 14

ccgtaggcac actcaaacaa cga 23

<210> SEQ ID NO 15
<211> LENGTH: 21

-continued

<212> TYPE: DNA
 <213> ORGANISM: Artificial
 <220> FEATURE:
 <223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 15

ttatcagttg tgagtgagga c

21

We claim:

1. A method for identifying prostate cancer in a patient comprising:

- (a) providing a sample containing prostate cells from the patient;
- (b) detecting in the sample containing prostate cells overexpression of SPINK1 compared to normal expression of SPINK1; and
- (c) detecting in the sample containing prostate cells normal expression of ERG and/or ETV1,

wherein detecting in the sample containing prostate cells mutually exclusive overexpression of SPINK1 compared to normal expression of ERG and/or ETV1 identifies prostate cancer in the patient.

2. The method of claim 1, wherein step (b) comprises detecting overexpression of SPINK1 RNA.

3. The method of claim 1, wherein step (b) comprises detecting overexpression of SPINK1 protein.

4. The method of claim 1, wherein the sample containing prostate cells is prostate tissue, blood, urine, semen, prostatic secretions or isolated prostate cells.

5. The method of claim 1, wherein detecting in the sample containing prostate cells overexpression of SPINK1 identifies invasive prostate cancer in the patient.

6. The method of claim 1, wherein the sample containing prostate cells is from a patient following radical prostatectomy and wherein overexpression of SPINK1 identifies recurrence of prostate cancer in the patient following radical prostatectomy.

7. A method for identifying prostate cancer in a patient comprising:

- (a) providing a sample containing prostate cells from the patient; and
- (b) detecting in the sample containing prostate cells:
 - (i) overexpression of SPINK1 compared to normal expression of SPINK1 and overexpression of PCA3 compared to normal expression of PCA3;
 - (ii) overexpression of SPINK1 compared to normal expression of SPINK1 and overexpression of GOLPH2 compared to normal expression of GOLPH2;
 - (iii) overexpression of SPINK1 compared to normal expression of SPINK1 and presence of TMPRSS2:ERG;
 - (iv) overexpression of SPINK1 compared to normal expression of SPINK1, overexpression of PCA3 compared to normal expression of PCA3 and overexpression of GOLPH2 compared to normal expression of GOLPH2;
 - (v) overexpression of SPINK1 compared to normal expression of SPINK1, overexpression of PCA3 compared to normal expression of PCA3 and presence of TMPRSS2:ERG;

(vi) overexpression of SPINK1 compared to normal expression of SPINK1, overexpression of GOLPH2 compared to normal expression of GOLPH2 and presence of TMPRSS2:ERG; or

(vii) overexpression of SPINK1 compared to normal expression of SPINK1, overexpression of PCA3 compared to normal expression of PCA3, overexpression of GOLPH2 compared to normal expression of GOLPH2 and presence of TMPRSS2:ERG,

wherein detecting in the sample containing prostate cells overexpression of SPINK1 identifies prostate cancer in the patient.

8. The method of claim 7, wherein step (b) comprises detecting overexpression of SPINK1 RNA.

9. The method of claim 7, wherein step (b) comprises detecting overexpression of SPINK1 protein.

10. The method of claim 7, wherein the sample containing prostate cells is prostate tissue, blood, urine, semen, prostatic secretions or isolated prostate cells.

11. A method for identifying prostate cancer in a patient comprising:

- (a) providing a sample containing prostate cells from the patient; and
- (b) detecting in the sample containing prostate cells:
 - (i) overexpression of SPINK1 compared to normal expression of SPINK1;
 - (ii) overexpression of PCA3 compared to normal expression of PCA3;
 - (iii) overexpression of GOLPH2 compared to normal expression of GOLPH2; and
 - (iv) presence of TMPRSS2:ERG,

wherein detecting in the sample containing prostate cells overexpression of SPINK1 identifies prostate cancer in the patient.

12. A composition comprising at least one of the following:

- (a) a first oligonucleotide probe comprising a sequence that hybridizes specifically to SPINK1 RNA or cDNA, a second oligonucleotide probe comprising a sequence that hybridizes specifically to ERG RNA or cDNA, and a third oligonucleotide probe comprising a sequence that hybridizes specifically to ETV1 RNA or cDNA;
- (b) a first pair of amplification oligonucleotides wherein each amplification oligonucleotide in the first pair comprises a sequence that hybridizes specifically to SPINK1 RNA or cDNA, a second pair of amplification oligonucleotides wherein each amplification oligonucleotide in the second pair comprises a sequence that hybridizes specifically to ERG RNA or cDNA, and a third pair of

- amplification oligonucleotides wherein each amplification oligonucleotide comprises a sequence that hybridizes specifically to ETV1 RNA or cDNA; or
- (c) a first antibody that binds specifically to SPINK1 protein, a second antibody that binds specifically to ERG protein, and a third antibody that binds specifically to ETV1 protein.
- 13.** A composition comprising at least one of the following:
- (a) at least two oligonucleotide probes comprising:
- (i) an oligonucleotide probe comprising a sequence that hybridizes specifically to SPINK1 RNA or cDNA; and
- (ii) at least one other oligonucleotide probe comprising a sequence that hybridizes specifically to:
- (A) PCA3 RNA or cDNA;
- (B) GOLPH2 RNA or cDNA; or
- (C) a junction of a chimeric RNA or cDNA in which a 5' portion of the chimeric RNA is transcribed from a TMPRSS2 gene and a 3' portion of the chimeric RNA is from transcribed from an ERG gene;
- (b) at least two pairs of amplification oligonucleotides comprising:
- (i) a pair of amplification oligonucleotides wherein each amplification oligonucleotide comprises a sequence that hybridizes specifically to SPINK1 RNA or cDNA; and
- (ii) at least one other pair of amplification oligonucleotides wherein:
- (A) each amplification oligonucleotide comprises a sequence that hybridizes specifically to PCA3 RNA or cDNA;
- (B) each amplification oligonucleotide comprises a sequence that hybridizes specifically to GOLPH2 RNA or cDNA; or
- (C) a first amplification oligonucleotide comprises a sequence that hybridizes specifically to a 5' portion of a chimeric RNA transcribed from a TMPRSS2 gene or its corresponding cDNA and a second amplification oligonucleotide comprises a sequence that hybridizes specifically to a 3' portion of the chimeric RNA transcribed from an ERG gene or its corresponding cDNA; or
- (c) at least two antibodies comprising:
- (i) an antibody that binds specifically to SPINK1 protein; and
- (ii) at least one other antibody that binds specifically to:
- (A) GOLPH2 protein;
- (B) a native ERG protein;
- (C) an amino-terminally truncated ERG protein encoded by a fusion of a TMPRSS2 gene and an ERG gene; or
- (D) a chimeric protein having an amino-terminal portion encoded by a TMPRSS2 gene and a carboxy-terminal portion encoded by an ERG gene.

* * * * *

专利名称(译)	Spink1作为前列腺癌标志物及其用途		
公开(公告)号	US20100129804A1	公开(公告)日	2010-05-27
申请号	US12/513637	申请日	2007-11-08
[标]申请(专利权)人(译)	CHINNAIYAN ARUL中号 TOMLINS斯科特 RHODES丹尼尔 - [R 梅赫拉ROHIT		
申请(专利权)人(译)	CHINNAIYAN ARUL中号 TOMLINS斯科特 RHODES丹尼尔 - [R 梅赫拉ROHIT		
当前申请(专利权)人(译)	密歇根大学董事会		
[标]发明人	CHINNAIYAN ARUL M TOMLINS SCOTT A RHODES DANIEL R MEHRA ROHIT		
发明人	CHINNAIYAN, ARUL M. TOMLINS, SCOTT A. RHODES, DANIEL R. MEHRA, ROHIT		
IPC分类号	C12Q1/68 G01N33/53		
CPC分类号	G01N33/57434 C12Q2600/178 C12Q2600/118 G01N2333/8135 C12Q2600/136 C12Q2600/112 C12Q2600/16 C12Q1/6886 C12Q2545/114 C12Q2545/101 C12Q2600/158		
优先权	60/906876 2007-03-14 US 60/857758 2006-11-08 US		
外部链接	Espacenet USPTO		

摘要(译)

提供了用于癌症研究，诊断和治疗的组合物和方法，包括但不限于癌症标志物。特别地，提供了SPINK1和用于前列腺癌的其他标志物。

Figure 1A

Meta COPA Rank	Gene	# of Studies	Avg Rank
1	ERG	7	19.3
2	SPINK1	5	28.8
3	SPRY1c	5	46
4	CIRY1	4	10
5	STY1	4	23
6	MYL2	4	28.8
7	NEB	4	27
8	TCF4	4	35.8
9	NEL2	4	35.8
10	RRT13	4	49
11	SUCSD4	4	63.3

Figure 1B

