



(86) Date de dépôt PCT/PCT Filing Date: 2012/02/16
(87) Date publication PCT/PCT Publication Date: 2012/09/20
(85) Entrée phase nationale/National Entry: 2013/09/11
(86) N° demande PCT/PCT Application No.: CA 2012/000141
(87) N° publication PCT/PCT Publication No.: 2012/122626
(30) Priorité/Priority: 2011/03/14 (US61/452,439)

(51) Cl.Int./Int.Cl. *C12Q 1/68* (2006.01),
C40B 30/04 (2006.01), *C40B 40/06* (2006.01)
(71) Demandeur/Applicant:
NATIONAL RESEARCH COUNCIL OF CANADA, CA
(72) Inventeurs/Inventors:
LI, JIE, CA;
O'CONNOR-MCCOURT, MAUREEN, CA;
WANG, EDWIN, CA
(74) Agent: PATENAUE, SONIA

(54) Titre : JEUX DE MARQUEURS DE PRONOSTIC DU CANCER DE LA PROSTATE
(54) Title: PROGNOSTIC MARKER SETS FOR PROSTATE CANCER

(57) **Abrégé/Abstract:**

Prostate cancer marker sets consisting of particular genes differentially expressed in prostate tumours provide improved accuracy of prostate cancer prognosis. The prostate cancer marker sets of the present invention, one of which consists of 30 genes related to apoptosis, one of which consists of 22 genes related to cell cycle and one of which consists of 30 genes related to response to external stimulus, may be used in a clinical setting to provide information about the likelihood of a prostate cancer patient to survive without treatment (i.e. whether the prostate tumour is "good" or "bad").



ABSTRACT

Prostate cancer marker sets consisting of particular genes differentially expressed in prostate tumours provide improved accuracy of prostate cancer prognosis. The prostate cancer marker sets of the present invention, one of which consists of 30 genes related to apoptosis, one of which consists of 22 genes related to cell cycle and one of
5 which consists of 30 genes related to response to external stimulus, may be used in a clinical setting to provide information about the likelihood of a prostate cancer patient to survive without treatment (i.e. whether the prostate tumour is "good" or "bad")

PROGNOSTIC MARKER SETS FOR PROSTATE CANCER

Cross-reference to Related Applications

This application claims the benefit of United States Provisional Patent Application
USSN 61/452,439 filed March 14, 2011, the entire contents of which is herein
5 incorporated by reference.

Field of the Invention

The present invention is related to prostate cancer, more particularly to methods
and markers for predicting prostate cancer risk.

Background of the Invention

10 There has been significant effort in the past directed to the diagnosis of prostate
cancer. The well known prostate specific antigen (PSA) test is one diagnostic test.
Another test (Belacel 2010) describes the use of eight different marker genes for
diagnosing prostate cancer. Although a variety of tests have been developed for
diagnosing prostate cancer, there have been relatively few efforts directed to developing
15 prognostic tests for predicting low-risk patients in order to determine the proper treatment
regime for patients diagnosed with prostate cancer. Two large scale studies of prostate
cancer recently showed that there is significant over-diagnosis and overtreatment of
prostate cancer patients (Andriole 2009; Schröder 2009). Many prostate cancer patients
suffer from the side effects of treatment and society is bearing the related costs. Most of
20 these treatments are unnecessary.

Recently, an algorithm (Multiple Survival Screening (MSS)) has been developed
for identifying high-quality cancer prognostic markers and this algorithm was applied for
identifying robust marker sets for breast cancer prognosis (Li 2010; Wang 2010).

25 There is a need to find new markers and develop new tests which are able to
more accurately predict low-risk patients for prostate cancer who should receive little or
no treatment.

Summary of the Invention

It has now been found that prostate cancer marker sets consisting of particular
genes differentially expressed in prostate tumours advantageously provide improved
30 accuracy of prostate cancer prognosis. The prostate cancer marker sets of the present

invention, one of which consists of 30 genes related to apoptosis, one of which consists of 22 genes related to cell cycle and one of which consists of 30 genes related to response to external stimulus, may be used in a clinical setting to provide information about the likelihood of a prostate cancer patient to survive without treatment (i.e. whether the prostate tumour is “good” or “bad”).

In one aspect of the present invention, there is provided a method of assessing likelihood of a patient having a prostate tumour benefiting from prostate cancer treatment, the method comprising: obtaining a sample of the prostate tumour or an extract thereof having message RNA therein of the patient; determining a gene expression profile of the sample for genes of a gene marker set; and, comparing the gene expression profile of the sample to standardized “good” and “bad” profiles of the marker set to determine whether the gene expression profile of the sample predicts that the tumour is “good” or “bad”, wherein “good” indicates that the patient is predicted to be at low-risk and would not likely benefit from prostate cancer treatment, “bad” indicates that the patient is predicted to be at high-risk and would likely benefit from prostate cancer treatment, and the gene marker set is Set 1, Set 2 or Set 3, wherein

Set 1 consists of apoptosis-related genes as follows:

<u>Gene</u>	<u>EntrezGene ID</u>	<u>Full Name of Gene</u>
COL4A3	1285	type IV collagen
BIRC5	332	baculoviral IAP repeat containing 5
TOP2A	7153	topoisomerase (DNA) II alpha
CDC2	983	cyclin-dependent kinase 1 (CDK1)
NRAS	4893	neuroblastoma RAS viral (v-ras) oncogene homolog
GAS1	2619	growth arrest-specific 1
LIG4	3981	ligase IV, DNA, ATP-dependent
OSM	5008	oncostatin M
PML	5371	promyelocytic leukemia
TP53	7157	tumour protein p53
NF1	4763	neurofibromin 1
SIAH1	6477	seven in absentia homolog 1 (Drosophila)
MALT1	10892	mucosa associated lymphoid tissue lymphoma translocation gene 1
KIT	3815	v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog
RHOA	387	ras homolog gene family, member A
ESR1	2099	estrogen receptor 1

RARB	5915	retinoic acid receptor, beta
VAV1	7409	vav 1 guanine nucleotide exchange factor
WRN	7486	Werner syndrome, RecQ helicase-like
TNFRSF10A	8797	tumour necrosis factor receptor superfamily, member 10a
RIPK1	8737	receptor (TNFRSF)-interacting serine-threonine kinase 1
ABL1	25	c-abl oncogene 1, non-receptor tyrosine kinase
TERT	7015	telomerase reverse transcriptase
GLI3	2737	GLI family zinc finger 3
JUN	3725	jun proto-oncogene
NFKBIA	4792	nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha
LCK	3932	lymphocyte-specific protein tyrosine kinase
CASP3	836	caspase 3, apoptosis-related cysteine peptidase
E2F2	1870	E2F transcription factor 2
LTA	4049	lymphotoxin alpha (TNF superfamily, member 1)

Set 2 consists of cell cycle-related genes as follows:

<u>Gene Name</u>	<u>EntrezGene ID</u>	<u>Description</u>
BCL2	596	B-cell CLL/lymphoma 2
RAD51	5888	RAD51 homolog (RecA homolog, E. coli) (S. cerevisiae)
CDKN2B	1030	cyclin-dependent kinase inhibitor 2B (p15, inhibits CDK4)
GML	2765	glycosylphosphatidylinositol anchored molecule like protein
E2F1	1869	E2F transcription factor 1
IKZF1	10320	IKAROS family zinc finger 1 (Ikaros)
BLM	641	Bloom syndrome, RecQ helicase-like
ABL1	25	c-abl oncogene 1, non-receptor tyrosine kinase
LIG4	3981	ligase IV, DNA, ATP-dependent
CCNA2	890	cyclin A2
NUMA1	4926	nuclear mitotic apparatus protein 1
CCNC	892	cyclin C
RBL2	5934	retinoblastoma-like 2 (p130)
LTA	4049	lymphotoxin alpha (TNF superfamily, member 1)
ERCC2	2068	excision repair cross-complementing rodent repair deficiency, complementation group 2
CASP3	836	caspase 3, apoptosis-related cysteine peptidase
TP53	7157	tumour protein p53

RAD54L	8438	RAD54-like (<i>S. cerevisiae</i>)
CCND3	896	cyclin D3
WEE1	7465	WEE1 homolog (<i>S. pombe</i>)
BIRC5	332	baculoviral IAP repeat containing 5
HDAC1	3065	histone deacetylase 1

Set 3 consists of response to external stimulus-related genes as follows:

<u>Gene Name</u>	<u>EntrezGene ID</u>	<u>Description</u>
COL4A3	1285	Type IV collagen
TOP2A	7153	topoisomerase (DNA) II alpha
CDC2	983	cyclin-dependent kinase 1 (CDK1)
LYN	4067	v-yes-1 Yamaguchi sarcoma viral related oncogene homolog
PXN	5829	paxillin
NTRK3	4916	neurotrophic tyrosine kinase, receptor, type 3
PDGFRA	5156	platelet-derived growth factor receptor, alpha polypeptide
NRAS	4893	neuroblastoma RAS viral (v-ras) oncogene homolog
CHEK1	1111	CHK1 checkpoint homolog (<i>S. pombe</i>)
PARP1	142	poly (ADP-ribose) polymerase 1
KIT	3815	v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog
TGFBR3	7049	transforming growth factor, beta receptor III
CCNA2	890	cyclin A2
NF1	4763	neurofibromin 1
MAPK10	5602	mitogen-activated protein kinase 10
CD9	928	CD9 molecule
ESR1	2099	estrogen receptor 1
FRAP1	2475	mechanistic target of rapamycin (serine/threonine kinase) (MTOR)
PML	5371	promyelocytic leukemia
ABL1	25	c-abl oncogene 1, non-receptor tyrosine kinase
TP53	7157	tumour protein p53
LIG4	3981	ligase IV, DNA, ATP-dependent
WEE1	7465	WEE1 homolog (<i>S. pombe</i>)
SYK	6850	spleen tyrosine kinase
MALT1	10892	mucosa associated lymphoid tissue lymphoma translocation gene 1
PTCH1	5727	patched 1

CASP3	836	caspase 3, apoptosis-related cysteine peptidase
BLM	641	Bloom syndrome, RecQ helicase-like
FYN	2534	FYN oncogene related to SRC, FGR, YES
WRN	7486	Werner syndrome, RecQ helicase-like

The genes in the prostate cancer marker sets of the present invention are individually known and are individually known to be differentially expressed in prostate tumour cells. How they are differentially expressed and whether their differential expression generally correlates to “good” or “bad” tumours can also be determined from publicly available datasets. However, the specific combination of the genes in each marker set of the present invention unexpectedly provides for more robust marker sets having improved prognostic accuracy for prostate cancer survival. The marker sets of the present invention consisting of the specific combination of genes that gives rise to the improved prognostic accuracy may be generated using the Multiple Survival Screening (MSS) method previously developed (Li 2010; Wang 2010).

The sample comprises a sample of the prostate tumour of the patient or an extract thereof, which contains the genes in the marker set or message RNA that hybridizes to the genes in the marker set. Preferably, the sample comprises a sample of the prostate tumour of the patient.

Preferably, all three sets are used together to make risk predictions. Thus, gene expression profiles of the sample are preferably determined for the genes in each of Sets 1, 2 and 3. In this case, the gene expression profiles are compared to standardized “good” and “bad” profiles of each respective gene marker set to determine whether each of the gene expression profiles predicts that the tumour is “good” or “bad”. If all three marker sets predict that the tumour is “good” then the patient is predicted to be at low-risk and would not likely benefit from prostate cancer treatment. If all three marker sets predict that the tumour is “bad” then the patient is predicted to be at high-risk and would likely benefit from prostate cancer treatment. If one or two of the marker sets predict that the tumour is “good” or one or two of the marker sets predict that the tumour is “bad” then the patient is predicted to be at intermediate-risk and may or may not benefit from prostate cancer treatment. Using all three marker sets improves accuracy of the prognosis.

In a particular embodiment, each gene in the gene expression profile has a gene expression value and a modified gene expression profile is obtained by multiplying the gene expression value by its marker-factor. Standardized “good” and “bad” profiles are

determined by computing standardized centroids for both “good” and “bad” classes using prediction analysis for microarrays method (Tibshirani 2002). Modified class centroids of the marker set are obtained by multiplying the standardized centroids for each class by the marker-factor. The modified gene expression profile of the sample is compared to
5 each modified class centroid to determine the tumour is “good” or “bad”. The class whose centroid is closest to the modified gene expression profile, in Pearson correlation distance, is predicted to be the class for the sample.

Gene expression profiles of a patient’s prostate tumour may be readily obtained by any number of methods known in the art, for example microarray analysis, individual
10 gene screening, etc. In a preferred embodiment, the sample is screened that against a microarray on which gene probes of the marker sets are printed. An output of the gene expression profile of the sample is preferably obtained before comparing the gene expression profile to the standardized “good” and “bad” profiles of the marker set. To obtain the output, message RNA in the sample may be hybridized to the genes on the
15 microarray, the hybridized microarray may be scanned to get all the readouts of marker genes for the sample, the readouts may be normalized and the gene expression profile of the marker set for the sample is thereby obtained. Detailed information for making microarray gene chip, scanning and normalization of array data is generally known in the art and can be found in the publicly available literature
20 (http://en.wikipedia.org/wiki/DNA_microarray). It is also possible to obtain the gene expression profile by RNA-sequencing and related sequencing technologies as these technologies become more accessible (<http://en.wikipedia.org/wiki/RNA-Seq>).

In another embodiment, kits or commercial packages are provided, which comprise gene probes for each of the genes in a gene marker set of the present invention
25 along with instructions for obtaining a gene expression profile of a sample for the gene marker set. The kit or commercial package may further comprise instructions for comparing the gene expression profile of the sample to standardized “good” and “bad” profiles of the marker set to determine whether the gene expression profile of the sample predicts that the tumour is “good” or “bad”. Preferably, the kit or commercial package
30 comprises gene probes for all three gene marker sets of the present invention. The kit or commercial package may further comprise means for obtaining a sample of a prostate tumour having message RNA therein from a patient, for example suitable syringes, fluid and/or tissue separation means, etc. In addition to the gene probes, the kit or commercial package may further comprise reagents and/or equipment useful for screening the
35 sample against the gene probes for obtaining the gene expression profile of the sample.

Various standard elements of such kits or commercial packages are generally known in the art.

Further features of the invention will be described or will become apparent in the course of the following detailed description.

5 Brief Description of the Drawings

In order that the invention may be more clearly understood, embodiments thereof will now be described in detail by way of example, with reference to the accompanying drawings, in which:

Fig. 1A provides gene names and EntrezGene ID numbers for genes in the
10 GSE10645 prostate cancer gene expression dataset which are deposited in a public database (<http://www.ncbi.nlm.nih.gov/geo/>) that belong to apoptosis GO term;

Fig. 1B provides gene names and EntrezGene ID numbers for genes in the GSE10645 prostate cancer gene expression dataset that belong to cell cycle GO term; and,

15 Fig. 1C provides gene names and EntrezGene ID numbers for genes in the GSE10645 prostate cancer gene expression dataset that belong to response to external stimulus GO term.

Description of Preferred Embodiments

Example 1: Generation of Prostate Cancer Marker Sets

20 To develop the prostate cancer marker sets of the present invention, the Multiple Survival Screening (MSS) method (Li 2010; Wang 2010) was used. In applying this method, a training set of 189 samples was selected from the GSE10645 GEO dataset (Nakagawa 2008). This prostate cancer gene expression dataset is from the population-based Swedish-Watchful Waiting cohort. The cohort consists of men with localized
25 prostate cancer (clinical stage T1-T2, Mx, N0). The GSE10645 GEO dataset contains information about genes that are differentially expressed in prostate tumours. The dataset identifies whether each of these genes is up-regulated or down-regulated in tumours and correlates these genes to patient survival (i.e. “good” vs. “bad” tumours).

30 The 189 samples from GSE10645 were randomly divided into three groups of 63 samples, each group retaining the same proportion of “good” vs. “bad” tumours that was

identified in the original GSE10645 dataset. Array-wide screening of the genes was performed on each of the three groups as described in the art (Li 2010; Wang 2010) to obtain survival genes, which are genes whose differential expression values are correlated with prostate cancer patient survivals. It is not relevant whether the expression of each gene is upregulated or downregulated so long as the differential expression is correlated to patient survival. Merging the results from each of the three groups yielded a survival gene set, which includes 133 survival genes.

Using the survival gene set, Gene Ontology (GO) analysis (using GO annotation software, David, <http://david.abcc.ncifcrf.gov/>) was performed to identify only those genes that belong to GO terms that are known to be associated with prostate cancer, such as apoptosis (cell death), cell adhesion, cell cycle, phosphorylation, response to external stimulus, cell motility and cell assembly. Table 1 lists the cancer-related GO term gene sets. One million distinct random-gene-sets were generated by randomly picking 30 genes from each cancer-related GO term gene set.

15

Table 1

<u>GO Term</u>	<u>Number of genes</u>
Apoptosis	47
Cell adhesion	68
Cell cycle	36
Phosphorylation	72
Response to external stimulus	67
Cell motility	49
Cell assembly	67

Of the 189 samples selected from the GSE10645 GEO dataset to form the training set, 36 random datasets were generated by randomly picking 60 samples from the training set while retaining in each random dataset the same proportion of “good” vs. “bad” tumours that was identified in the original GSE10645 dataset.

For a given GO term gene set, survival screening was then conducted using the 1 million random-gene-sets against all the 36 random datasets. For each random dataset, the statistical significance of the correlation between the expression values of each random-gene-set (30 genes) and patient survival status (“good” or “bad”) was examined by Kaplan–Meier analysis by implementing the Cox–Mantel log-rank test (Cui 2007). If

the P value was less than a cut-off for a survival screening using one random-gene-set against one random dataset, that random-gene-set was said to have passed. When a few thousands of random-gene-sets had passed 32 or more random datasets (the detailed parameters are shown in Table 5), the random-gene-sets that had passed were retained for further analysis. The genes in the retained random-gene-sets were then ranked based on their frequency of appearance in the passed random-gene-sets. The top 30 genes were chosen as a potential-marker-set. A similar survival screening of random-gene-sets against random datasets was performed for each of the other selected GO term gene sets.

For each GO term gene set another 1 million distinct random-gene-sets were generated and the survival screening process using the random datasets mentioned above was repeated. If the gene members for the top 30 were substantially the same as those in the potential-marker-set generated by the first screening, then the potential-marker-set is stable and can be used as a real prostate cancer marker set. If the genes for the two potential marker sets were not substantially the same, then these GO term genes are unsuitable for finding a real marker set and the potential marker set was dropped from further analysis. In some cases somewhat fewer than 30 genes may be the same in the two potential marker sets, in which case the smaller set may be designated as a marker set.

In this way, three prostate cancer marker sets were generated having stable signatures, one related to apoptosis, one related to cell cycle and one related to response to external stimulus. The genes, EntrezGene ID and full names of the genes in each of the three marker sets are given in the Tables 2-4 below. More details of each gene, including the nucleotide sequence of each gene, are known in the art and may be conveniently found in the National Center for Biotechnology Information (NCBI) Databases at <http://www.ncbi.nlm.nih.gov/>.

Table 2 - Marker Set Related to Apoptosis (30 genes)

<u>Gene</u>	<u>EntrezGene ID</u>	<u>Full Name of Gene</u>
COL4A3	1285	type IV collagen
BIRC5	332	baculoviral IAP repeat containing 5
TOP2A	7153	topoisomerase (DNA) II alpha
CDC2	983	cyclin-dependent kinase 1 (CDK1)
NRAS	4893	neuroblastoma RAS viral (v-ras) oncogene homolog
GAS1	2619	growth arrest-specific 1

LIG4	3981	ligase IV, DNA, ATP-dependent
OSM	5008	oncostatin M
PML	5371	promyelocytic leukemia
TP53	7157	tumour protein p53
NF1	4763	neurofibromin 1
SIAH1	6477	seven in absentia homolog 1 (Drosophila)
MALT1	10892	mucosa associated lymphoid tissue lymphoma translocation gene 1
KIT	3815	v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog
RHOA	387	ras homolog gene family, member A
ESR1	2099	estrogen receptor 1
RARB	5915	retinoic acid receptor, beta
VAV1	7409	vav 1 guanine nucleotide exchange factor
WRN	7486	Werner syndrome, RecQ helicase-like
TNFRSF10A	8797	tumour necrosis factor receptor superfamily, member 10a
RIPK1	8737	receptor (TNFRSF)-interacting serine-threonine kinase 1
ABL1	25	c-abl oncogene 1, non-receptor tyrosine kinase
TERT	7015	telomerase reverse transcriptase
GLI3	2737	GLI family zinc finger 3
JUN	3725	jun proto-oncogene
NFKBIA	4792	nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha
LCK	3932	lymphocyte-specific protein tyrosine kinase
CASP3	836	caspase 3, apoptosis-related cysteine peptidase
E2F2	1870	E2F transcription factor 2
LTA	4049	lymphotoxin alpha (TNF superfamily, member 1)

Table 3 - Marker Set Related to Cell Cycle (22 genes)

<u>Gene Name</u>	<u>EntrezGene ID</u>	<u>Description</u>
BCL2	596	B-cell CLL/lymphoma 2
RAD51	5888	RAD51 homolog (RecA homolog, E. coli) (S. cerevisiae)
CDKN2B	1030	cyclin-dependent kinase inhibitor 2B (p15, inhibits CDK4)
GML	2765	glycosylphosphatidylinositol anchored molecule like protein
E2F1	1869	E2F transcription factor 1
IKZF1	10320	IKAROS family zinc finger 1 (Ikaros)

BLM	641	Bloom syndrome, RecQ helicase-like
ABL1	25	c-abl oncogene 1, non-receptor tyrosine kinase
LIG4	3981	ligase IV, DNA, ATP-dependent
CCNA2	890	cyclin A2
NUMA1	4926	nuclear mitotic apparatus protein 1
CCNC	892	cyclin C
RBL2	5934	retinoblastoma-like 2 (p130)
LTA	4049	lymphotoxin alpha (TNF superfamily, member 1)
ERCC2	2068	excision repair cross-complementing rodent repair deficiency, complementation group 2
CASP3	836	caspase 3, apoptosis-related cysteine peptidase
TP53	7157	tumour protein p53
RAD54L	8438	RAD54-like (<i>S. cerevisiae</i>)
CCND3	896	cyclin D3
WEE1	7465	WEE1 homolog (<i>S. pombe</i>)
BIRC5	332	baculoviral IAP repeat containing 5
HDAC1	3065	histone deacetylase 1

Table 4 - Marker Set Related to Response to External Stimulus (30 genes)

<u>Gene Name</u>	<u>EntrezGene ID</u>	<u>Description</u>
COL4A3	1285	Type IV collagen
TOP2A	7153	topoisomerase (DNA) II alpha
CDC2	983	cyclin-dependent kinase 1 (CDK1)
LYN	4067	v-yes-1 Yamaguchi sarcoma viral related oncogene homolog
PXN	5829	paxillin
NTRK3	4916	neurotrophic tyrosine kinase, receptor, type 3
PDGFRA	5156	platelet-derived growth factor receptor, alpha polypeptide
NRAS	4893	neuroblastoma RAS viral (v-ras) oncogene homolog
CHEK1	1111	CHK1 checkpoint homolog (<i>S. pombe</i>)
PARP1	142	poly (ADP-ribose) polymerase 1
KIT	3815	v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog
TGFBR3	7049	transforming growth factor, beta receptor III
CCNA2	890	cyclin A2
NF1	4763	neurofibromin 1
MAPK10	5602	mitogen-activated protein kinase 10

CD9	928	CD9 molecule
ESR1	2099	estrogen receptor 1
FRAP1	2475	mechanistic target of rapamycin (serine/threonine kinase) (MTOR)
PML	5371	promyelocytic leukemia
ABL1	25	c-abl oncogene 1, non-receptor tyrosine kinase
TP53	7157	tumour protein p53
LIG4	3981	ligase IV, DNA, ATP-dependent
WEE1	7465	WEE1 homolog (S. pombe)
SYK	6850	spleen tyrosine kinase
MALT1	10892	mucosa associated lymphoid tissue lymphoma translocation gene 1
PTCH1	5727	patched 1
CASP3	836	caspase 3, apoptosis-related cysteine peptidase
BLM	641	Bloom syndrome, RecQ helicase-like
FYN	2534	FYN oncogene related to SRC, FGR, YES
WRN	7486	Werner syndrome, RecQ helicase-like

Table 5 - Parameters for Screening of the Marker Sets

	Number of Passed Sample Sets	Number of Passed Gene Sets	Cut-off P value
Apoptosis	32	4674	0.00001
Cell cycle	32	5548	0.0001
Response to external stimulus	35	4142	0.00001

Example 2: Validating Effectiveness of the Marker Sets in Prostate Cancer Prognosis

5 The effectiveness of the three marker sets generated in Example 1 was validated against three separate GEO datasets containing prostate cancer gene expression data from sample populations. One of the three datasets against which the markers were validated was the GSE16560 dataset described above except that 261 samples from that dataset were used. The other two test datasets were GEO datasets GSE21034 (Taylor

10 2010) and GSE10645 (Nakagawa 2008, the validation samples marked by the authors). In all three cases, test datasets were constructed by selecting samples from the GEO datasets so that the test datasets contained 90% “good” tumours and 10% “bad” tumours,

based on ultimate patient survival outcomes, in order to simulate the suggestion that over 90% of prostate cancer patients do not actually need to be treated.

To perform the validation for a given test dataset containing 'n' samples, the gene expression profile of the marker set was extracted. For each gene expression value its marker-factor was multiplied to obtain a modified gene expression profile of the testing sample. Standardized centroids were computed for both "good" and "bad" classes from n-1 samples for the marker set using the Prediction Analysis for Microarrays (PAM) method (Tibshirani 2002). The marker-factor of each gene was multiplied to the class centroids to get modified class centroids of the marker set. For predicting the recurrence of the targeted testing sample using the marker set the modified gene expression profile of the sample was compared to each of these modified class centroids. The class whose centroid that it is closest to, in Pearson correlation distance, is the predicted class for that sample. If the sample is predicted to be a "good" tumour, it is denoted as 0, otherwise it is denoted as 1. If all three marker sets predict that a particular prostate cancer sample is "good" (i.e. denoted as 0 for all 3 marker sets), the sample is assigned to low-risk group. If all three marker sets predict that a particular prostate cancer sample is "bad" (i.e. denoted as 1 for all 3 marker sets), the sample is assigned to high-risk group. If a sample is not assigned to low-risk or high-risk group, it is assigned to intermediate-risk group.

This validation process was carried out in all three of the test datasets. Table 6 shows the results for the low-risk group in comparison to the GSE16560 training set originally used to generate the three marker sets (see Example 1). As would be expected, the accuracy of the marker sets against the training set is 100%. The accuracy of the marker sets against the test datasets derived from the three GEO datasets is remarkably high.

Table 6 – Predicting Accuracy of the Marker Sets

<u>Dataset</u>	<u>No. of Samples</u>	<u>Accuracy (low-risk group)</u>
GSE10645 (training set)	189	100%
GSE16560	261	95.58%
GSE21034	140	99.31%
GSE10645 (the validation samples marked by the authors, Nakagawa 2008)	205	98.24%

The accuracy of the present marker sets can be compared to the prior art. Table 7 provides the performance of several markers and marker sets of the prior art. Table 7 is derived from Table 5 of Nakagawa 2008. The clinical models used and the nature of the various markers and marker sets listed in Table 7 below are explained in Nakagawa 2008. It is clear comparing Table 6 to Table 7 that the prognostic accuracy of the present marker sets for determining the expected survival of a prostate cancer patient is substantially greater than the prior art markers and marker sets.

Table 7 – AUC's of Prior Art Markers and Marker Sets

<u>Marker or Marker Set</u>	<u>Probes alone</u>	<u>Clinical model</u>		
		A	B	C
Clinical model alone	NA	0.736	0.757	0.783
Nakagawa 2008 - Final 17 gene/probe	0.852	0.857	0.873	0.883
Glinsky 2004 - Signature 1	0.665	0.762	0.776	0.798
Glinsky 2004 - Signature 2	0.638	0.764	0.781	0.798
Glinsky 2004 - Signature 3	0.669	0.770	0.788	0.810
Glinksy 2005	0.729	0.780	0.800	0.811
Lapointe 2004 - Tumor Recurrence Sig.	0.789	0.825	0.838	0.855
Lapointe 2004 - MUC1 and AZGP1	0.660	0.767	0.777	0.793
Singh 2002	0.783	0.824	0.838	0.851
Yu 2004	0.725	0.797	0.815	0.830

10 *References:* The contents of the entirety of each of which are incorporated by this reference.

Andriole GL, Crawford ED, Grubb III RL, et al. (2009) Mortality Results from a Randomized Prostate-Cancer Screening Trial. *The New England Journal of Medicine*. 360(13), 1310-1319.

15 Belacel N, Cuperlovic-Culf M, Ouellette R. (2010) Molecular Method for Diagnosis of Prostate Cancer. United States Patent 7,759,060 issued July 20, 2010.

Cui Q, Ma Y, Jaramillo M, Bari H, Awan A, Yang S, Zhang S, Liu L, Lu M, O'Connor-McCourt M, Purisima EO, Wang E. (2007) A map of human cancer signaling. *Molecular Systems Biology*. 3:152, 13 pages.

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

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

GO annotation software, David. <http://david.abcc.ncifcrf.gov/>.

Lapointe J, Li C, Higgins JP, van de Rijn M, Bair E, et al. (2004) Gene expression
profiling identifies clinically relevant subtypes of prostate cancer. *Proc Natl Acad Sci USA.*
10 101, 811-6.

Li J, Lenferink AEG, Deng Y, Collins C, Cui Q, Purisima EO, O'Connor-McCourt MD, Wang E. (2010) Identification of high-quality cancer prognostic markers and metastasis network modules. *Nature Communications.* 1:34, DOI: 10.1038/ncomms1033.

Nakagawa T, Kollmeyer TM, Morlan BW, et al. (2008) A Tissue Biomarker Panel
15 Predicting Systemic Progression after PSA Recurrence Post-Definitive Prostate Cancer Therapy. *PLoS one.* 3(5), e2318.

National Center for Biotechnology Information (NCBI) Databases.
<http://www.ncbi.nlm.nih.gov/>.

Sboner A, Demichelis F, Calza S, et al. (2010) Molecular Sampling of Prostate Cancer: A
20 Dilemma for Predicting Disease Progression. *BMC Medical Genomics.* 3-8. (GEO Gene Expression Omnibus GSE16560).

Schröder FH, Hugosson J, Roobol MJ, et al. (2009) Screening and Prostate-Cancer Mortality in a Randomized European Study. *The New England Journal of Medicine.* 360(13), 1320-1328.

25 Singh D, Febbo PG, Ross K, Jackson DG, Manola J, et al. (2002) Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell.* 1, 203-9.

Taylor BS, Schultz N, Hieronymus H, et al. (2010) Integrative Genomic Profiling of Human Prostate Cancer. *Cancer Cell.* 8(1), 11-22.

Tibshirani R, Hastie T, Narasimhan B, Chu G. (2002) Diagnosis of multiple cancer types
30 by shrunken centroids of gene expression. *PNAS.* 99, 6567-6572.

Wang E, Li J, Deng Y, Lenferink AEG, O'Connor-McCourt MD, Purisima EO. (2010) Process for Tumour Characteristic and Marker Set Identification, Tumour Classification and Marker Sets for Cancer. International Patent Application WO 2010/118520 published October 21, 2010.

5 Wikipedia, the free encyclopedia. (2010a) DNA Microarray. http://en.wikipedia.org/wiki/DNA_microarray.

Wikipedia, the free encyclopedia. (2010b) RNA-Seq. <http://en.wikipedia.org/wiki/RNA-Seq>.

10 Yu YP, Landsittel D, Jing L, Nelson J, Ren B, et al. (2004) Gene expression alterations in prostate cancer predicting tumour aggression and preceding development of malignancy. *J Clin Oncol*. 22, 2790-9.

Other advantages that are inherent to the structure are obvious to one skilled in the art. The embodiments are described herein illustratively and are not meant to limit the scope of the invention as claimed. Variations of the foregoing embodiments will be
15 evident to a person of ordinary skill and are intended by the inventor to be encompassed by the following claims.

Claims:

1. A method of assessing likelihood of a patient having a prostate tumour benefiting from prostate cancer treatment, the method comprising: obtaining a sample of the prostate tumour or an extract thereof having message RNA therein of the patient;
- 5 determining a gene expression profile of the sample for genes of a gene marker set; and, comparing the gene expression profile of the sample to standardized “good” and “bad” profiles of the marker set to determine whether the gene expression profile of the sample predicts that the tumour is “good” or “bad”,

wherein

- 10 “good” indicates that the patient is predicted to be at low-risk and would not likely benefit from prostate cancer treatment,

“bad” indicates that the patient is predicted to be at high-risk and would likely benefit from prostate cancer treatment, and

the gene marker set is Set 1, Set 2 or Set 3, wherein

- 15 Set 1 consists of apoptosis-related genes as follows:

<u>Gene</u>	<u>EntrezGene ID</u>	<u>Full Name of Gene</u>
COL4A3	1285	type IV collagen
BIRC5	332	baculoviral IAP repeat containing 5
TOP2A	7153	topoisomerase (DNA) II alpha
CDC2	983	cyclin-dependent kinase 1 (CDK1)
NRAS	4893	neuroblastoma RAS viral (v-ras) oncogene homolog
GAS1	2619	growth arrest-specific 1
LIG4	3981	ligase IV, DNA, ATP-dependent
OSM	5008	oncostatin M
PML	5371	promyelocytic leukemia
TP53	7157	tumour protein p53
NF1	4763	neurofibromin 1
SIAH1	6477	seven in absentia homolog 1 (Drosophila)
MALT1	10892	mucosa associated lymphoid tissue lymphoma translocation gene 1
KIT	3815	v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog
RHOA	387	ras homolog gene family, member A

ESR1	2099	estrogen receptor 1
RARB	5915	retinoic acid receptor, beta
VAV1	7409	vav 1 guanine nucleotide exchange factor
WRN	7486	Werner syndrome, RecQ helicase-like
TNFRSF10A	8797	tumour necrosis factor receptor superfamily, member 10a
RIPK1	8737	receptor (TNFRSF)-interacting serine-threonine kinase 1
ABL1	25	c-abl oncogene 1, non-receptor tyrosine kinase
TERT	7015	telomerase reverse transcriptase
GLI3	2737	GLI family zinc finger 3
JUN	3725	jun proto-oncogene
NFKBIA	4792	nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha
LCK	3932	lymphocyte-specific protein tyrosine kinase
CASP3	836	caspase 3, apoptosis-related cysteine peptidase
E2F2	1870	E2F transcription factor 2
LTA	4049	lymphotoxin alpha (TNF superfamily, member 1)

Set 2 consists of cell cycle-related genes as follows:

<u>Gene Name</u>	<u>EntrezGene ID</u>	<u>Description</u>
BCL2	596	B-cell CLL/lymphoma 2
RAD51	5888	RAD51 homolog (RecA homolog, E. coli) (S. cerevisiae)
CDKN2B	1030	cyclin-dependent kinase inhibitor 2B (p15, inhibits CDK4)
GML	2765	glycosylphosphatidylinositol anchored molecule like protein
E2F1	1869	E2F transcription factor 1
IKZF1	10320	IKAROS family zinc finger 1 (Ikaros)
BLM	641	Bloom syndrome, RecQ helicase-like
ABL1	25	c-abl oncogene 1, non-receptor tyrosine kinase
LIG4	3981	ligase IV, DNA, ATP-dependent
CCNA2	890	cyclin A2
NUMA1	4926	nuclear mitotic apparatus protein 1
CCNC	892	cyclin C
RBL2	5934	retinoblastoma-like 2 (p130)
LTA	4049	lymphotoxin alpha (TNF superfamily, member 1)
ERCC2	2068	excision repair cross-complementing rodent repair deficiency, complementation group 2
CASP3	836	caspase 3, apoptosis-related cysteine peptidase

TP53	7157	tumour protein p53
RAD54L	8438	RAD54-like (<i>S. cerevisiae</i>)
CCND3	896	cyclin D3
WEE1	7465	WEE1 homolog (<i>S. pombe</i>)
BIRC5	332	baculoviral IAP repeat containing 5
HDAC1	3065	histone deacetylase 1

Set 3 consists of response to external stimulus-related genes as follows:

<u>Gene Name</u>	<u>EntrezGene ID</u>	<u>Description</u>
COL4A3	1285	Type IV collagen
TOP2A	7153	topoisomerase (DNA) II alpha
CDC2	983	cyclin-dependent kinase 1 (CDK1)
LYN	4067	v-yes-1 Yamaguchi sarcoma viral related oncogene homolog
PXN	5829	paxillin
NTRK3	4916	neurotrophic tyrosine kinase, receptor, type 3
PDGFRA	5156	platelet-derived growth factor receptor, alpha polypeptide
NRAS	4893	neuroblastoma RAS viral (v-ras) oncogene homolog
CHEK1	1111	CHK1 checkpoint homolog (<i>S. pombe</i>)
PARP1	142	poly (ADP-ribose) polymerase 1
KIT	3815	v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog
TGFBR3	7049	transforming growth factor, beta receptor III
CCNA2	890	cyclin A2
NF1	4763	neurofibromin 1
MAPK10	5602	mitogen-activated protein kinase 10
CD9	928	CD9 molecule
ESR1	2099	estrogen receptor 1
FRAP1	2475	mechanistic target of rapamycin (serine/threonine kinase) (MTOR)
PML	5371	promyelocytic leukemia
ABL1	25	c-abl oncogene 1, non-receptor tyrosine kinase
TP53	7157	tumour protein p53
LIG4	3981	ligase IV, DNA, ATP-dependent
WEE1	7465	WEE1 homolog (<i>S. pombe</i>)
SYK	6850	spleen tyrosine kinase
MALT1	10892	mucosa associated lymphoid tissue lymphoma translocation gene 1

PTCH1	5727	patched 1
CASP3	836	caspase 3, apoptosis-related cysteine peptidase
BLM	641	Bloom syndrome, RecQ helicase-like
FYN	2534	FYN oncogene related to SRC, FGR, YES
WRN	7486	Werner syndrome, RecQ helicase-like

2. The method according to claim 1, wherein the sample comprises a sample of the prostate tumour of the patient.

3. The method according to claim 1 or 2, wherein gene expression profiles of the sample are determined for the genes in each of Sets 1, 2 and 3 and the gene expression profiles are compared to standardized “good” and “bad” profiles of each respective gene marker set to determine whether each of the gene expression profiles predicts that the tumour is “good” or “bad”, whereby if all three marker sets predict that the tumour is “good” then the patient is predicted to be at low-risk and would not likely benefit from prostate cancer treatment, if all three marker sets predict that the tumour is “bad” then the patient is predicted to be at high-risk and would likely benefit from prostate cancer treatment and if one or two of the marker sets predict that the tumour is “good” or one or two of the marker sets predict that the tumour is “bad” then the patient is predicted to be at intermediate-risk and may or may not benefit from prostate cancer treatment.

4. The method according to any one of claims 1 to 3, wherein

each gene in the gene expression profile has a gene expression value and a modified gene expression profile is obtained by multiplying the gene expression value by its marker-factor,

the standardized “good” and “bad” profiles are determined by computing standardized centroids for both “good” and “bad” classes using prediction analysis for microarrays method,

modified class centroids of the marker set are obtained by multiplying the standardized centroids for each class by the marker-factor, and

the modified gene expression profile of the sample is compared to each modified class centroid to determine the tumour is “good” or “bad”, wherein the class whose centroid is closest to the modified gene expression profile, in Pearson correlation distance, is predicted to be the class for the sample.

5. The method according to any one of claims 1 to 4, further comprising obtaining an output of the gene expression profile of the sample before comparing the gene expression profile to the standardized “good” and “bad” profiles of the marker set.
6. The method according to any one of claims 1 to 5, wherein the gene expression
5 profile of the sample is determined by screening the sample against a microarray on which gene probes of the marker set are printed.
7. Use of one or more of the gene marker sets as defined in claim 1 for predicting prostate cancer risk in a patient having a prostate tumour.
8. The use according to claim 7, wherein all three of the gene marker sets are used
10 for predicting the prostate cancer risk.
9. A kit for predicting prostate cancer risk in a patient having a prostate tumour, the kit comprising gene probes for each of the genes in a gene marker set as defined in claim 1 along with instructions for obtaining a gene expression profile of a sample for the gene marker set.
- 15 10. The kit according to claim 9 comprising gene probes for all three gene marker sets as defined in claim 1.
11. The kit according to any one of claims 9 to 10, further comprising instructions for comparing the gene expression profile of the sample to standardized “good” and “bad” profiles of the marker set to determine whether the gene expression profile of the sample
20 predicts that the tumour is “good” or “bad”.

Fig. 1B

[illegible]

Response to External Stimulus			Response to External Stimulus	
EntrezGene ID	Gene		EntrezGene ID	Gene
1285	COL4A3		4283	CXCL9
7153	TOP2A		7248	TSC1
983	CDC2		8737	RIPK1
4067	LYN		1277	COL1A1
5829	PXN		6774	STAT3
4916	NTRK3		3725	JUN
5156	PDGFRA		5378	PMS1
4893	NRAS		8438	RAD54L
1111	CHEK1		2547	XRCC6
142	PARP1		7515	XRCC1
3815	KIT		5008	OSM
7049	TGFBR3		1509	CTSD
890	CCNA2		4792	NFKBIA
4763	NF1		5888	RAD51
5602	MAPK10		4851	NOTCH1
928	CD9		6696	SPP1
2099	ESR1		1030	CDKN2B
2475	FRAP1		4968	OGG1
5371	PML		3557	IL1RN
25	ABL1		1032	CDKN2D
7157	TP53		2071	ERCC3
3981	LIG4		1462	VCAN
7465	WEE1		7472	WNT2
6850	SYK		3485	IGFBP2
10892	MALT1		405	ARNT
5727	PTCH1		999	CDH1
836	CASP3		598	BCL2L1
641	BLM		572	BAD
2534	FYN		2068	ERCC2
7486	WRN		2067	ERCC1
596	BCL2		5914	RARA
2353	FOS		7046	TGFBR1
4049	LTA		2064	ERBB2
2765	GML			

Fig. 1C