US 20030050470A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2003/0050470 A1**

An et al. (43) **Pub. Date: Mar. 13, 2003**

(54) **BIOMARKERS AND TARGETS FOR DIAGNOSIS, PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE, BLADDER AND BREAST CANCER**

(75) Inventors: **Gang An**, Oklahoma City, OK (US); **S. Mark O'Hara**, Oklahoma City, OK (US); **David Ralph**, Edmund, OK (US); **Robert Veltri**, Oklahoma City, OK (US)

Correspondence Address:
Gina N. Shishima
FULBRIGHT & JAWORSKI, L.L.P.
Ste. 1900
600 Congress Ave.
Austin, TX 78701 (US)

(73) Assignee: **UroCor, Inc.**, 800 Research Pkwy, #200, Oklahoma City, OK 73104-3699

(21) Appl. No.: **09/974,546**

(22) Filed: **Oct. 10, 2001**

Related U.S. Application Data

(60) Division of application No. 09/662,270, filed on Sep. 14, 2000, which is a division of application No. 09/097,199, filed on Jun. 12, 1998, now Pat. No. 6,218,529, which is a continuation-in-part of application No. 08/692,787, filed on Jul. 31, 1996, now Pat. No. 5,882,864.

Publication Classification

(51) **Int. Cl.**[7] **C12Q 1/68**; C07H 21/04

(52) **U.S. Cl.** **536/24.3**; 435/6

(57) **ABSTRACT**

Disclosed are diagnostic techniques for the detection of human prostate, bladder and breast cancer. Genetic probes and methods useful in monitoring the progression and diagnosis of prostate, bladder and breast cancer are described. The invention relates particularly to probes and methods for evaluating the presence of RNA species that are differentially expressed in prostate, bladder and breast cancer compared to normal human prostate, benign prostatic hyperplasia, or normal bladder or breast tissue.

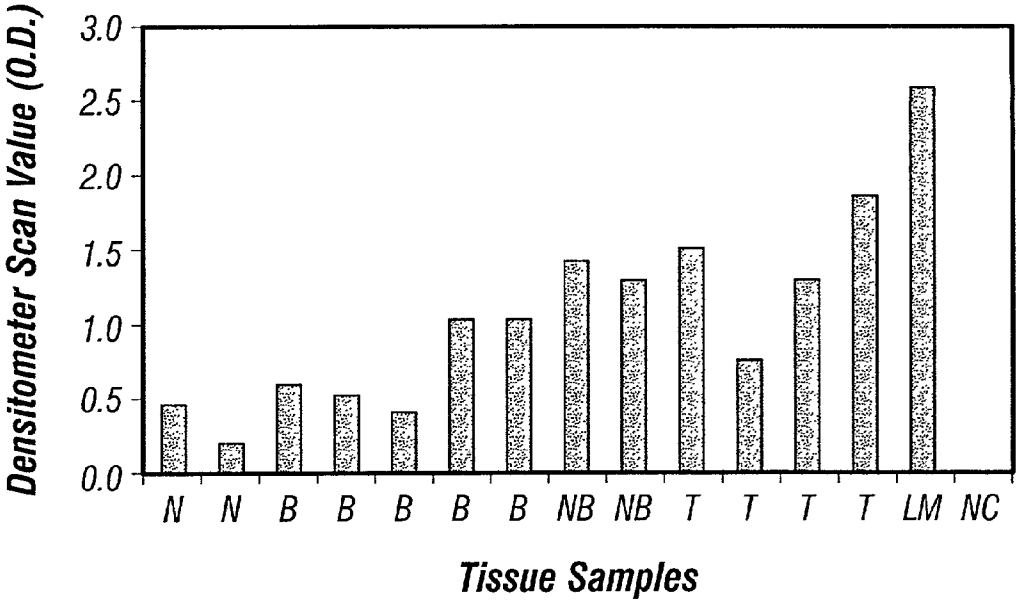
Densitometer Scan Value (O.D.)
3.0
2.5
2.0
1.5
1.0
0.5
0.0
N N B B B B B NB NB T T T T LM NC
Tissue Samples

FIG. 1

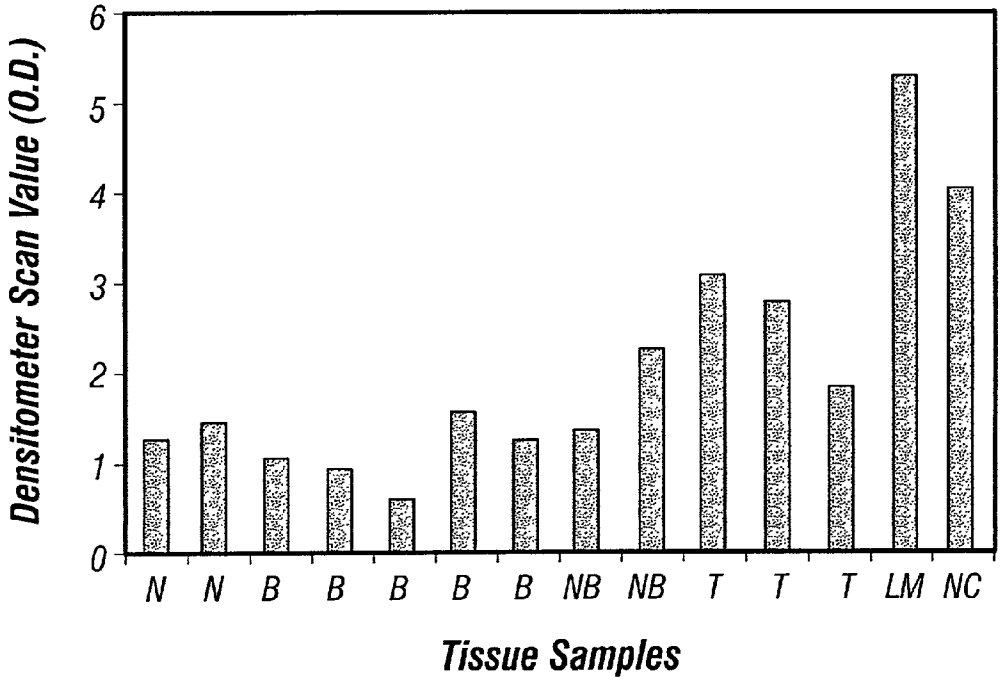
Densitometer Scan Value (O.D.)
6
5
4
3
2
1
0
N N B B B B B NB NB T T T LM NC
Tissue Samples

FIG. 2

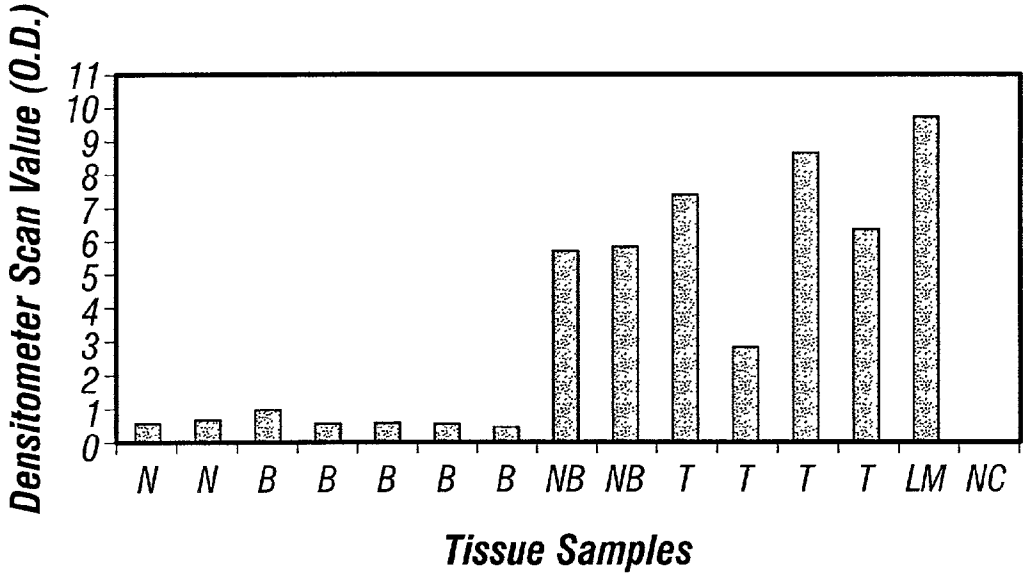
Densitometer Scan Value (O.D.)
11
10
9
8
7
6
5
4
3
2
1
0
N N B B B B B NB NB T T T T LM NC
Tissue Samples

FIG. 3

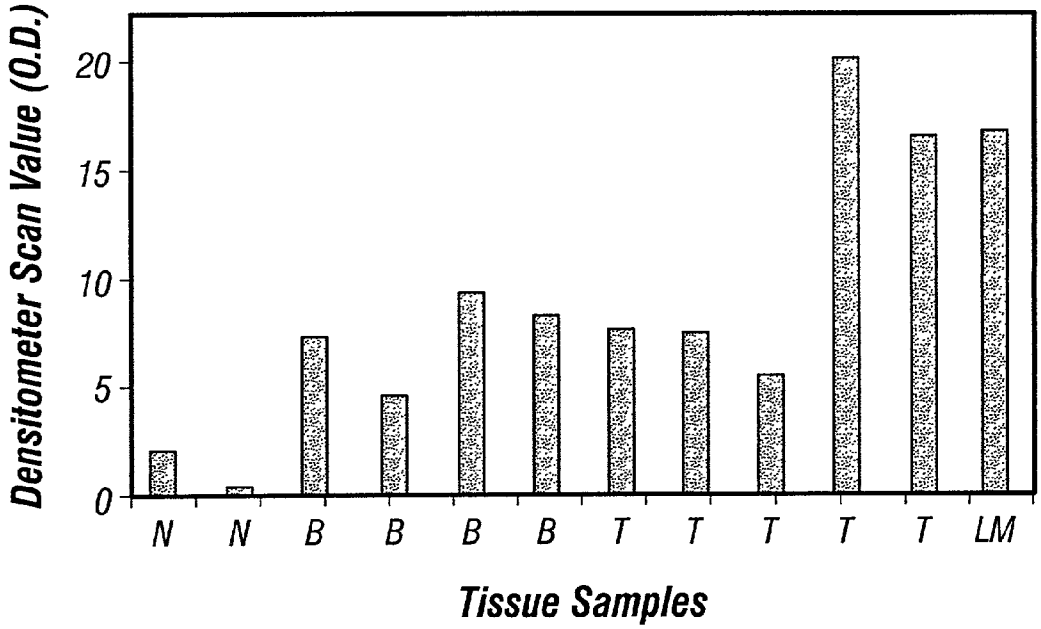
Densitometer Scan Value (O.D.)
20
15
10
5
0
N N B B B B T T T T T LM
Tissue Samples

FIG. 4

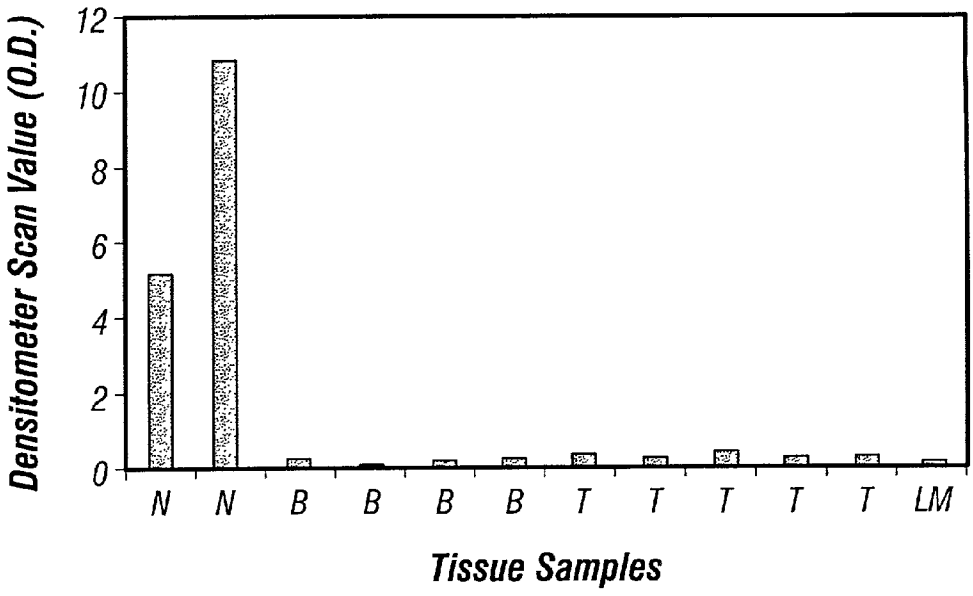
Densitometer Scan Value (O.D.)
12
10
8
6
4
2
0
N N B B B B T T T T T LM
Tissue Samples

FIG. 5

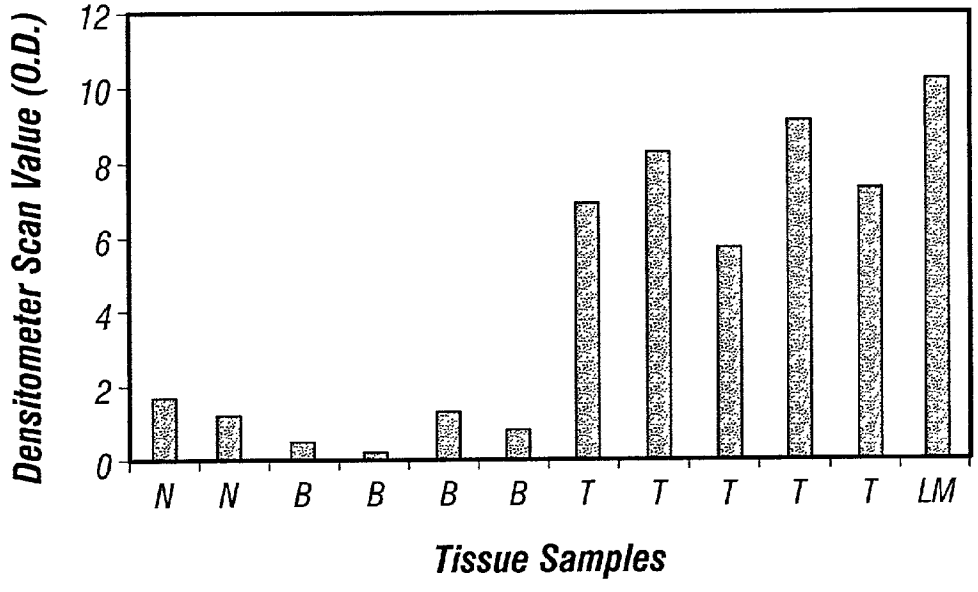
Densitometer Scan Value (O.D.)
12
10
8
6
4
2
0
N N B B B B T T T T T LM
Tissue Samples

FIG. 6

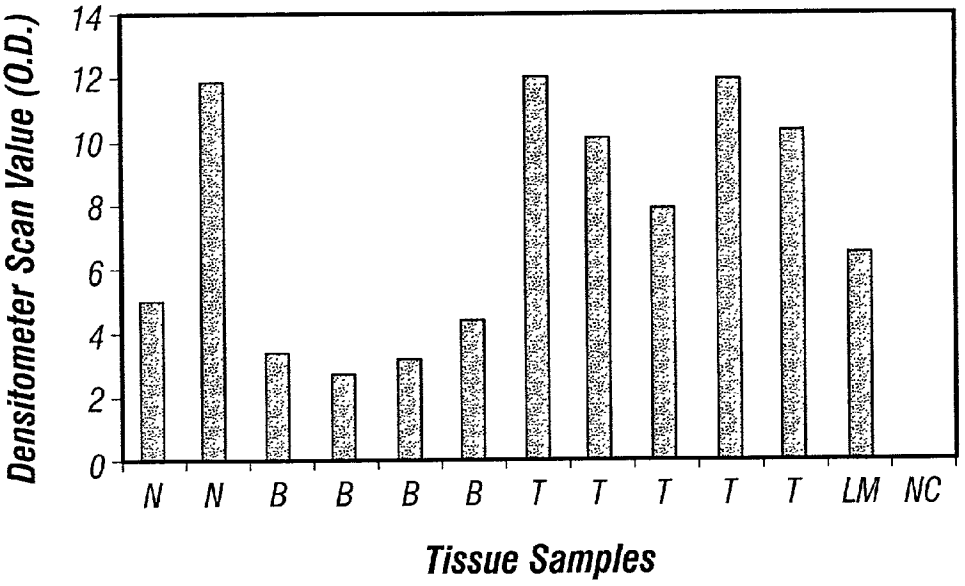
Densitometer Scan Value (O.D.)
14
12
10
8
6
4
2
0
N N B B B B T T T T T LM NC
Tissue Samples

FIG. 7

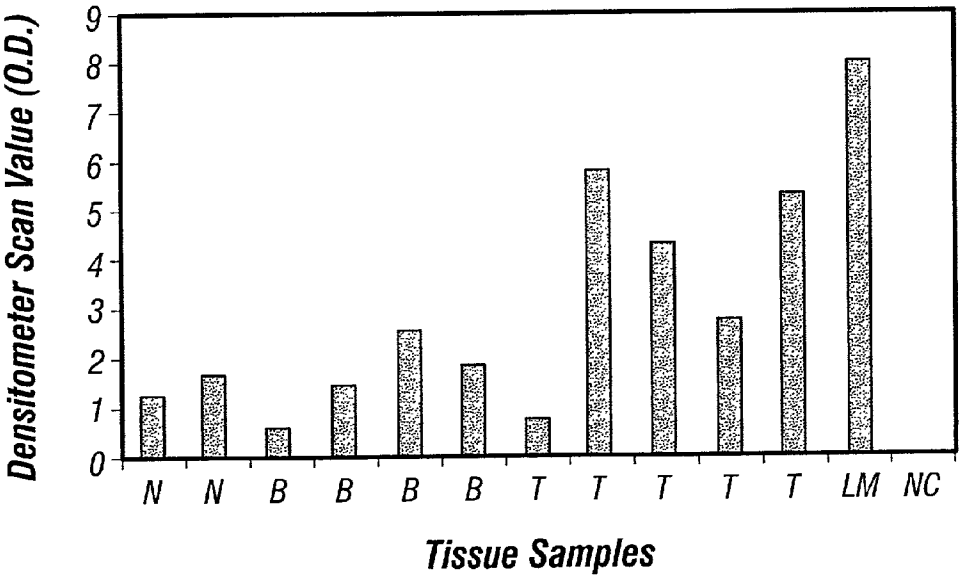
Densitometer Scan Value (O.D.)
9
8
7
6
5
4
3
2
1
0
N N B B B B T T T T T LM NC
Tissue Samples

FIG. 8

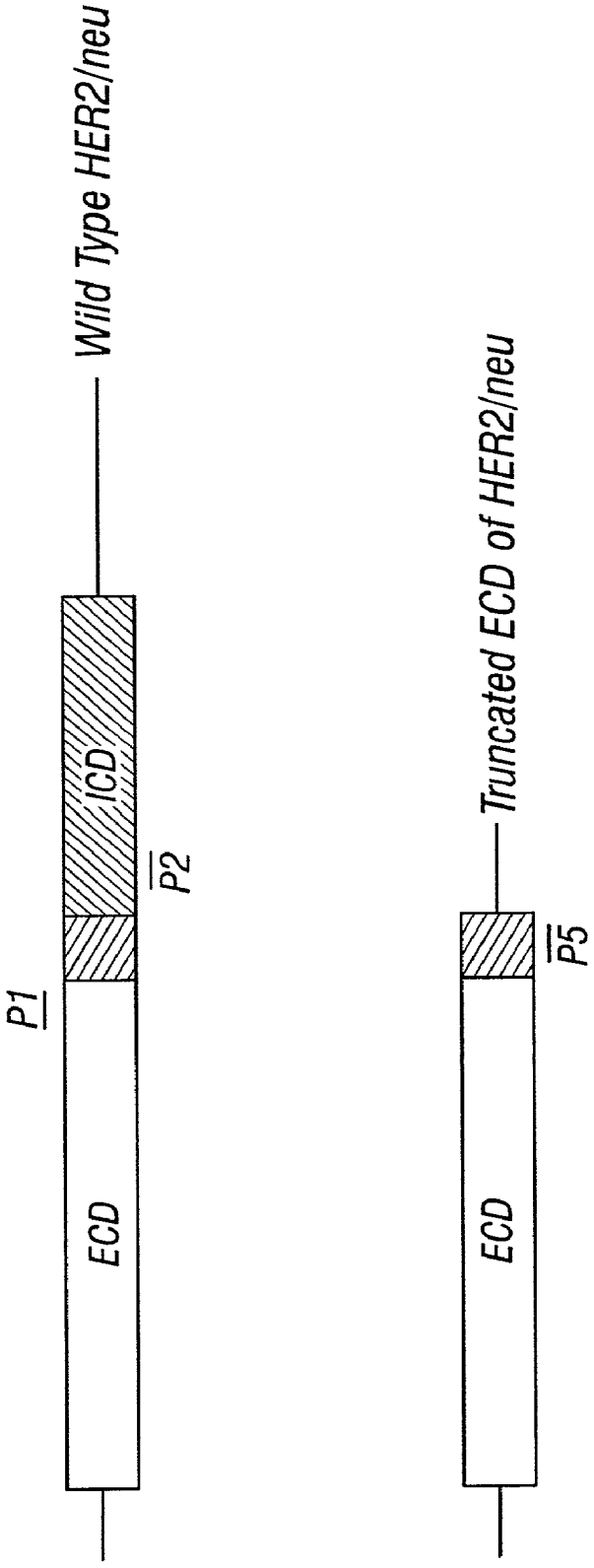
Wild Type HER2/neu
ICD
P2
P1
ECD
Truncated ECD of HER2/neu
P5
ECD

FIG. 9

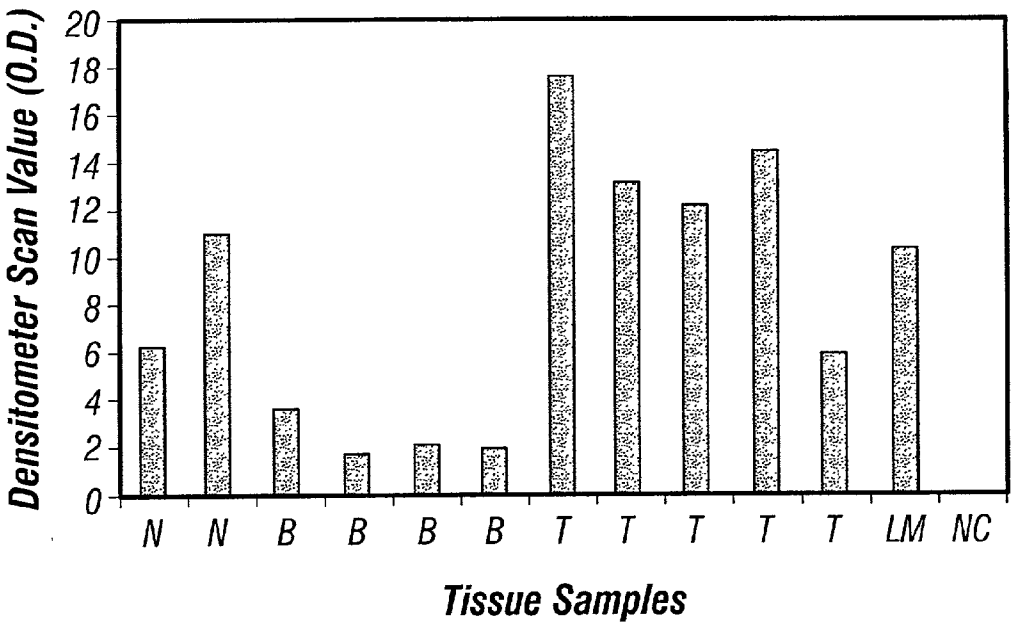
Densitometer Scan Value (O.D.)
20
18
16
14
12
10
8
6
4
2
0
N N B B B B T T T T T LM NC
Tissue Samples

FIG. 10

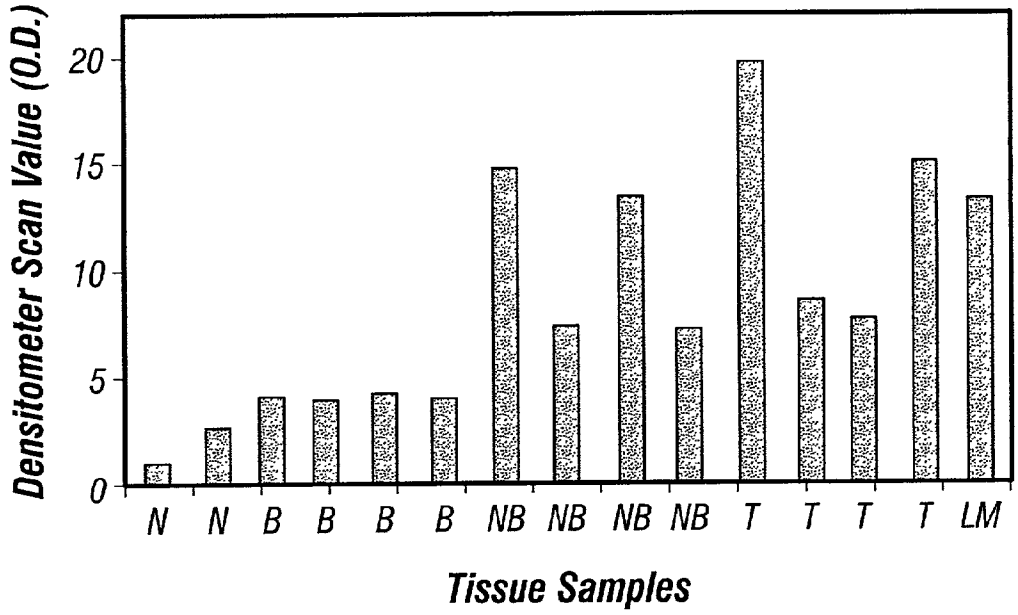
Densitometer Scan Value (O.D.)
20
15
10
5
0
N N B B B B NB NB NB NB T T T T LM
Tissue Samples

FIG. 11

β-*Actin* Normalization of cDNAs from Individual Normal Prostates, Glands with BPH, and Prostate Tumors

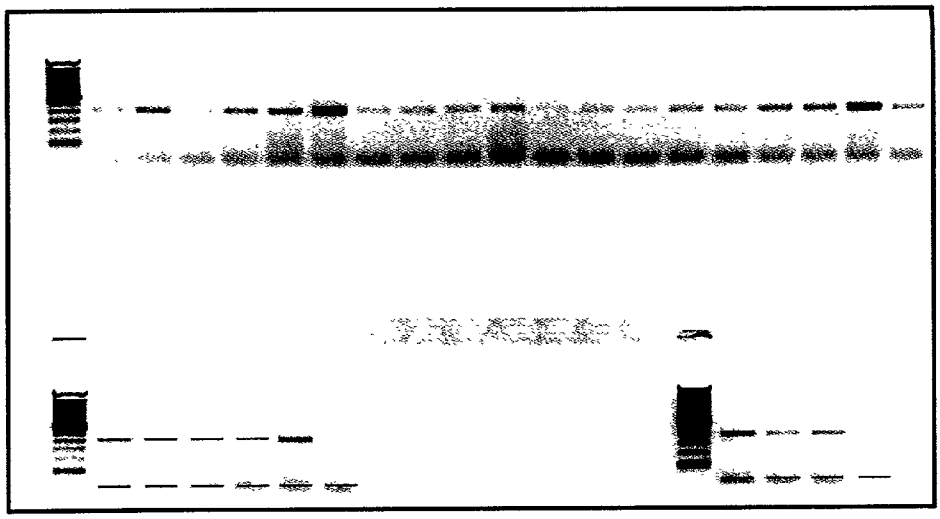

FIG. 12

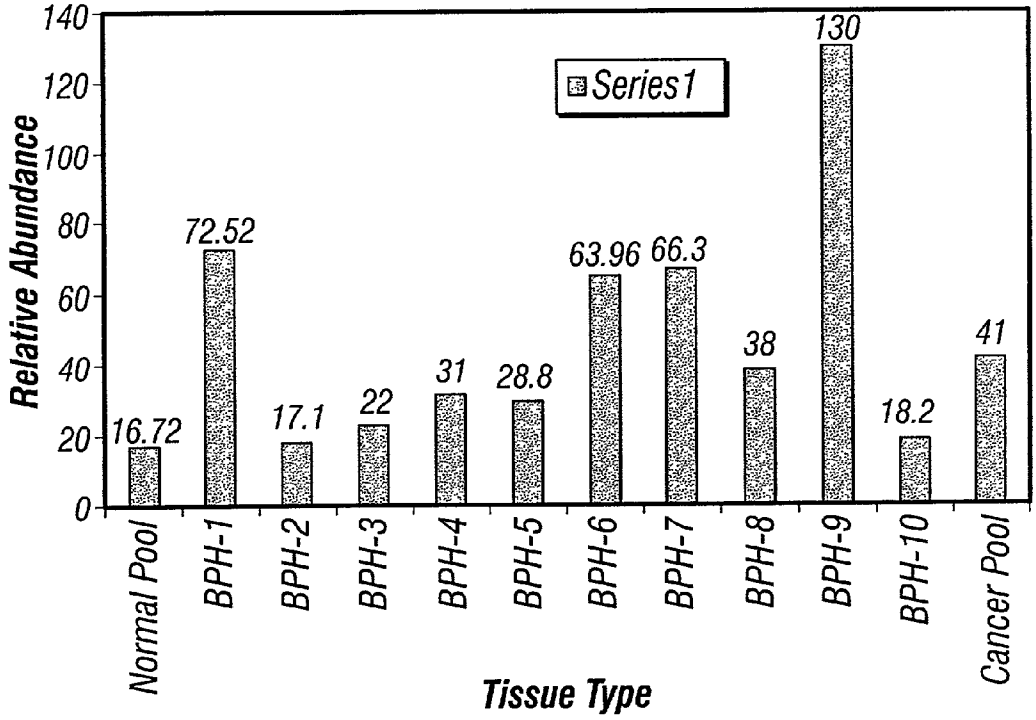

FIG. 13

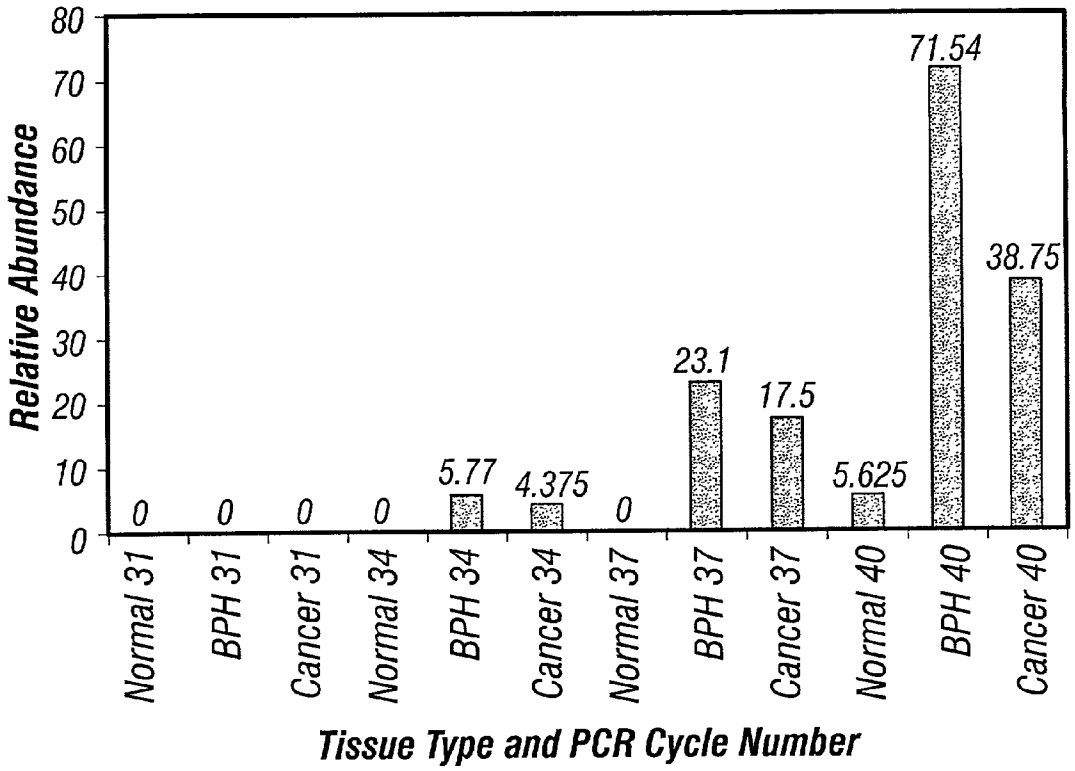
Relative Abundance
80
70
60
50
40
30
20
10
0
0
0
0
0
5.77
4.375
0
23.1
17.5
5.625
71.54
38.75
Normal 31
BPH 31
Cancer 31
Normal 34
BPH 34
Cancer 34
Normal 37
BPH 37
Cancer 37
Normal 40
BPH 40
Cancer 40
Tissue Type and PCR Cycle Number

FIG. 14

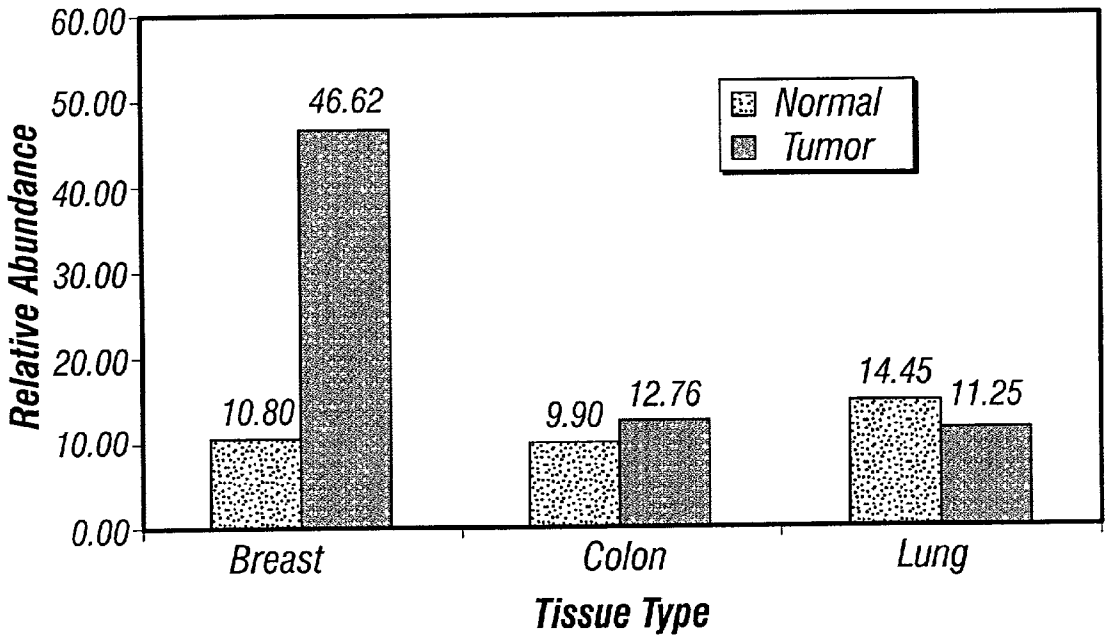
Relative Abundance
60.00
50.00
40.00
30.00
20.00
10.00
0.00
46.62
10.80
9.90
12.76
14.45
11.25
Normal
Tumor
Breast
Colon
Lung
Tissue Type

FIG. 15

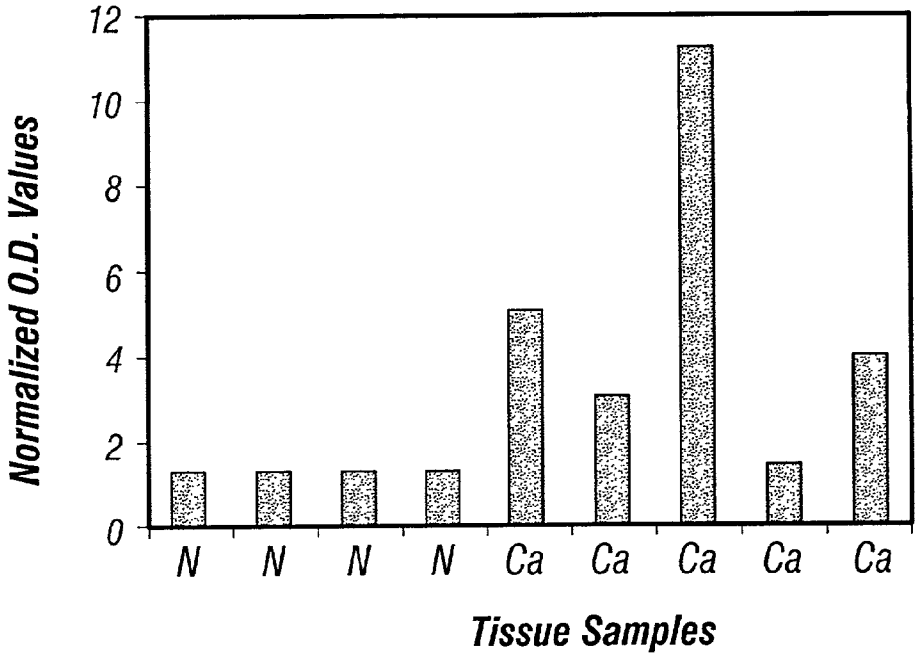
12
10
8
6
4
2
0
Normalized O.D. Values
N
N
N
N
Ca
Ca
Ca
Ca
Ca
Tissue Samples

FIG. 16

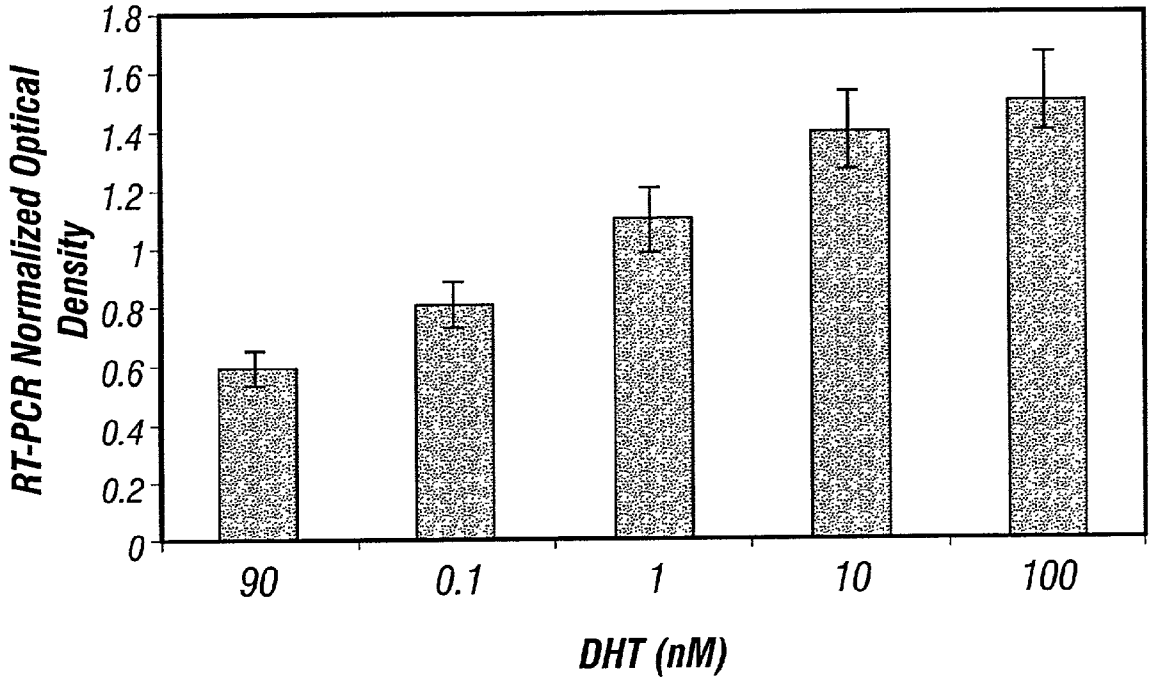
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
RT-PCR Normalized Optical Density
90
0.1
1
10
100
DHT (nM)

FIG. 17

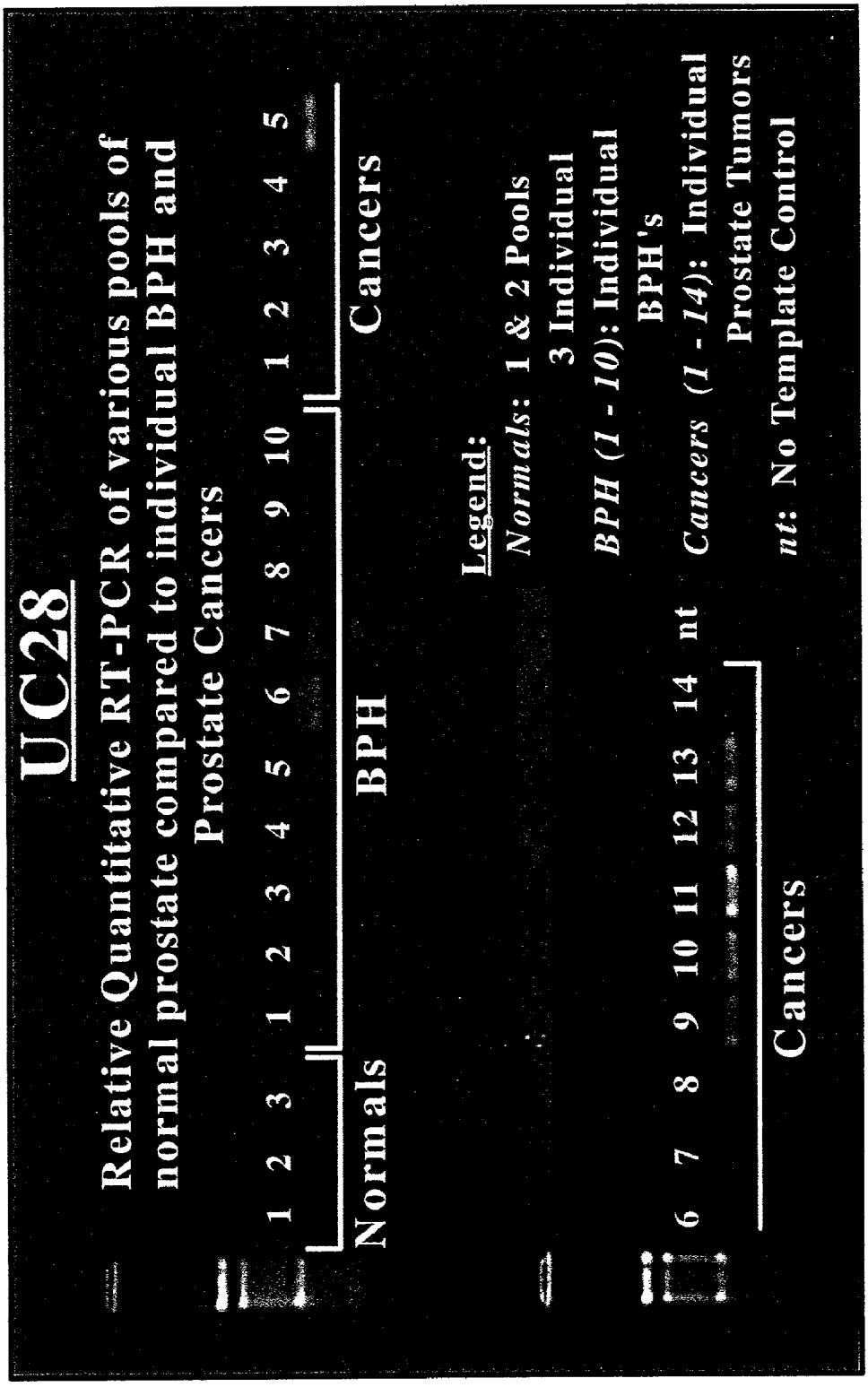
UC28
Relative Quantitative RT-PCR of various pools of normal prostate compared to individual BPH and Prostate Cancers
1 2 3 1 2 3 4 5 6 7 8 9 10 1 2 3 4 5
Normals
BPH
Cancers
6 7 8 9 10 11 12 13 14 nt
Cancers
Legend:
Normals: 1 & 2 Pools
3 Individual
BPH (1 - 10): Individual
BPH's
Cancers (1 - 14): Individual
Prostate Tumors
nt: No Template Control

FIG. 18

UC28 Human Genomic DNA Southern Blot

Experiment showing the digestion of human genomic DNA using two different restriction enzymes. Upon hybridization with the UC28 gene probe, a single band for each restriction enzyme is revealed, suggesting that there is a single copy of the UC28 gene in human genomic DNA.

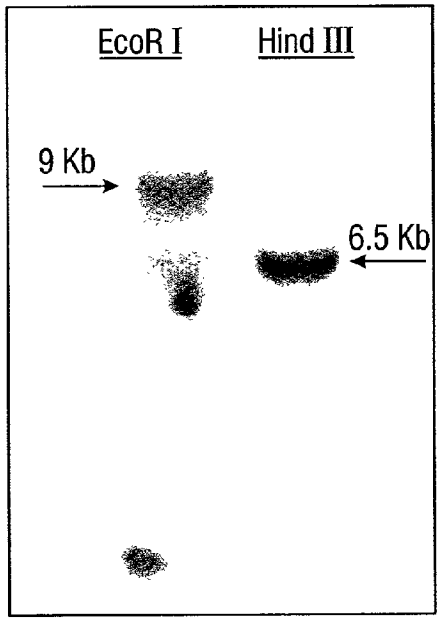

FIG. 19



FIG. 20C

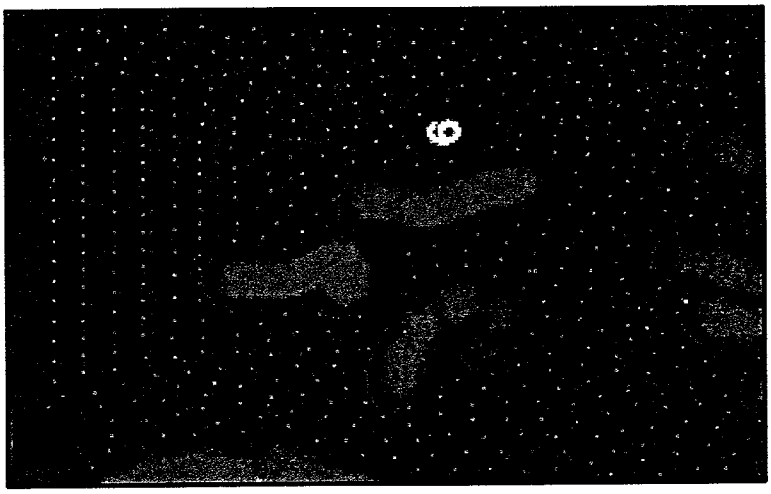

FIG. 20B

UC28 Chromosome Localization

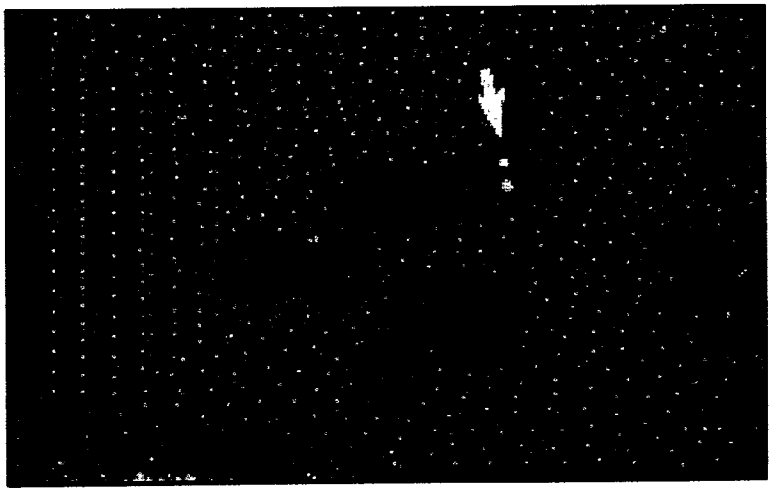

FIG. 20A

BIOMARKERS AND TARGETS FOR DIAGNOSIS, PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE, BLADDER AND BREAST CANCER

[0001] The present application is a continuation-in-part of co-pending U.S. patent application Ser. No. 08/692,787 filed Jul. 31, 1996. The entire text of the above-referenced disclosure is specifically incorporated by reference herein without disclaimer.

BACKGROUND OF THE INVENTION

[0002] A. Field of the Invention

[0003] The present invention relates generally to nucleic acid sequences useful as probes for the diagnosis of cancer and methods relating thereto. More particularly, the present invention concerns probes and methods useful in diagnosing, identifying and monitoring the progression of prostate cancer, benign prostatic hyperplasia, bladder cancer or breast cancer through measurements of gene products.

[0004] B. Description of the Related Art

[0005] Genetic detection of human disease states is a rapidly developing field (Taparowsky et al., 1982; Slamon et al., 1989; Sidransky et al., 1992; Miki et al., 1994; Dong et al., 1995; Morahan et al., 1996; Lifton, 1996; Barinaga, 1996). However, some problems exist with this approach. A number of known genetic lesions merely predispose to development of specific disease states. Individuals carrying the genetic lesion may not develop the disease state, while other individuals may develop the disease state without possessing a particular genetic lesion. In human cancers, genetic defects may potentially occur in a large number of known tumor suppresser genes and proto-oncogenes.

[0006] The genetic detection of cancer has a long history. One of the earliest genetic lesions shown to predispose to cancer was transforming point mutations in the ras oncogenes (Taparowsky et al., 1982). Transforming ras point mutations may be detected in the stool of individuals with benign and malignant colorectal tumors (Sidransky et al., 1992). However, only 50% of such tumors contained a ras mutation (Sidransky et al., 1992). Similar results have been obtained with amplification of HER-2/neu in breast and ovarian cancer (Slamon et al., 1989), deletion and mutation of p53 in bladder cancer (Sidransky et al., 1991), deletion of DCC in colorectal cancer (Fearon et al., 1990) and mutation of BRCA1 in breast and ovarian cancer (Miki et al., 1994).

[0007] None of these genetic lesions are capable of predicting a majority of individuals with cancer and most require direct sampling of a suspected tumor, making screening difficult.

[0008] Further, none of the markers described above are capable of distinguishing between metastatic and non-metastatic forms of cancer. In effective management of cancer patients, identification of those individuals whose tumors have already metastasized or are likely to metastasize is critical. Because metastatic cancer kills 560,000 people in the US each year (ACS home page), identification of markers for metastatic cancer, such as metastatic prostate and breast cancer, would be an important advance.

[0009] A particular problem in cancer detection and diagnosis occurs with prostate cancer. Carcinoma of the prostate (PCA) is the most frequently diagnosed cancer among men in the United States (Veltri et al., 1996). Prostate cancer was diagnosed in approximately 210,000 men in 1997 and about 39,000 men succumbed to the malignancy (Parker et al., 1996; Wingo et al., 1997). The American Cancer Society expects that more than 340,000 new cases of prostate cancer will be diagnosed in 1998 (Orozco et al., 1998). Although relatively few prostate tumors progress to clinical significance during the lifetime of the patient, those which are progressive in nature are likely to have metastasized by the time of detection. Survival rates for individuals with metastatic prostate cancer are quite low. Between these extremes are patients with prostate tumors that will metastasize but have not yet done so, for whom surgical prostate removal is curative. Determination of which group a patient falls within is critical in determining optimal treatment and patient survival.

[0010] The FDA approval of the serum prostate specific antigen (PSA) test in 1984 has subsequently changed the way prostate disease was managed (Allhoff et al., 1989; Cooner et al., 1990; Jacobson et al., 1995; Orozco et al., 1998). PSA is widely used as a serum biomarker to detect and monitor therapeutic response in prostate cancer patients (Badalament et al., 1996; O'Dowd et al., 1997). Several modifications in PSA assays (Partin and Oesterling, 1994; Babian et al., 1996; Zlotta et al., 1997) have resulted in earlier diagnoses and improved treatment.

[0011] While an effective indicator of prostate cancer when serum levels are relatively high, PSA serum levels are more ambiguous indicators of prostate cancer when only modestly elevated, for example when levels are between 2-10 ng/ml. At these modest elevations, serum PSA may have originated from non-cancerous disease states such as BPH (benign prostatic hyperplasia), prostatitis or physical trauma (McCormack et al., 1995). Although application of the lower 2.0 ng/ml cancer detection cutoff concentration of serum PSA has increased the diagnosis of prostate cancer, especially in younger men with non-palpable early stage tumors (Stage Tlc) (Soh et al., 1997; Carter et al., 1997; Harris et al., 1997; Orozco et al., 1998), the specificity of the PSA assay for prostate cancer detection at low serum PSA levels remains a problem.

[0012] In current clinical practice, the serum PSA assay and digital rectal exam (DRE) is used to indicate which patients should have a prostate biopsy (Lithrup et al., 1994; Orozco et al., 1998). Histological examination of the biopsied tissue is used to make the diagnosis of prostate cancer. Based upon the American Cancer Society estimate of 340,000 cases of diagnosed prostate cancer in 1998 and a known cancer detection rate of about 35% (Parker et al., 1996), it is estimated that in 1998 over one million prostate biopsies will be performed in the United States (Orozco et al., 1998). Clearly, there would be much benefit derived from a serological test that was sensitive enough to detect small and early stage prostate tumors that also had sufficient specificity to exclude a greater portion of patients with noncancerous or clinically insignificant conditions.

[0013] Several investigators have sought to improve upon the specificity of serologic detection of prostate cancer by examining a variety of other biomarkers besides serum PSA concentration (Ralph and Veltri, 1997). One of the most heavily investigated of these other biomarkers is the ratio of free versus total PSA (f/t PSA) in a patient's blood. Most

PSA in serum is in a molecular form that is bound to other proteins such as α1-antichymotrypsin (ACT) or α2-macroglobulin (Christensson et al., 1993; Stenman et al., 1991; Lilja et al., 1991). Free PSA is not bound to other proteins. The ratio of free to total PSA (f/tPSA) is usually significantly higher in patients with BPH compared to those with organ confined prostate cancer (Marley et al., 1996; Oesterling et al., 1995; Pettersson et al., 1995). When an appropriate cutoff is determined for the f/tPSA assay, the f/tPSA assay can help distinguish patients with BPH from those with prostate cancer in cases in which serum PSA levels are only modestly elevated (Marley et al., 1996; Partin and Oesterling, 1996). Unfortunately, while f/tPSA may improve on the detection of prostate cancer, information in the f/tPSA ratio is insufficient to improve the sensitivity and specificity of serologic detection of prostate cancer to desirable levels.

[0014] Genetic changes reported to be associated with prostate cancer include: allelic loss (Bova, et al., 1993; Macoska et al., 1994; Carter et al., 1990); DNA hypermethylation (Isaacs et al., 1994); point mutations or deletions of the retinoblastoma (Rb) and p53 genes (Bookstein et al., 1990a; Bookstein et al., 1990b; Isaacs et al., 1991); and aneuploidy and aneusomy of chromosomes detected by fluorescence in situ hybridization (FISH) (Macoska et al., 1994; Visakorpi et al., 1994; Takahashi et al., 1994; Alcaraz et al., 1994).

[0015] A recent development in this field was the identification of a prostate metastasis suppresser gene, KAI1 (Dong et al., 1995). Insertion of wild-type KAI1 gene into a rat prostate cancer line caused a significant decrease in metastatic tumor formation (Dong et al., 1995). However, detection of KAI1 mutations is dependent upon direct sampling of mutant prostate cells. Thus, either a primary prostate tumor must be sampled or else sufficient transformed cells must be present in blood, lymph nodes or other tissues to detect the missing or abnormal gene. Further, the presence of a deleted gene may frequently be masked by large numbers of untransformed cells that may be present in a given tissue sample.

[0016] The most commonly utilized current tests for prostate cancer are digital rectal examination (DRE) and analysis of serum prostate specific antigen (PSA). Although PSA has been widely used as a clinical marker of prostate cancer since 1988 (Partin & Oesterling, 1994), screening programs utilizing PSA alone or in combination with digital rectal examination have not been successful in improving the survival rate for men with prostate cancer (Partin & Oesterling, 1994). While PSA is specific to prostate tissue, it is produced by normal and benign as well as malignant prostatic epithelium, resulting in a high false-positive rate for prostate cancer detection (Partin & Oesterling, 1994).

[0017] Other markers that have been used for prostate cancer detection include prostatic acid phosphatase (PAP) and prostate secreted protein (PSP). PAP is secreted by prostate cells under hormonal control (Brawn et al., 1996). It has less specificity and sensitivity than does PSA. As a result, it is used much less now, although PAP may still have some applications for monitoring metastatic patients that have failed primary treatments. In general, PSP is a more sensitive biomarker than PAP, but is not as sensitive as PSA (Huang et al., 1993). Like PSA, PSP levels are frequently elevated in patients with BPH as well as those with prostate cancer.

[0018] Another serum marker associated with prostate disease is prostate specific membrane antigen (PSMA) (Horoszewicz et al., 1987; Carter et al., 1996; Murphy et al., 1996). PSMA is a Type II cell membrane protein and has been identified as Folic Acid Hydrolase (FAH) (Carter et al., 1996). Antibodies against PSMA react with both normal prostate tissue and prostate cancer tissue (Horoszewicz et al., 1987). Murphy et al. (1995) used ELISA to detect serum PSMA in advanced prostate cancer. As a serum test, PSMA levels are a relatively poor indicator of prostate cancer. However, PSMA may have utility in certain circumstances. PSMA is expressed in metastatic prostate tumor capillary beds (Silver et al., 1997) and is reported to be more abundant in the blood of metastatic cancer patients (Murphy et al., 1996). PSMA messenger RNA (mRNA) is down-regulated 8-10 fold in the LNCAP prostate cancer cell line after exposure to 5-α-dihydroxytestosterone (DHT) (Israeli et al., 1994).

[0019] A relatively new potential biomarker for prostate cancer is human kallekrein 2 (HK2) (Piironen et al., 1996). HK2 is a member of the kallekrein family that is secreted by the prostate gland. In theory, serum concentrations of HK2 may be of utility in prostate cancer detection or diagnosis, but the usefulness of this marker is still being evaluated.

[0020] As prostate cancer is one of the most prevalent forms of cancer in men, breast cancer is one of the most prevalent forms of cancer in women. Breast cancer is the leading cause of death for women between 30-50 years of age in the United States. Pathological breast cancer staging (tumor size, nodal status) is still the most reliable method for predicting outcome. In contrast to other forms of cancer, only a few tumor markers have been identified for breast cancer (e.g., estrogen receptor, progesterone receptor, S-phase, P53, Erb-2, cathepsin D) (see, e.g. Slamon et al., 1987).

[0021] Mutational analysis of important tumor suppressor genes such as p53 (Elledge, 1994) and BRCA1 (Miki et al., 1994) has recently been introduced as a diagnostic and prognostic test for breast cancer. However, many of those markers are not reliable enough to be used for routine purposes in the clinic. Two tumor suppressor genes that are mutated in a number of other cancers (Rb and p53) show a frequency of mutation of only about 30% in breast cancer (Cox et al., 1994). Mutations in the recently identified breast cancer susceptibility genes BRCA1 (chromosome 17q21) and BRCA2 (chromosome 13q13) are associated with familial breast cancer, accounting for about 5% of total breast cancer cases, but have not been found in sporadic breast cancer (Stratton and Wooster, 1996). There has yet to be found a single genetic change that accounts for the majority of sporadic breast cancers. Therefore, there is an urgent need for better prognostic markers in breast cancer diagnosis.

[0022] It is known that the processes of transformation and tumor progression are associated with changes in the levels of messenger RNA species (Slamonet al., 1984; Sager et al., 1993; Mok et al., 1994; Watson et al., 1994). Recently, a variation on PCR analysis known as RNA fingerprinting has been used to identify messages differentially expressed in ovarian or breast carcinomas (Liang et al., 1992; Sager et al., 1993; Mok et al., 1994; Watson et al., 1994). By using arbitrary primers to generate "fingerprints" from total cell RNA, followed by separation of the amplified fragments by

high resolution gel electrophoresis, it is possible to identify RNA species that are either up-regulated or down-regulated in cancer cells. Results of these studies indicated the presence of several markers of potential utility for diagnosis of breast or ovarian cancer, including a6-integrin (Sager et al., 1993), DEST001 and DEST002 (Watson et al., 1994), and LF4.0 (Mok et al., 1994).

[0023] There remain, however, deficiencies in the prior art with respect to the identification of the genes linked with the progression of prostate, bladder or breast cancer and the development of diagnostic methods to monitor disease progression. Likewise, the identification of genes which are differentially expressed in prostate, bladder, breast and other forms of cancer would be of considerable importance in the development of a rapid, inexpensive method to diagnose cancer.

SUMMARY OF THE INVENTION

[0024] The present invention addresses deficiencies in the prior art by identifying and characterizing RNA species that are differentially expressed in human prostate disease, bladder cancer or breast cancer, along with providing methods for identifying such RNA species. These RNA species and the corresponding encoded protein species have utility, for example, as markers of prostate cancer, benign prostatic hyperplasia (BPH), bladder cancer or breast cancer, and as targets for therapeutic intervention in prostate cancer, BPH, bladder cancer or breast cancer. The disclosed methods may also be applied to other tissues in order to identify differentially expressed genes that are markers of different physiological states of that tissue.

[0025] The identified markers of prostate cancer, BPH, bladder cancer or breast cancer can in turn be used to design specific oligonucleotide probes and primers. When used in combination with nucleic acid hybridization and amplification procedures, these probes and primers permit the rapid analysis of prostate, bladder or breast biopsy core specimens, serum samples, etc. This will assist physicians in diagnosing prostate disease, bladder cancer or breast cancer, and in determining optimal treatment courses for individuals with bladder cancer, breast cancer or with prostate tumors of varying malignancy. The same probes and primers may also be used for in situ hybridization or in situ PCR detection and diagnosis of prostate cancer, BPH, bladder cancer or breast cancer.

[0026] The identified markers of prostate cancer, BPH, bladder cancer or breast cancer may also be used to identify and isolate full length gene sequences, including regulatory elements for gene expression, from genomic human DNA libraries. The cDNA sequences identified in the present invention are first used as hybridization probes to screen genomic human DNA libraries by standard techniques. Once partial genomic clones have been identified, full-length genes are isolated by "chromosomal walking" (also called "overlap hybridization"). See, Chinault & Carbon "Overlap Hybridization Screening: Isolation and Characterization of Overlapping DNA Fragments Surrounding the LEU2 Gene on Yeast Chromosome III."*Gene* 5: 111-126, 1979. Nonrepetitive sequences at or near the ends of the partial genomic clones are then used as hybridization probes in further genomic library screening, ultimately allowing the isolation of entire gene sequences for the cancer markers of interest. Those experienced in the art will realize that full length genes may be obtained using the small expressed sequence tags (ESTs) described herein using technology currently available (Sambrook et al., 1989; Chinault & Carbon, 1979), as illustrated in Example 5 of the instant application.

[0027] The identified markers may also be used to identify and isolate cDNA sequences. In the practice of this method, the EST sequences identified in the present disclosure are used as hybridization probes to screen human cDNA libraries by standard techniques. In a preferred practice, a high quality human cDNA library is obtained from commercial or other sources. The library is plated on, for example, agarose plates containing nutrients, antibiotics and other standard ingredients. Individual colonies are transferred to nylon or nitrocellulose membranes and the EST probes are hybridized to complementary sequences on the membranes. Hybridization is detected by radioactive or enzyme-linked tags associated with the hybridized probes. Positive colonies are grown up and sequenced by, for example, dideoxy nucleotide sequencing or similar methods well known in the art. Comparison of cloned cDNA sequences with known human or animal cDNA or genomic sequences is performed using computer programs and databases well known to the skilled practitioner.

[0028] In one embodiment of the present invention, the isolated nucleic acids of the present invention are incorporated into expression vectors and expressed as the encoded proteins or peptides. Such proteins or peptides may in certain embodiments be used as antigens for induction of monoclonal or polyclonal antibody production.

[0029] One aspect of the present invention is thus, oligonucleotide hybridization probes and primers that hybridize selectively to specific markers of prostate cancer, BPH, bladder cancer or breast cancer. These probes and primers are selected from those sequences designated herein as SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:83 and SEQ ID NO:85. Probes and primers selected from those sequences designated herein as SEQ ID NO:3, SEQ ID NO:83 and SEQ ID NO:85 are preferred in hybridization to markers of prostate disease, bladder cancer or breast cancer. The availability of probes and primers specific for such unique markers provides the basis for diagnostic kits useful for distinguishing between BPH, prostate organ confined cancer and prostate tumors with potential for metastatic progression. Alternatively, the availability of probes and primers that hybridize to one or more markers of breast cancer, bladder cancer or prostate disease provide the basis for diagnostic kits useful in the detection of breast cancer, bladder cancer or prostate disease.

[0030] In one broad aspect, the present invention encompasses kits for use in detecting prostate cancer, BPH, bladder cancer or breast cancer cells in a biological sample. Such a kit may comprise one or more pairs of primers for amplifying nucleic acids corresponding to one or more prostate cancer, BPH, bladder cancer or breast cancer marker genes. The kit may further comprise samples of total mRNA

derived from tissue of various physiological states, such as normal, BPH, confined tumor and metastatically progressive tumor, for example, to be used as controls. The kit may also comprise buffers, nucleotide bases, and other compositions to be used in hybridization and/or amplification reactions. Each solution or composition may be contained in a vial or bottle and all vials held in close confinement in a box for commercial sale. Another embodiment of the present invention encompasses a kit for use in detecting prostate, bladder or breast cancer cells in a biological sample comprising oligonucleotide probes effective to bind with high affinity to markers of prostate disease, bladder cancer or breast cancer, in a Northern blot assay and containers for each of these probes. In a further embodiment, the invention encompasses a kit for use in detecting prostate cancer, BPH, bladder cancer or breast cancer cells in a biological sample comprising antibodies specific for proteins encoded by the nucleic acid markers of prostate cancer, BPH, bladder cancer or breast cancer, identified in the present invention.

[0031] In one broad aspect, the present invention encompasses methods for treating prostate cancer patients by administration of effective amounts of antibodies specific for the peptide products of prostate cancer markers identified herein, or by administration of effective amounts of vectors producing anti-sense messenger RNAs that bind to the nucleic acid products of prostate cancer markers, thereby inhibiting expression of the protein products of prostate cancer marker genes. In another broad aspect, the present invention encompasses methods for treating breast or bladder cancer patients by administration of effective amounts of antibodies specific for the peptide products of breast or bladder cancer markers identified herein, or by administration of effective amounts of vectors producing anti-sense messenger RNAs that bind to the nucleic acid products of breast or bladder cancer markers, thereby inhibiting expression of the protein products of breast or bladder cancer marker genes. Antisense nucleic acid molecules may also be provided as RNAs, as some stable forms of RNA with a long half-life that may be administered directly without the use of a vector are now known in the art. In addition, DNA constructs may be delivered to cells by liposomes, receptor mediated transfection and other methods known in the art. The method of delivery does not, in and of itself, constitute the present invention, but it is the delivery of an agent that will inhibit or disrupt expression of the targeted mRNAs as defined herein that constitute a critical step of this embodiment of the invention. Therefore, delivery of those agents, by any means known in the art would be encompassed by the present claims.

[0032] One aspect of the present invention is novel isolated nucleic acid segments that are useful as described herein as hybridization probes and primers that specifically hybridize to prostate cancer, BPH, bladder cancer or breast cancer markers. These disease markers, including both known genes and previously undescribed genes, are described herein as those mRNA species shown to be differentially expressed (either up- or down-regulated) in a prostate cancer, BPH, bladder cancer or breast cancer state as compared to a normal prostate, bladder or breast tissue. The novel isolated segments are designated herein as SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:19,SEQ ID NO:20,SEQ ID NO:21, SEQ ID NO:22,SEQ ID NO:23, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:83 and SEQ ID NO:85. The invention further comprises an isolated nucleic acid of between about 14 and about 100 bases in length, either identical to or complementary to a portion of the same length occurring within the disclosed sequences.

[0033] The present invention comprises proteins and peptides with amino acid sequences encoded by the aforementioned isolated nucleic acid segments. The invention also comprises methods for identifying biomarkers for prognostic or diagnostic assays of human prostate cancer, BPH, bladder cancer or breast cancer, using the techniques of RNA fingerprinting to identify RNAs that are differentially expressed between prostate, bladder or breast cancers versus normal or benign tissues of the same origin. Such fnger-printing techniques may utilize an oligo dT primer and an arbitrary primer, an oligo dT primer alone or random hexamers or any other method known in the art.

[0034] The invention further comprises methods for detecting prostate, bladder or breast cancer cells in biological samples, using hybridization primers and probes designed to specifically hybridize to prostate, bladder or breast cancer markers. The hybridization probes are identified and designated herein as SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:83 and SEQ ID NO:85. This method further comprises measuring the amounts of nucleic acid amplification products formed when primers selected from the designated sequences are used.

[0035] The invention further comprises the prognosis and/or diagnosis of prostate, bladder or breast cancer by measuring the amounts of nucleic acid amplification products formed as above. The invention comprises methods of treating individuals with prostate, bladder or breast cancer by providing effective amounts of antibodies and/or antisense DNA molecules which bind to the products of the above mentioned isolated nucleic acids. Preferred methods of prognosis and/or diagnosis of breast cancer utilize nucleic acid amplification products formed from the sequences designated as SEQ ID NO:3, SEQ ID NO:83 and SEQ ID NO:85 and methods of treating individuals with breast cancer comprise providing effective amounts of antibodies and/or antisense DNA molecules which bind to the products of SEQ ID NO:3, SEQ ID NO:83 and SEQ ID NO:85. The invention further comprises kits for performing the above-mentioned procedures, containing antibodies, amplification primers and/or hybridization probes.

[0036] The present invention further comprises production of antibodies specific for proteins or peptides encoded by SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:83 and SEQ ID NO:85, and the use of those antibodies for diagnostic applications in detecting prostate, bladder or breast cancer. The production of

antibodies specific for proteins or peptides encoded by SEQ ID NO:3, SEQ ID NO:83 and SEQ ID NO:85 is preferred in the use of those antibodies for diagnostic applications in detecting bladder cancer or breast cancer. The invention further comprises therapeutic treatment of prostate, bladder or breast cancer by administration of effective doses of inhibitors specific for the aforementioned encoded proteins.

BRIEF DESCRIPTION OF THE DRAWINGS

[0037] **FIG. 1**. Normalized quantitative RT-PCR of UC Band #25 (SEQ ID NO:1) shows that it is overexpressed in prostate cancers and benign prostate compared with normal prostate tissues. The levels are particularly high in metastatic prostate cancer. N=normal prostate, B=benign prostatic hyperplasia (BPH), NB=needle core biopsy of prostate cancer, T=primary prostate cancer, LM=metastatic lymph node prostate cancer, NC=negative control.

[0038] **FIG. 2**. Normalized quantitative RT-PCR of UC Band #27 (SEQ ID NO:2) shows that it is elevated in prostate cancers compared with normal or benign prostates. N=normal prostate, B=benign prostatic hyperplasia (BPH), NB=needle core biopsy of prostate cancer, T=primary prostate cancer, LM=metastatic lymph node prostate cancer, NC=negative control.

[0039] **FIG. 3**. Normalized quantitative RT-PCR of UC Band #28 (SEQ ID NO:3) shows that it is elevated in prostate cancers, particularly in metastatic prostate cancer, compared with normal or benign prostates. N=normal prostate, B=benign prostatic hyperplasia (BPH), NB=needle core biopsy of prostate cancer, T=primary prostate cancer, LM=metastatic lymph node prostate cancer, NC=negative control.

[0040] **FIG. 4**. Normalized quantitative RT-PCR of UC Band #31 (SEQ ID NO:4) shows that it is overexpressed in benign and malignant prostate compared with normal prostate. N=normal prostate, B=benign prostatic hyperplasia (BPH), NB=needle core biopsy of prostate cancer, T=primary prostate cancer, LM=metastatic lymph node prostate cancer, NC=negative control.

[0041] **FIG. 5**. Normalized quantitative RT-PCR of a sequence from the human fibronectin gene (SEQ ID NO:7) shows that it is down regulated in BPH and prostate cancer compared with normal prostate. N=normal prostate, B=benign prostatic hyperplasia (BPH), NB=needle core biopsy of prostate cancer, T=primary prostate cancer, LM=metastatic lymph node prostate cancer, NC=negative control.

[0042] **FIG. 6**. Normalized quantitative RT-PCR of UC Band #33 (SEQ ID NO:5) shows that it is overexpressed in prostate cancers compared with normal or benign prostate. N=normal prostate, B=benign prostatic hyperplasia (BPH), NB=needle core biopsy of prostate cancer, T primary prostate cancer, LM=metastatic lymph node prostate cancer, NC=negative control.

[0043] **FIG. 7**. Quantitative RT-PCR of TGF-β1 shows that it is overexpressed in prostate cancer compared to benign prostatic hyperplasia. N=normal prostate, B=benign prostatic hyperplasia (BPH), NB=needle core biopsy of prostate cancer, T=primary prostate cancer, LM=metastatic lymph node prostate cancer, NC=negative control.

[0044] **FIG. 8**. Quantitative RT-PCR of Cyclin A (SEQ ID NO:8) shows that it is overexpressed in prostate cancer compared to normal prostate and benign prostatic hyperplasia. N=normal prostate, B=benign prostatic hyperplasia (BPH), NB=needle core biopsy of prostate cancer, T=primary prostate cancer, LM=metastatic lymph node prostate cancer, NC=negative control.

[0045] **FIG. 9**. Oligonucleotides used in RT-PCR investigations of Her2/neu and a truncated form of Her2/neu. The binding sites for PCR primers are marked as P1 (Neu5'), P2 (Neu3') and P5 (Neu3'). The truncated form of Her2/neu also contains the P1 binding site. The regions within the Her2/neu coding sequence are: ECD (extracellular domain), MD (membrane domain), and ICD (intracellular domain).

[0046] **FIG. 10**. Normalized quantitative RT-PCR for the full length Her2/neu transcript shows that it is overexpressed in prostate cancers compared to normal prostate and benign prostatic hyperplasia. N=normal prostate, B=benign prostatic hyperplasia (BPH), NB=needle core biopsy of prostate cancer, T=primary prostate cancer, LM=metastatic lymph node prostate cancer, NC=negative control.

[0047] **FIG. 11**. Normalized quantitative RT-PCR for the truncated form of the Her2/neu transcript (SEQ ID NO:9) shows that it is overexpressed in prostate cancers compared to normal prostate and benign prostatic hyperplasia. N=normal prostate, B=benign prostatic hyperplasia (BPH), NB=needle core biopsy of prostate cancer, T=primary prostate cancer, LM=metastatic lymph node prostate cancer, NC=negative control.

[0048] **FIG. 12**. Amplification of β-actin cDNA from 25 cDNAs synthesized from various prostate tissues. The physiological states of these tissues, being either normal prostates, glands with BPH or prostate tumors are given in Table 2. Also shown are molecular weight markers displayed as "ladders" and three isolated bands representing the PCR products from pools of (left to right) normal, BPH and prostate cancers.

[0049] **FIG. 13** Amplification of a cDNA fragment derived from the Hek (UC205) mRNA in the individual prostate cancers described in Table 2. Many, but not all, prostate glands with BPH are seen to have higher levels of expression of Hek than seen in a pool of normal glands. Examination of a gel also indicated that some of the PCRs are not in the linear phase of their amplification curves. Data was captured on the IS1000 and normalized as described in Table 2.

[0050] **FIG. 14**. Amplification of Hek Using Pooled cDNAs normalized by β-actin. Pools of cDNAs synthesized from either normal prostates (N), prostate glands with BPH (B) or prostate tumors (C) were used as templates for β-actin cDNA amplification. Four identical sets of PCRs were set up. These were stopped and examined after differing numbers of PCR cycles. The data for the 22 cycles were numerically captured by the IS1000 and used to derive normalizing statistics. The normalizing statistics are obtained by dividing the average intensity of the three captured bands by the value of the three bands separately. These normalizing statistics were then used to normalize the data obtained from the mRNA of Hek (UC205). Hek mRNA is more abundant in the BPH and prostate cancer pools than in the pool of normal prostates. At 34 and 37 cycles, the PCRs for the BPH and cancer pools are observed in the linear phase of their amplification curves. The data was normalized to the β-actin data.

[0051] **FIG. 15**. Normalized quantitative RT-PCR of UC 28 in breast, colon, and lung cancers. UC 28 (SEQ ID NO:3) is overexpressed in breast cancer compared to normal tissue, but not overexpressed in colon and lung cancer.

[0052] **FIG. 16**. Differential expression of UC 28 in bladder cancer. Expression of UC 28 was examined in four normal bladder tissues and five bladder cancer tissues by relative quantitative RT-PCR using β-actin as a control for normalization.

[0053] **FIG. 17**. Stimulation of UC 28 gene expression by Dihydrotestosterone (DHT). LnCaP cells were cultured in RPMI-1640 with 10% charcoal stripped serum for 48 hours prior to treatment. The cells were incubated in DHT for 24 hours. RT-PCR of the UC 28 gene message showed increasing UC 28 message level in LnCaP cells incubated in increasing concentrations (0, 0.1, 1, 10, and 100 nM) of DHT. RT-PCR message was normalized to β-actin.

[0054] **FIG. 18**. Differential expression of UC 28 in prostate cancer by relative quantitative RT-PCR. Expression of UC 28 was examined in pooled samples of normal prostate tissues compared to individual BPH and prostate cancer samples.

[0055] **FIG. 19**. Determination of copy number of UC 28 by human genomic DNA Southern analyses. Genomic human DNA was digested with either Eco RI or Hind III and labeled with a UC 28 specific probe. A single band was observed with each restriction endonuclease.

[0056] **FIG. 20**. Localization of the UC 28 gene to human chromosome 6q23-24 by FISH chromosome mapping.

DETAILED DESCRIPTION OF THE INVENTION

[0057] The present invention concerns the early detection, diagnosis, prognosis and treatment of bladder or breast cancer or prostate diseases, such as prostate cancer or benign prostatic hyperplasia (BPH). Markers of prostate disease, bladder cancer or breast cancer, in the form of nucleic acid sequences isolated from human prostate tumors or prostate cancer cell lines are disclosed. These markers are indicators of malignant transformation of prostate, bladder or breast tissues and are diagnostic of the potential for metastatic spread of malignant prostate tumors.

[0058] Those skilled in the art will realize that the nucleic acid sequences disclosed herein will find utility in a variety of applications in prostate, bladder or breast cancer detection, diagnosis, prognosis and treatment. Examples of such applications within the scope of the present invention comprise amplification of one or more markers of prostate disease, bladder cancer or breast cancer, using specific primers; detection of markers of prostate disease, bladder cancer or breast cancer, by hybridization with oligonucleotide probes; incorporation of isolated nucleic acids into vectors; expression of RNA, peptides or polypeptides from the vectors; development of immunologic reagents corresponding to marker encoded products; and therapeutic treatments of prostate, bladder or breast cancer using antibodies, anti-sense nucleic acids, or other inhibitors specific for the identified prostate, bladder or breast cancer markers.

[0059] A. Nucleic Acids

[0060] As described herein, an aspect of the present disclosure is 26 markers of prostate disease, including one gene marker for prostate disease, bladder cancer and breast cancer, identified by RNA fingerprinting or quantitative RT-PCR. These include 20 previously unknown gene products, including the gene marker for prostate, bladder and breast cancer, as well as nucleic acid products of the PAP, fibronectin and cyclin A genes and a truncated nucleic acid product of the Her2/neu gene. The latter three gene products have been identified in other forms of cancer, but the present invention is the first report of overexpressionin prostate cancer.

[0061] The SEQ ID NOs corresponding to the identified markers are listed below.

UC 25	SEQ ID NO:1
UC 27	SEQ ID NO:2
UC 28	SEQ ID NO:3, SEQ ID NO:83, SEQ ID NO:85
UC 31	SEQ ID NO:4
UC 32	SEQ ID NO:7 (fibronectin)
UC 33	SEQ ID NO:5
Cyclin A	SEQ ID NO:8
truncated neu	SEQ ID NO:9
UC 38	SEQ ID NO:10
UC 40	SEQ ID NO:11
UC 41	SEQ ID NO:12
UC 43	SEQ ID NO:19
UC 47	SEQ ID NO:47 (prostatic acid phosphatase)
UC 201	SEQ ID NO:13
UC 204	SEQ ID NO:20
UC 205	SEQ ID NO:14 (Hek)
UC 207	SEQ ID NO:15
UC 209	SEQ ID NO:16
UC 210	SEQ ID NO:17
UC 211	SEQ ID NO:21
UC 212	SEQ ID NO:22
UC 213	SEQ ID NO:23
UC 214	SEQ ID NO:45
UC 215	SEQ ID NO:46

[0062] The biomarkers, primers and amino acid sequences corresponding to each SEQ ID NO are identified below (first number of each column is SEQ ID NO).

1. UC 25
2. UC 27
3. UC 28
4. UC 31
5. UC 33
6. UC 214 primer
7. UC 32 fibronectin
8. Cyclin A
9. Truncated NEU
10. UC 38
11. UC 40
12. UC 41
13. UC 201
14. UC 205 (human HEK)
15. UC 207
16. UC 209
30. UC 27 primer
31. UC 27 primer
32. UC 28 primer
33. UC 28 primer
34. UC 31 primer
35. UC 31 primer
36. UC 32 primer
37. UC 32 primer
38. UC 33 primer
39. UC 33 primer
40. β-Actin primer
41. β-Actin primer
42. 5' primer UC 28 mRNA
43. 3' primer UC 28 2.1 kb mRNA
44. NEU T3' primer
55. UC 28 ISH probe
56. UC 28 antigenic peptide
57. UC 43 primer
58. UC 43 primer
59. UC 47 primer
60. UC 47 primer
61. UC 201 primer
62. UC 201 primer
63. UC 204 primer
64. UC 204 primer
65. UC 205 primer
66. UC 205 primer
67. UC 207 primer
68. UC 207 primer
69. UC 209 primer
70. UC 209 primer

-continued

17. UC 210	45. UC 214	71. UC 210 primer
18. UC 214 primer	46. UC 215	72. UC 210 primer
19. UC 43	47. UC 47 (prostatic	73. UC 211 primer
20. UC 204	acid phosphatase)	74. UC 211 primer
21. UC 211	48. Amino acid	75. UC 212 primer
22. UC 212	sequence of UC 47	76. UC 212 primer
23. UC 213	49. UC 38 primer	77. UC 213 primer
24. UC 215 primer	50. UC 38 primer	78. UC 213 primer
25. UC 215 primer	51. UC 40 primer	79. PSA primer
26. cyclin A primer	52. UC 40 primer	80. PSA primer
27. cyclin A primer	53. UC 41 primer	81. β-Actin primer
28. UC 25 primer	54. UC 41 primer	82. β-Actin primer
29. UC 25 primer		
83. UC 28	85. UC 28	87. primer UC 28 2.5 kb
84. UC 28 amino acid sequence	86. UC 28 amino acid sequence	mRNA

[0063] In one embodiment, the nucleic acid sequences disclosed herein will find utility as hybridization probes or amplification primers. These nucleic acids may be used, for example, in diagnostic evaluation of tissue samples or employed to clone full length cDNAs or genomic clones corresponding thereto. In certain embodiments, these probes and primers consist of oligonucleotide fragments. Such fragments should be of sufficient length to provide specific hybridization to a RNA or DNA tissue sample. The sequences typically will be 10-20 nucleotides, but may be longer. Longer sequences, e.g., 40, 50, 100,500 and even up to full length, are preferred for certain embodiments.

[0064] Nucleic acid molecules having contiguous stretches of about 10, 15, 17, 20, 30, 40, 50, 60, 75 or 100 or 500 nucleotides from a sequence selected from SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:83 and SEQ ID NO:85 are contemplated. Molecules that are complementary to the above mentioned sequences and that bind to these sequences under high stringency conditions also are contemplated. These probes will be useful in a variety of hybridization embodiments, such as Southern and Northern blotting. In some cases, it is contemplated that probes may be used that hybridize to multiple target sequences without compromising their ability to effectively diagnose cancer.

[0065] Various probes and primers can be designed around the disclosed nucleotide sequences. Primers may be of any length but, typically, are 10-20 bases in length. By assigning numeric values to a sequence, for example, the first residue is 1, the second residue is 2, etc., an algorithm defining all primers can be proposed:

[0066] n to n+y

[0067] where n is an integer from 1 to the last number of the sequence and y is the length of the primer minus one (9 to 19), where n+y does not exceed the last number of the sequence. Thus, for a 10-mer, the probes correspond to bases 1 to 10, 2 to 11, 3 to 12 . . . and so on. For a 15-mer, the probes correspond to bases 1 to 15, 2 to 16, 3 to 17 . . . and so on. For a 20-mer, the probes correspond to bases 1 to 20, 2 to 21, 3 to 22 . . . and so on.

[0068] The values of n in the algorithm above for each of the nucleic acid sequences is: SEQ ID NO:1, n=391; SEQ ID NO:2, n=614; SEQ ID NO:3, n=757; SEQ ID NO:4, n=673; SEQ ID NO:5, n—358; SEQ ID NO:10, n=166; SEQ ID NO:11, n=107; SEQ ID NO:12, n=183; SEQ ID NO:13, n=92; SEQ ID NO:15, n—174; SEQ ID NO:16, n=132; SEQ ID NO:17, n—135; SEQ ID NO:19, n=471; SEQ ID NO:20, n=209, SEQ ID NO:21, n=407, SEQ ID NO:22, n=267, SEQ ID NO:23, n=333, SEQ ID NO:45, n=369, SEQ ID NO:46, n=301, SEQ ID NO:83, n=2087, SEQ ID NO:85, n=2505.

[0069] In certain embodiments, it is contemplated that multiple probes may be used for hybridization to a single sample. For example, a truncated form of Her2/neu could be detected by probing human tissue samples with oligonucleotides specific for the 5' and 3' ends of the full-length Her2/neu transcript. A full-length Her2/neu transcript would bind both probes, while a truncated form of the Her2/neu transcript, indicative of transformed cells, would bind to the 5' probe but not to the 3' probe.

[0070] The use of a hybridization probe of between 14 and 100 nucleotides in length allows the formation of a duplex molecule that is both stable and selective. Molecules having complementary sequences over stretches greater than 20 bases in length are generally preferred, in order to increase stability and selectivity of the hybrid, and thereby improve the quality and degree of particular hybrid molecules obtained. One will generally prefer to design nucleic acid molecules having stretches of 20 to 30 nucleotides, or even longer where desired. Such fragments may be readily prepared by, for example, directly synthesizing the fragment by chemical means or by introducing selected sequences into recombinant vectors for recombinant production.

[0071] Accordingly, the nucleotide sequences of the invention may be used for their ability to selectively form duplex molecules with complementary stretches of genes or RNAs or to provide primers for amplification of DNA or RNA from tissues. Depending on the application envisioned, one will desire to employ varying conditions of hybridization to achieve varying degrees of selectivity of probe towards target sequence.

[0072] For applications requiring high selectivity, one will typically desire to employ relatively stringent conditions to form the hybrids, e.g., one will select relatively low salt and/or high temperature conditions, such as provided by about 0.02 M to about 0.10 M NaCl at temperatures of about 50° C. to about 70° C. Such high stringency conditions tolerate little, if any, mismatch between the probe and the template or target strand, and would be particularly suitable for isolating specific genes or detecting specific mRNA transcripts. It is generally appreciated that conditions can be rendered more stringent by the addition of increasing amounts of formamide.

[0073] For certain applications, for example, substitution of amino acids by site-directed mutagenesis, it is appreciated that lower stringency conditions are required. Under these conditions, hybridization may occur even though the sequences of probe and target strand are not perfectly complementary, but are mismatched at one or more posi-

tions. Conditions may be rendered less stringent by increasing salt concentration and decreasing temperature. For example, a medium stringency condition could be provided by about 0.1 to 0.25 M NaCl at temperatures of about 37° C. to about 55° C., while a low stringency condition could be provided by about 0.15 M to about 0.9 M salt, at temperatures ranging from about 20° C. to about 55° C. Thus, hybridization conditions can be readily manipulated, and thus will generally be a method of choice depending on the desired results.

[0074] The following codon chart may be used, in a site-directed mutagenic scheme, to produce nucleic acids encoding the same or slightly different amino acid sequences of a given nucleic acid:

TABLE 1

Amino Acids			Codons					
Alanine	Ala	A	GCA	GCC	GCG	GCU		
Cysteine	Cys	C	UGC	UGU				
Aspartic acid	Asp	D	GAC	GAU				
Glutamic acid	Glu	E	GAA	GAG				
Phenylalanine	Phe	F	UUC	UUU				
Glycine	Gly	G	GGA	GGC	GGG	GGU		
Histidine	His	H	CAC	CAU				
Isoleucine	Ile	I	AUA	AUC	AUU			
Lysine	Lys	K	AAA	AAG				
Leucine	Leu	L	UUA	UUG	CUA	CUC	CUG	CUU
Methionine	Met	M	AUG					
Asparagine	Asn	N	AAC	AAU				
Proline	Pro	P	CCA	CCC	CCG	CCU		
Glutamine	Gln	Q	CAA	CAG				
Arginine	Arg	R	AGA	AGG	CGA	CGC	CGG	CGU
Serine	Ser	S	AGC	AGU	UCA	UCC	UCG	UCU
Threonine	Thr	T	ACA	ACC	ACG	ACU		
Valine	Val	V	GUA	GUC	GUG	GUU		
Tryptophan	Trp	W	UGG					
Tyrosine	Tyr	Y	UAC	UAU				

[0075] In other embodiments, hybridization may be achieved under conditions of, for example, 50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM $MgCl_2$, 10 mM dithiothreitol, at temperatures between approximately 20° C. to about 37° C. Other hybridization conditions utilized could include approximately 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 μM $MgCl_2$, at temperatures ranging from approximately 40° C. to about 72° C.

[0076] In certain embodiments, it will be advantageous to employ nucleic acid sequences of the present invention in combination with an appropriate means, such as a label, for determining hybridization. A wide variety of appropriate indicator means are known in the art, including fluorescent, radioactive, enzymatic or other ligands, such as avidin/biotin, which are capable of being detected. In preferred embodiments, one may desire to employ a fluorescent label or an enzyme tag such as urease, alkaline phosphatase or peroxidase, instead of radioactive or other environmentally undesirable reagents. In the case of enzyme tags, colorimetric indicator substrates are known which can be employed to provide a detection means visible to the human eye or spectrophotometrically, to identify specific hybridization with complementary nucleic acid-containing samples.

[0077] In general, it is envisioned that the hybridization probes described herein will be useful both as reagents in solution hybridization, as in PCR, for detection of expression of corresponding genes, as well as in embodiments employing a solid phase. In embodiments involving a solid phase, the test DNA (or RNA) is adsorbed or otherwise affixed to a selected matrix or surface. This fixed, single-stranded nucleic acid is then subjected to hybridization with selected probes under desired conditions. The selected conditions will depend on the particular circumstances based on the particular criteria required (depending, for example, on the G+C content, type of target nucleic acid, source of nucleic acid, size of hybridization probe, etc.). Following washing of the hybridized surface to remove non-specifically bound probe molecules, hybridization is detected, or even quantified, by means of the label.

[0078] It will be understood that this invention is not limited to the particular probes disclosed herein and particularly is intended to encompass at least nucleic acid sequences that are hybridizable to the disclosed sequences or are functional sequence analogs of these sequences. For example, a partial sequence may be used to identify a structurally-related gene or the fill length genomic or cDNA clone from which it is derived. Those of skill in the art are well aware of the methods for generating cDNA and genomic libraries which can be used as a target for the above-described probes (Sambrook et al., 1989).

[0079] For applications in which the nucleic acid segments of the present invention are incorporated into vectors, such as plasmids, cosmids or viruses, these segments may be combined with other DNA sequences, such as promoters, polyadenylation signals, restriction enzyme sites, multiple cloning sites, other coding segments, and the like, such that their overall length may vary considerably. It is contemplated that a nucleic acid fragment of almost any length may be employed, with the total length preferably being limited by the ease of preparation and use in the intended recombinant DNA protocol.

[0080] DNA segments encoding a specific gene may be introduced into recombinant host cells and employed for expressing a specific structural or regulatory protein. Alternatively, through the application of genetic engineering techniques, subportions or derivatives of selected genes may be employed. Upstream regions containing regulatory regions such as promoter regions may be isolated and subsequently employed for expression of the selected gene.

[0081] Where an expression product is to be generated, it is possible for the nucleic acid sequence to be varied while retaining the ability to encode the same product. Reference to the codon chart, provided above, will permit those of skill in the art to design any nucleic acid encoding for the product of a given nucleic acid.

[0082] B. Encoded Proteins

[0083] Once the entire coding sequence of a marker-associated gene has been determined, the gene can be inserted into an appropriate expression system. The gene can be expressed in any number of different recombinant DNA expression systems to generate large amounts of the polypeptide product, which can then be purified and used to vaccinate animals to generate antisera with which further studies may be conducted.

[0084] Examples of expression systems known to the skilled practitioner in the art include bacteria such as *E. coli*, yeast such as *Pichia pastoris*, baculovirus, and mammalian

expression systems such as in Cos or CHO cells. A complete gene can be expressed or, alternatively, fragments of the gene encoding portions of polypeptide can be produced.

[0085] In certain broad applications of the invention, the gene sequence encoding the polypeptide is analyzed to detect putative transmembrane sequences. Such sequences are typically very hydrophobic and are readily detected by the use of standard sequence analysis software, such as MacVector (IBI, New Haven, CT). The presence of transmembrane sequences is often deleterious when a recombinant protein is synthesized in many expression systems, especially *E. coli*, as it leads to the production of insoluble aggregates which are difficult to renature into the native conformation of the protein. Deletion of transmembrane sequences typically does not significantly alter the conformation of the remaining protein structure.

[0086] Moreover, transmembrane sequences, being by definition embedded within a membrane, are inaccessible. Antibodies to these sequences may not, therefore, prove useful in in vivo or in situ studies. Deletion of transmembrane-encoding sequences from the genes used for expression can be achieved by standard techniques. For example, fortuitously-placed restriction enzyme sites can be used to excise the desired gene fragment, or PCR-type amplification can be used to amplify only the desired part of the gene.

[0087] Computer sequence analysis may be used to determine the location of the predicted major antigenic determinant epitopes of the polypeptide. Software capable of carrying out this analysis is readily available commercially, for example MacVector (IBI, New Haven, Conn.). The software typically uses standard algorithms such as the Kyte/Doolittle or Hopp/Woods methods for locating hydrophilic sequences may be found on the surface of proteins and are, therefore, likely to act as antigenic determinants.

[0088] Once this analysis is made, polypeptides may be prepared which contain at least the essential features of the antigenic determinant and which may be employed in the generation of antisera against the polypeptide. Minigenes or gene fusions encoding these determinants may be constructed and inserted into expression vectors by standard methods, for example, using PCR cloning methodology.

[0089] The gene or gene fragment encoding a polypeptide may be inserted into an expression vector by standard subcloning techniques. An *E. coli* expression vector may be used which produces the recombinant polypeptide as a fusion protein, allowing rapid affinity purification of the protein. Examples of such fusion protein expression systems are the glutathione S-transferase system (Pharmacia, Piscataway, N.J.), the maltose binding protein system (NEB, Beverley, Mass.), the FLAG system (IBI, New Haven, Conn.), and the 6xHis system (Qiagen, Chatsworth, Calif.).

[0090] Some of these systems produce recombinant polypeptides bearing only a small number of additional amino acids, which are unlikely to affect the antigenic ability of the recombinant polypeptide. For example, both the FLAG system and the 6xHis system add only short sequences, both of which are known to be poorly antigenic and which do not adversely affect folding of the polypeptide to its native conformation. Other fusion systems are designed to produce fusions wherein the fusion partner is easily excised from the desired polypeptide. In one embodiment, the fusion partner is linked to the recombinant polypeptide by a peptide sequence containing a specific recognition sequence for a protease. Examples of suitable sequences are those recognized by the Tobacco Etch Virus protease (Life Technologies, Gaithersburg, Md.) or Factor Xa (New England Biolabs, Beverley, Mass.).

[0091] The expression system used may also be one driven by the baculovirus polyhedron promoter. The gene encoding the polypeptide may be manipulated by standard techniques in order to facilitate cloning into the baculovirus vector. One baculovirus vector is the pBlueBac vector (Invitrogen, Sorrento, Calif.). The vector carrying the gene for the polypeptide is transfected into *Spodoptera frugiperda* (Sf9) cells by standard protocols, and the cells are cultured and processed to produce the recombinant antigen. See Summers et al., A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures, Texas Agricultural Experimental Station; U.S. Pat. No. 4,215,051 (incorporated by reference).

[0092] As an alternative to recombinant polypeptides, synthetic peptides corresponding to the antigenic determinants may be prepared. Such peptides are at least six amino acid residues long, and may contain up to approximately 35 residues, which is the approximate upper length limit of automated peptide synthesis machines, such as those available from Applied Biosystems (Foster City, Calif.). Use of such small peptides for vaccination typically requires conjugation of the peptide to an immunogenic carrier protein such as hepatitis B surface antigen, keyhole limpet hemocyanin or bovine serum albumin. Methods for performing this conjugation are well known in the art.

[0093] Amino acid sequence variants of the polypeptide may also be prepared. These may, for instance, be minor sequence variants of the polypeptide which arise due to natural variation within the population or they may be homologues found in other species. They also may be sequences which do not occur naturally but which are sufficiently similar that they function similarly and/or elicit an immune response that cross-reacts with natural forms of the polypeptide. Sequence variants may be prepared by standard methods of site-directed mutagenesis such as those described herein for removing the transmembrane sequence.

[0094] Amino acid sequence variants of the polypeptide may be substitutional, insertional or deletion variants. Deletion variants lack one or more residues of the native protein which are not essential for function or immunogenic activity, and are exemplified by the variants lacking a transmembrane sequence. Another common type of deletion variant is one lacking secretory signal sequences or signal sequences directing a protein to bind to a particular part of a cell. An example of the latter sequence is the SH2 domain, which induces protein binding to phosphotyrosine residues.

[0095] Substitutional variants typically contain an alternative amino acid at one or more sites within the protein, and may be designed to modulate one or more properties of the polypeptide such as stability against proteolytic cleavage. Substitutions preferably are conservative, that is, one amino acid is replaced with one of similar size and charge. Conservative substitutions are well known in the art and include, for example, the changes of: alanine to serine; arginine to lysine; asparagine to glutamine or histidine; aspartate to glutamate; cysteine to serine; glutamine to asparagine; glutamate to aspartate; glycine to proline; histidine to aspar-

agine or glutamine; isoleucine to leucine or valine; leucine to valine or isoleucine; lysine to arginine, glutamine, or glutamate; methionine to leucine or isoleucine; phenylalanine to tyrosine, leucine or methionine; serine to threonine; threonine to serine; tryptophan to tyrosine; tyrosine to tryptophan or phenylalanine; and valine to isoleucine or leucine.

[0096] Insertional variants include fusion proteins such as those used to allow rapid purification of the polypeptide and also may include hybrid proteins containing sequences from other proteins and polypeptides which are homologues of the polypeptide. For example, an insertional variant may include portions of the amino acid sequence of the polypeptide from one species, together with portions of the homologous polypeptide from another species. Other insertional variants may include those in which additional amino acids are introduced within the coding sequence of the polypeptide. These typically are smaller insertions than the fusion proteins described above and are introduced, for example, to disrupt a protease cleavage site.

[0097] Major antigenic determinants of the polypeptide may be identified by an empirical approach in which portions of the gene encoding the polypeptide are expressed in a recombinant host, and the resulting proteins tested for their ability to elicit an immune response. For example, PCR may be used to prepare a range of peptides lacking successively longer fragments of the C-terminus of the protein. The immunoprotective activity of each of these peptides then identifies those fragments or domains of the polypeptide which are essential for this activity. Further studies in which only a small number of amino acids are removed at each iteration then allows the location of the antigenic determinants of the polypeptide.

[0098] Another method for the preparation of the polypeptides according to the invention is the use of peptide mimetics. Mimetics are peptide-containing molecules which mimic elements of protein secondary structure. See, for example, Johnson et al., "Peptide Turn Mimetics" in *BIOTECHNOLOGY AND PHARMACY*, Pezzuto et al., Eds., Chapman and Hall, New York (1993). The underlying rationale behind the use of peptide mimetics is that the peptide backbone of proteins exists chiefly to orient amino acid side chains in such a way as to facilitate molecular interactions, such as those of antibody and antigen. A peptide mimetic is expected to permit molecular interactions similar to the natural molecule.

[0099] Successful applications of the peptide mimetic concept have thus far focused on mimetics of β-turns within proteins, which are known to be highly antigenic. Likely β-turn structure within a polypeptide may be predicted by computer-based algorithms as discussed herein. Once the component amino acids of the turn are determined, peptide mimetics may be constructed to achieve a similar spatial orientation of the essential elements of the amino acid side chains.

[0100] C. Preparation of Antibodies Specific for Encoded Proteins

[0101] 1. Expression of Proteins from Cloned cDNAs

[0102] The cDNA species specified in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:83 and SEQ ID NO:85 may be expressed as encoded peptides or proteins. The engineering of DNA segment(s) for expression in a prokaryotic or eukaryotic system may be performed by techniques generally known to those of skill in recombinant expression. It is believed that virtually any expression system may be employed in the expression of the claimed nucleic acid sequences.

[0103] Both cDNA and genomic sequences are suitable for eukaryotic expression, as the host cell will generally process the genomic transcripts to yield functional mRNA for translation into protein. In addition, it is possible to use partial sequences for generation of antibodies against discrete portions of a gene product, even when the entire sequence of that gene product remains unknown. Computer programs are available to aid in the selection of regions which have potential immunologic significance. For example, software capable of carrying out this analysis is readily available commercially from MacVector (IBI, New Haven, Conn.). The software typically uses standard algorithms such as the Kyte/Doolittle or Hopp/Woods methods for locating hydrophilic sequences which are characteristically found on the surface of proteins and are, therefore, likely to act as antigenic determinants.

[0104] As used herein, the terms "engineered" and "recombinant" cells are intended to refer to a cell into which an exogenous DNA segment or gene, such as a cDNA or gene has been introduced through the hand of man. Therefore, engineered cells are distinguishable from naturally occurring cells which do not contain a recombinantly introduced exogenous DNA segment or gene. Recombinant cells include those having an introduced cDNA or genomic gene, and also include genes positioned adjacent to a heterologous promoter not naturally associated with the particular introduced gene.

[0105] To express a recombinant encoded protein or peptide, whether mutant or wild-type, in accordance with the present invention one would prepare an expression vector that comprises one of the claimed isolated nucleic acids under the control of, or operatively linked to, one or more promoters. To bring a coding sequence "under the control of" a promoter, one positions the 5' end of the transcription initiation site of the transcriptional reading frame generally between about 1 and about 50 nucleotides "downstream" (i.e., 3') of the chosen promoter. The "upstream" promoter stimulates transcription of the DNA and promotes expression of the encoded recombinant protein. This is the meaning of "recombinant expression" in this context.

[0106] Many standard techniques are available to construct expression vectors containing the appropriate nucleic acids and transcriptional/translational control sequences in order to achieve protein or peptide expression in a variety of host-expression systems. Cell types available for expression include, but are not limited to, bacteria, such as *E. coli* and *B. subtilis* transformed with recombinant bacteriophage DNA, plasmid DNA or cosmid DNA expression vectors.

[0107] Certain examples of prokaryotic hosts are *E. coli* strain RR1, *E. coli* LE392, *E. coli* B, *E. coli* X 1776 (ATCC

No. 31537) as well as *E. coli* W3110 (F-, lambda-, prototrophic, ATCC No. 273325); bacilli such as *Bacillus subtilis*; and other enterobacteriaceae such as *Salmonella typhimurium, Serratia marcescens*, and various Pseudomonas species.

[0108] In general, plasmid vectors containing replicon and control sequences which are derived from species compatible with the host cell are used in connection with these hosts. The vector ordinarily carries a replication site, as well as marking sequences which are capable of providing phenotypic selection in transformed cells. For example, *E. coli* is often transformed using pBR322, a plasmid derived from an *E. coli* species. pBR322 contains genes for ampicillin and tetracycline resistance and thus provides easy means for identifying transformed cells. The pBR plasmid, or other microbial plasmid or phage must also contain, or be modified to contain, promoters which may be used by the microbial organism for expression of its own proteins.

[0109] In addition, phage vectors containing replicon and control sequences that are compatible with the host microorganism may be used as transforming vectors in connection with these hosts. For example, the phage lambda GEM™-11 may be utilized in making a recombinant phage vector which may be used to transform host cells, such as *E. coli* LE392.

[0110] Further useful vectors include pIN vectors (Inouye et al., 1985); and pGEX vectors, for use in generating glutathione S-transferase (GST) soluble fusion proteins for later purification and separation or cleavage. Other suitable fusion proteins are those with β-galactosidase, ubiquitin, or the like.

[0111] Promoters that are most commonly used in recombinant DNA construction include the β-lactamase (penicillinase), lactose and tryptophan (trp) promoter systems. While these are the most commonly used, other microbial promoters have been discovered and utilized, and details concerning their nucleotide sequences have been published, enabling those of skill in the art to ligate them functionally with plasmid vectors.

[0112] For expression in Saccharomyces, the plasmid YRp7, for example, is commonly used (Stinchcomb et al., 1979; Kingsman et al., 1979; Tschemper et al., 1980). This plasmid already contains the trp1 gene which provides a selection marker for a mutant strain of yeast lacking the ability to grow in tryptophan, for example ATCC No. 44076 or PEP4-1 (Jones, 1977). The presence of the trp1 lesion as a characteristic of the yeast host cell genome then provides an effective environment for detecting transformation by growth in the absence of tryptophan.

[0113] Suitable promoting sequences in yeast vectors include the promoters for 3-phosphoglycerate kinase (Hitzeman et al., 1980) or other glycolytic enzymes (Hess et al., 1968; Holland et al., 1978), such as enolase, glyceraldehyde-3-phosphated ehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triosephosphate isomerase, phosphoglucose isomerase, and glucokinase. In constructing suitable expression plasmids, the termination sequences associated with these genes are also ligated into the expression vector 3' of the sequence desired to be expressed to provide polyadenylation of the mRNA and termination.

[0114] Other suitable promoters, which have the additional advantage of transcription controlled by growth conditions, include the promoter region for alcohol dehydrogenase 2, isocytochrome C, acid phosphatase, degradative enzymes associated with nitrogen metabolism, and the aforementioned glyceraldehyde-3-phosphate dehydrogenase, and enzymes responsible for maltose and galactose utilization.

[0115] In addition to micro-organisms, cultures of cells derived from multicellular organisms may also be used as hosts. In principle, any such cell culture is workable, whether from vertebrate or invertebrate culture. In addition to mammalian cells, these include insect cell systems infected with recombinant virus expression vectors (e.g., baculovirus); and plant cell systems infected with recombinant virus expression vectors (e.g., cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or transformed with recombinant plasmid expression vectors (e.g., Ti plasmid) containing one or more coding sequences.

[0116] In a useful insect system, *Autographa californica* nuclear polyhidrosis virus (AcNPV) is used as a vector to express foreign genes. The virus grows in *Spodoptera frugiperda* cells. The isolated nucleic acid coding sequences are cloned into non-essential regions (for example the polyhedrin gene) of the virus and placed under control of an AcNPV promoter (for example the polyhedrin promoter). Successful insertion of the coding sequences results in the inactivation of the polyhedrin gene and production of non-occluded recombinant virus (i.e., virus lacking the proteinaceous coat coded for by the polyhedrin gene). These recombinant viruses are then used to infect *Spodoptera frugiperda* cells in which the inserted gene is expressed (e.g., U.S. Pat. No. 4,215,051 (Smith)).

[0117] Examples of useful mammalian host cell lines are VERO and HeLa cells, Chinese hamster ovary (CHO) cell lines, WI 38, BHK, COS-7, 293, HepG2, 3T3, RIN and MDCK cell lines. In addition, a host cell strain may be chosen that modulates the expression of the inserted sequences, or modifies and processes the gene product in the specific fashion desired. Such modifications (e.g., glycosylation) and processing (e.g., cleavage) of protein products may be important for the function of the encoded protein.

[0118] Different host cells have characteristic and specific mechanisms for the post-translational processing and modification of proteins. Appropriate cells lines or host systems may be chosen to ensure the correct modification and processing of the foreign protein expressed. Expression vectors for use in mammalian cells ordinarily include an origin of replication (as necessary), a promoter located in front of the gene to be expressed, along with any necessary ribosome binding sites, RNA splice sites, polyadenylation site, and transcriptional terminator sequences. The origin of replication may be provided either by construction of the vector to include an exogenous origin, such as may be derived from SV40 or other viral (e.g., Polyoma, Adeno, VSV, BPV) source, or may be provided by the host cell chromosomal replication mechanism. If the vector is integrated into the host cell chromosome, the latter is often sufficient.

[0119] The promoters may be derived from the genome of mammalian cells (e.g., metallothionein promoter) or from mammalian viruses (e.g., the adenovirus late promoter; the

vaccinia virus 7.5K promoter). Further, it is also possible, and may be desirable, to utilize promoter or control sequences normally associated with the desired gene sequence, provided such control sequences are compatible with the host cell systems.

[0120] A number of viral based expression systems may be utilized, for example, commonly used promoters are derived from polyoma, Adenovirus 2, and most frequently Simian Virus 40 (SV40). The early and late promoters of SV40 virus are particularly useful because both are obtained easily from the virus as a fragment which also contains the SV40 viral origin of replication. Smaller or larger SV40 fragments may also be used, provided there is included the approximately 250 bp sequence extending from the Hind III site toward the Bgl I site located in the viral origin of replication.

[0121] In cases where an adenovirus is used as an expression vector, the coding sequences may be ligated to an adenovirus transcription/translation control complex, e.g., the late promoter and tripartite leader sequence. This chimeric gene may then be inserted in the adenovirus genome by in vitro or in vivo recombination. Insertion in a non-essential region of the viral genome (e.g., region E1 or E3) will result in a recombinant virus that is viable and capable of expressing proteins in infected hosts.

[0122] Specific initiation signals may also be required for efficient translation of the claimed isolated nucleic acid coding sequences. These signals include the ATG initiation codon and adjacent sequences. Exogenous translational control signals, including the ATG initiation codon, may additionally need to be provided. One of ordinary skill in the art would readily be capable of determining this and providing the necessary signals. It is well known that the initiation codon must be in-frame (or in-phase) with the reading frame of the desired coding sequence to ensure translation of the entire insert. These exogenous translational control signals and initiation codons may be of a variety of origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of appropriate transcription enhancer elements or transcription terminators (Bittner et al., 1987).

[0123] In eukaryotic expression, one will also typically desire to incorporate into the transcriptional unit an appropriate polyadenylation site (e.g., 5'-AATAAA-3') if one was not contained within the original cloned segment. Typically, the poly A addition site is placed about 30 to 2000 nucleotides "downstream" of the termination site of the protein at a position prior to transcription termination.

[0124] For long-term, high-yield production of recombinant proteins, stable expression is preferred. For example, cell lines that stably express constructs encoding proteins may be engineered. Rather than using expression vectors that contain viral origins of replication, host cells may be transformed with vectors controlled by appropriate expression control elements (e.g., promoter, enhancer, sequences, transcription terminators, polyadenylation sites, etc.), and a selectable marker. Following the introduction of foreign DNA, engineered cells may be allowed to grow for 1-2 days in an enriched media, and then are switched to a selective media. The selectable marker in the recombinant plasmid confers resistance to the selection and allows cells to stably integrate the plasmid into their chromosomes and grow to form foci which in turn may be cloned and expanded into cell lines.

[0125] A number of selection systems may be used, including but not limited to, the herpes simplex virus thymidine kinase (Wigler et al., 1977), hypoxanthine-guanine phosphoribosyltransferase (Szybalska et al., 1962) and adenine phosphoribosyltransferase genes (Lowy et al., 1980), in tk-, hgprt- or aprt-cells, respectively. Also, antimetabolite resistance may be used as the basis of selection for dhfr, that confers resistance to methotrexate (Wigler et al., 1980; O'Hare et al., 1981); gpt, that confers resistance to mycophenolic acid (Mulligan et al., 1981); neo, that confers resistance to the aminoglycoside G-418 (Colberre-Garapinet al., 1981); and hygro, that confers resistance to hygromycin (Santerre et al., 1984).

[0126] It is contemplated that the isolated nucleic acids of the invention may be "overexpressed", i.e., expressed in increased levels relative to its natural expression in human prostate, bladder or breast cells, or even relative to the expression of other proteins in the recombinant host cell. Such overexpression may be assessed by a variety of methods, including radio-labeling and/or protein purification. However, simple and direct methods are preferred, for example, those involving SDS/PAGE and protein staining or Western blotting, followed by quantitative analyses, such as densitometric scanning of the resultant gel or blot. A specific increase in the level of the recombinant protein or peptide in comparison to the level in natural human prostate, bladder or breast cells is indicative of overexpression, as is a relative abundance of the specific protein in relation to the other proteins produced by the host cell and, e.g., visible on a gel.

[0127] 2. Purification of Expressed Proteins

[0128] Further aspects of the present invention concern the purification, and in particular embodiments, the substantial purification, of an encoded protein or peptide. The term "purified protein or peptide" as used herein, is intended to refer to a composition, isolatable from other components, wherein the protein or peptide is purified to any degree relative to its naturally-obtainable state, i.e., in this case, relative to its purity within a prostate, bladder or breast cell extract. A purified protein or peptide therefore also refers to a protein or peptide, free from the environment in which it may naturally occur.

[0129] Generally, "purified" will refer to a protein or peptide composition which has been subjected to fractionation to remove various other components, and which composition substantially retains its expressed biological activity. Where the term "substantially purified" is used, this will refer to a composition in which the protein or peptide forms the major component of the composition, such as constituting about 50% or more of the proteins in the composition.

[0130] Various methods for quantifying the degree of purification of the protein or peptide will be known to those of skill in the art in light of the present disclosure. These include, for example, determining the specific activity of an active fraction, or assessing the number of polypeptides within a fraction by SDS/PAGE analysis. A preferred method for assessing the purity of a fraction is to calculate the specific activity of the fraction, to compare it to the specific activity of the initial extract, and to thus calculate

the degree of purity, herein assessed by a "-fold purification number". The actual units used to represent the amount of activity will, of course, be dependent upon the particular assay technique chosen to follow the purification and whether or not the expressed protein or peptide exhibits a detectable activity.

[0131] Various techniques suitable for use in protein purification will be well known to those of skill in the art. These include, for example, precipitation with ammonium sulfate, PEG, antibodies and the like or by heat denaturation, followed by centrifugation; chromatography steps such as ion exchange, gel filtration, reverse phase, hydroxylapatite and affinity chromatography; isoelectric focusing; gel electrophoresis; and combinations of such and other techniques. As is generally known in the art, it is believed that the order of conducting the various purification steps may be changed, or that certain steps may be omitted, and still result in a suitable method for the preparation of a substantially purified protein or peptide.

[0132] There is no general requirement that the protein or peptide always be provided in the most purified state. Indeed, it is contemplated that less substantially purified products will have utility in certain embodiments. Partial purification may be accomplished by using fewer purification steps in combination, or by utilizing different forms of the same general purification scheme. For example, it is appreciated that a cation-exchange column chromatography performed utilizing an HPLC apparatus will generally result in a greater—fold purification than the same technique utilizing a low pressure chromatography system. Methods exhibiting a lower degree of relative purification may have advantages in total recovery of protein product, or in maintaining the activity of an expressed protein.

[0133] It is known that the migration of a polypeptide may vary, sometimes significantly, with different conditions of SDS/PAGE (Capaldi et al., *Biochem. Biophys. Res. Comm.*, 76:425, 1977). It will therefore be appreciated that under differing electrophoresis conditions, the apparent molecular weights of purified or partially purified expression products may vary.

[0134] 3. Antibody Generation

[0135] For some embodiments, it will be desirable to produce antibodies that bind with high specificity to the polypeptide product(s) of an isolated nucleic acid selected from SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:83 and SEQ ID NO:85. Means for preparing and characterizing antibodies are well known in the art (See, e.g., Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory, 1988; incorporated herein by reference).

[0136] Methods for generating polyclonal antibodies are well known in the art. Briefly, a polyclonal antibody is prepared by immunizing an animal with an immunogenic composition and collecting antisera from that immunized animal. A wide range of animal species may be used for the production of antisera. Typically the animal used for production of anti-antisera is a rabbit, a mouse, a rat, a hamster, a guinea pig or a goat. Because of the relatively large blood volume of rabbits, a rabbit is a preferred choice for production of polyclonal antibodies.

[0137] As is well known in the art, a given composition may vary in its immunogenicity. It is often necessary therefore to boost the host immune system, as may be achieved by coupling a peptide or polypeptide immunogen to a carrier. Exemplary and preferred carriers are keyhole limpet hemocyanin (KLH) and bovine serum albumin (BSA). Other albumins such as ovalbumin, mouse serum albumin or rabbit serum albumin may also be used as carriers. Means for conjugating a polypeptide to a carrier protein are well known in the art and include glutaraldehyde, m-maleimidobenzoyl-N-hydroxysuccinimide ester, carbodiimide and bis-biazotized benzidine.

[0138] As is also well known in the art, the immunogenicity of a particular immunogen composition may be enhanced by the use of non-specific stimulators of the immune response, known as adjuvants. Exemplary and preferred adjuvants include complete Freund's adjuvant (a non-specific stimulator of the immune response containing killed *Mycobacterium tuberculosis*), incomplete Freund's adjuvants and aluminum hydroxide adjuvant.

[0139] The amount of immunogen composition used in the production of polyclonal antibodies varies upon the nature of the immunogen as well as the animal used for immunization. A variety of routes may be used to administer the immunogen (subcutaneous, intramuscular, intradermal, intravenous and intraperitoneal). The production of polyclonal antibodies may be monitored by sampling blood of the immunized animal at various points following immunization. A second, booster injection, may also be given. The process of boosting and titering is repeated until a suitable titer is achieved. When a desired level of immunogenicity is obtained, the immunized animal may be bled and the serum isolated and stored, and/or the animal may be used to generate MAbs. For production of rabbit polyclonal antibodies, the animal may be bled through an ear vein or alternatively by cardiac puncture. The removed blood is allowed to coagulate and then centrifuged to separate serum components from whole cells and blood clots. The serum may be used as is for various applications or else the desired antibody fraction may be purified by well-known methods, such as affinity chromatography using another antibody or a peptide bound to a solid matrix.

[0140] Monoclonal antibodies (MAbs) may be readily prepared through use of well-known techniques, such as those exemplified in U.S. Pat. No. 4,196,265, incorporated herein by reference. Typically, this technique involves immunizing a suitable animal with a selected immunogen composition, e.g., a purified or partially purified expressed protein, polypeptide or peptide. The immunizing composition is administered in a manner effective to stimulate antibody producing cells.

[0141] The methods for generating monoclonal antibodies (MAbs) generally begin along the same lines as those for preparing polyclonal antibodies. Rodents such as mice and rats are preferred animals, however, the use of rabbit, sheep or frog cells is also possible. The use of rats may provide certain advantages (Goding, 1986, pp. 60-61), but mice are preferred, with the BALB/c mouse being most preferred as this is most routinely used and generally gives a higher percentage of stable fusions.

[0142] The animals are injected with antigen as described above. The antigen may be coupled to carrier molecules such as keyhole limpet hemocyanin if necessary. The antigen would typically be mixed with adjuvant, such as Freund's complete or incomplete adjuvant. Booster injections with the same antigen would occur at approximately two-week intervals.

[0143] Following immunization, somatic cells with the potential for producing antibodies, specifically B lymphocytes (B cells), are selected for use in the MAb generating protocol. These cells may be obtained from biopsied spleens, tonsils or lymph nodes, or from a peripheral blood sample. Spleen cells and peripheral blood cells are preferred, the former because they are a rich source of antibody-producing cells that are in the dividing plasmablast stage, and the latter because peripheral blood is easily accessible. Often, a panel of animals will have been immunized and the spleen of the animal with the highest antibody titer will be removed and the spleen lymphocytes obtained by homogenizing the spleen with a syringe. Typically, a spleen from an immunized mouse contains approximately 5×10^7 to 2×10^8 lymphocytes.

[0144] The antibody-producing B lymphocytes from the immunized animal are then fused with cells of an immortal myeloma cell, generally one of the same species as the animal that was immunized. Myeloma cell lines suited for use in hybridoma-producing fusion procedures preferably are non-antibody-producing, have high fusion efficiency, and enzyme deficiencies that render then incapable of growing in certain selective media which support the growth of only the desired fused cells (hybridomas).

[0145] Any one of a number of myeloma cells may be used, as are known to those of skill in the art (Goding, pp. 65-66, 1986; Campbell, pp. 75-83, 1984). For example, where the immunized animal is a mouse, one may use P3-X63/Ag8,x63-Ag8.653, NS1/1.Ag 41, Sp210-Ag14, FO, NSO/U, MPC-11, MPC11-X45-GTG 1.7 and S194/5XXO Bul; for rats, one may use R210.RCY3, Y3-Ag 1.2.3, IR983F and 4B210; and U-266, GM1500-GRG2, LICR-LON-HMy2 and UC729-6 are all useful in connection with human cell fusions.

[0146] One preferred murine myeloma cell is the NS-1 myeloma cell line (also termed P3-NS-1-Ag4-1), which is readily available from the NIGMS Human Genetic Mutant Cell Repository by requesting cell line repository number GM3573. Another mouse myeloma cell line that may be used is the 8-azaguanine-resistant mouse murine myeloma SP2/0 non-producer cell line.

[0147] Methods for generating hybrids of antibody-producing spleen or lymph node cells and myeloma cells usually comprise mixing somatic cells with myeloma cells in a 2:1 proportion, though the proportion may vary from about 20:1 to about 1:1, respectively, in the presence of an agent or agents (chemical or electrical) that promote the fusion of cell membranes. Fusion methods using Sendai virus have been described by Kohler and Milstein (1975; 1976), and those using polyethylene glycol (PEG), such as 37% (v/v) PEG, by Gefter et al. (1977). The use of electrically induced fusion methods is also appropriate (Godingpp. 71-74, 1986).

[0148] Fusion procedures usually produce viable hybrids at low frequencies, about 1×10^{-6} to 1×10^{-8}. However, this does not pose a problem, as the viable, fused hybrids are differentiated from the parental, unfused cells (particularly the unfused myeloma cells that would normally continue to divide indefinitely) by culturing in a selective medium. The selective medium is generally one that contains an agent that blocks the de novo synthesis of nucleotides in the tissue culture media. Exemplary and preferred agents are aminopterin, methotrexate, and azaserine. Aminopterin and methotrexate block de novo synthesis of both purines and pyrimidines, whereas azaserine blocks only purine synthesis. Where aminopterin or methotrexate is used, the media is supplemented with hypoxanthine and thymidine as a source of nucleotides (HAT medium). Where azaserine is used, the media is supplemented with hypoxanthine.

[0149] The preferred selection medium is HAT. Only cells capable of operating nucleotide salvage pathways are able to survive in HAT medium. The myeloma cells are defective in key enzymes of the salvage pathway, e.g., hypoxanthine phosphoribosyl transferase (HPRT), and they cannot survive. The B cells may operate this pathway, but they have a limited life span in culture and generally die within about two weeks. Therefore, the only cells that can survive in the selective media are those hybrids formed from myeloma and B cells.

[0150] This culturing provides a population of hybridomas from which specific hybridomas are selected. Typically, selection of hybridomas is performed by culturing the cells by single-clone dilution in microtiter plates, followed by testing the individual clonal supernatants (after about two to three weeks) for the desired reactivity. The assay should be sensitive, simple and rapid, such as radioimmunoassays, enzyme immunoassays, cytotoxicity assays, plaque assays, dot immunobinding assays, and the like.

[0151] The selected hybridomas would then be serially diluted and cloned into individual antibody-producing cell lines, which clones may then be propagated indefinitely to provide MAbs. The cell lines may be exploited for MAb production in two basic ways. A sample of the hybridoma may be injected (often into the peritoneal cavity) into a histocompatible animal of the type that was used to provide the somatic and myeloma cells for the original fusion. The injected animal develops tumors secreting the specific monoclonal antibody produced by the fused cell hybrid. The body fluids of the animal, such as serum or ascites fluid, may then be tapped to provide MAbs in high concentration. The individual cell lines may also be cultured in vitro, where the MAbs are naturally secreted into the culture medium from which they may be readily obtained in high concentrations. MAbs produced by either means may be further purified, if desired, using filtration, centrifugation and various chromatographic methods such as HPLC or affinity chromatography.

[0152] Large amounts of the monoclonal antibodies of the present invention may also be obtained by multiplying hybridoma cells in vivo. Cell clones are injected into mammals which are histocompatible with the parent cells, e.g., syngeneic mice, to cause growth of antibody-producing tumors. Optionally, the animals are primed with a hydrocarbon, especially oils such as pristane (tetramethylpentadecane) prior to injection.

[0153] In accordance with the present invention, fragments of the monoclonal antibody of the invention may be

obtained from the monoclonal antibody produced as described above, by methods which include digestion with enzymes such as pepsin or papain and/or cleavage of disulfide bonds by chemical reduction. Alternatively, monoclonal antibody fragments encompassed by the present invention may be synthesized using an automated peptide synthesizer.

[0154] The monoclonal conjugates of the present invention are prepared by methods known in the art, e.g., by reacting a monoclonal antibody prepared as described above with, for instance, an enzyme in the presence of a coupling agent such as glutaraldehyde or periodate. Conjugates with fluorescein markers are prepared in the presence of these coupling agents or by reaction with an isothiocyanate. Conjugates with metal chelates are similarly produced. Other moieties to which antibodies may be conjugated include radionuclides such as ^{3}H, ^{125}I , ^{131}I, ^{32}P, ^{35}S, ^{14}C, ^{51}Cr, ^{36}Cl, ^{57}Co, ^{58}Co, ^{59}Fe, ^{75}Se, ^{152}Eu, and ^{99m}Tc. Radioactively labeled monoclonal antibodies of the present invention are produced according to well-known methods in the art. For instance, monoclonal antibodies may be iodinated by contact with sodium or potassium iodide and a chemical oxidizing agent such as sodium hypochlorite, or an enzymatic oxidizing agent, such as lactoperoxidase. Monoclonal antibodies according to the invention may be labeled with technetium-^{99} by ligand exchange process, for example, by reducing pertechnate with stannous solution, chelating the reduced technetium onto a Sephadex column and applying the antibody to this column or by direct labeling techniques, e.g., by incubating pertechnate, a reducing agent such as $SNCl_2$, a buffer solution such as sodium-potassium phthalate solution, and the antibody.

[0155] It will be appreciated by those of skill in the art that monoclonal or polyclonal antibodies specific for proteins that are preferentially expressed in metastatic or nonmetastatic human prostate, bladder or breast cancer will have utilities in several types of applications. These may include the production of diagnostic kits for use in detecting or diagnosing human prostate, bladder or breast cancer. An alternative use would be to link such antibodies to therapeutic agents, such as chemotherapeutic agents, followed by administration to individuals with prostate, bladder or breast cancer, thereby selectively targeting the prostate, bladder or breast cancer cells for destruction. The skilled practitioner will realize that such uses are within the scope of the present invention.

[0156] D. Immunodetection Assays

[0157] 1. Immunodetection Methods

[0158] In still further embodiments, the present invention concerns immunodetection methods for binding, purifying, removing, quantifying or otherwise generally detecting biological components. The encoded proteins or peptides of the present invention may be employed to detect antibodies having reactivity therewith, or, alternatively, antibodies prepared in accordance with the present invention, may be employed to detect the encoded proteins or peptides. The steps of various useful immunodetection methods have been described in the scientific literature, such as, e.g., Nakamura et al. (1987).

[0159] In general, the immunobinding methods include obtaining a sample suspected of containing a protein, peptide or antibody, and contacting the sample with an antibody or protein or peptide in accordance with the present invention, as the case may be, under conditions effective to allow the formation of immunocomplexes.

[0160] The immunobinding methods include methods for detecting or quantifying the amount of a reactive component in a sample, which methods require the detection or quantitation of any immune complexes formed during the binding process. Here, one would obtain a sample suspected of containing a prostate disease, bladder cancer or breast cancer marker encoded protein, peptide or a corresponding antibody, and contact the sample with an antibody or encoded protein or peptide, as the case may be, and then detect or quantify the amount of immune complexes formed under the specific conditions.

[0161] In terms of antigen detection, the biological sample analyzed may be any sample that is suspected of containing a prostate, bladder or breast cancer-specificantigen, such as a prostate, bladder, breast, or lymph node tissue section or specimen, a homogenized tissue extract, an isolated cell, a cell membrane preparation, separated or purified forms of any of the above protein-containing compositions, or even any biological fluid that comes into contact with prostate, bladder or breast tissues, including blood, lymphatic fluid, and even seminal or lactary fluids.

[0162] Contacting the chosen biological sample with the protein, peptide or antibody under conditions effective and for a period of time sufficient to allow the formation of immune complexes (primary immune complexes) is generally a matter of simply adding the composition to the sample and incubating the mixture for a period of time long enough for the antibodies to form immune complexes with, i.e., to bind to, any antigens present. After this time, the sample-antibody composition, such as a tissue section, ELISA plate, dot blot or Western blot, will generally be washed to remove any non-specifically bound antibody species, allowing only those antibodies specifically bound within the primary immune complexes to be detected.

[0163] In general, the detection of immunocomplex formation is well known in the art and may be achieved through the application of numerous approaches. These methods are generally based upon the detection of a label or marker, such as any radioactive, fluorescent, biological or enzymatic tags or labels of standard use in the art. U.S. Patents concerning the use of such labels include U.S. Pat. Nos. 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149 and 4,366,241, each incorporated herein by reference. Of course, one may find additional advantages through the use of a secondary binding ligand such as a second antibody or a biotin/avidin ligand binding arrangement, as is known in the art.

[0164] The encoded protein, peptide or corresponding antibody employed in the detection may itself be linked to a detectable label, wherein one would then simply detect this label, thereby allowing the amount of the primary immune complexes in the composition to be determined.

[0165] Alternatively, the first added component that becomes bound within the primary immune complexes may be detected by means of a second binding ligand that has binding affinity for the encoded protein, peptide or corresponding antibody. In these cases, the second binding ligand

may be linked to a detectable label. The second binding ligand is itself often an antibody, which may thus be termed a "secondary" antibody. The primary immune complexes are contacted with the labeled, secondary binding ligand, or antibody, under conditions effective and for a period of time sufficient to allow the formation of secondary immune complexes. The secondary immune complexes are then generally washed to remove any non-specifically bound labeled secondary antibodies or ligands, and the remaining label in the secondary immune complexes is then detected.

[0166] Further methods include the detection of primary immune complexes by a two step approach. A second binding ligand, such as an antibody, that has binding affinity for the encoded protein, peptide or corresponding antibody is used to form secondary immune complexes, as described above. After washing, the secondary immune complexes are contacted with a third binding ligand or antibody that has binding affinity for the second antibody, again under conditions effective and for a period of time sufficient to allow the formation of immune complexes (tertiary immune complexes). The third ligand or antibody is linked to a detectable label, allowing detection of the tertiary immune complexes thus formed. This system may provide for signal amplification if this is desired.

[0167] The immunodetection methods of the present invention have evident utility in the diagnosis of conditions such as prostate cancer, benign prostate hyperplasia, bladder cancer and breast cancer. Here, a biological or clinical sample suspected of containing either the encoded protein or peptide or corresponding antibody is used. However, these embodiments also have applications to non-clinical samples, such as in the titering of antigen or antibody samples, in the selection of hybridomas, and the like.

[0168] In the clinical diagnosis or monitoring of patients with prostate cancer, the detection of an antigen encoded by a prostate cancer marker nucleic acid, or an increase in the levels of such an antigen, in comparison to the levels in a corresponding biological sample from a normal subject is indicative of a patient with prostate cancer. The basis for such diagnostic methods lies, in part, with the finding that the nucleic acid prostate cancer markers identified in the present invention are overexpressed in prostate cancer tissue samples (see Examples below). By extension, it may be inferred that at least some of these markers produce elevated levels of encoded proteins, that may also be used as prostate cancer markers.

[0169] In the clinical diagnosis or monitoring of patients with breast cancer, the detection of an antigen encoded by a breast cancer marker nucleic acid, or an increase in the levels of such an antigen, in comparison to the levels in a corresponding biological sample from a normal subject is indicative of a patient with breast cancer. The basis for such diagnostic methods lies, in part, with the finding that the nucleic acid breast cancer marker identified in the present invention are overexpressed in breast cancer tissue samples (see Examples below). By extension, it may be inferred that this marker produces elevated levels of encoded protein, that may also be used as a breast cancer marker.

[0170] In the clinical diagnosis or monitoring of patients with bladder cancer, the detection of an antigen encoded by a bladder cancer marker nucleic acid, or an increase in the levels of such an antigen, in comparison to the levels in a corresponding biological sample from a normal subject is indicative of a patient with bladder cancer. The basis for such diagnostic methods lies, in part, with the finding that the nucleic acid bladder cancer marker identified in the present invention are overexpressed in bladder cancer tissue samples (see Examples below). By extension, it may be inferred that this marker produces elevated levels of encoded protein, that may also be used as a bladder cancer marker.

[0171] Those of skill in the art are very familiar with differentiating between significant expression of a biomarker, which represents a positive identification, and low level or background expression of a biomarker. Indeed, background expression levels are often used to form a "cut-off" above which increased staining will be scored as significant or positive. Significant expression may be represented by high levels of antigens in tissues or within body fluids, or alternatively, by a high proportion of cells from within a tissue that each give a positive signal.

[0172] 2. Immunohistochemistry

[0173] The antibodies of the present invention may be used in conjunction with both fresh-frozen and formalin-fixed, paraffin-embedded tissue blocks prepared by immunohistochemistry(IHC). Any IHC method well known in the art may be used such as those described in *Diagnostic Immunopathology,* 2nd edition. edited by, Robert B. Colvin, Atul K. Bhan and Robert T. McCluskey. Raven Press, New York., 1995, (incorporated herein by reference) and in particular, Chapter 31 of that reference entitled Gynecological and Genitourinary Tumors (pages 579-597), by Debra A. Bell, Robert H. Young and Robert E. Scully and references therein.

[0174] 3. ELISA

[0175] As noted, it is contemplated that the encoded proteins or peptides of the invention will find utility as immunogens, e.g., in connection with vaccine development, in immunohistochemistry and in ELISA assays. One evident utility of the encoded antigens and corresponding antibodies is in immunoassays for the detection of prostate disease, bladder cancer or breast cancer marker proteins, as needed in diagnosis and prognostic monitoring.

[0176] Immunoassays, in their most simple and direct sense, are binding assays. Certain preferred immunoassays are the various types of enzyme linked immunosorbent assays (ELISAs) and radioimmunoassays (RIA) known in the art. Immunohistochemical detection using tissue sections is also particularly useful. However, it will be readily appreciated that detection is not limited to such techniques, and Western blotting, dot blotting, FACS analyses, and the like may also be used.

[0177] In one exemplary ELISA, antibodies binding to the encoded proteins of the invention are immobilized onto a selected surface exhibiting protein affinity, such as a well in a polystyrene microtiter plate. Then, a test composition suspected of containing the prostate disease, bladder cancer or breast cancer marker antigen, such as a clinical sample, is added to the wells. After binding and washing to remove non-specifically bound immunecomplexes, the bound antigen may be detected. Detection is generally achieved by the addition of a second antibody specific for the target protein, that is linked to a detectable label. This type of ELISA is a simple "sandwich ELISA". Detection may also be achieved

by the addition of a second antibody, followed by the addition of a third antibody that has binding affinity for the second antibody, with the third antibody being linked to a detectable label.

[0178] In another exemplary ELISA, the samples suspected of containing the prostate disease, bladder cancer or breast cancer marker antigen are immobilized onto the well surface and then contacted with the antibodies of the invention. After binding and washing to remove non-specifically bound immunecomplexes, the bound antigen is detected. Where the initial antibodies are linked to a detectable label, the immunecomplexes may be detected directly. Again, the immunecomplexes may be detected using a second antibody that has binding affinity for the first antibody, with the second antibody being linked to a detectable label.

[0179] Another ELISA in which the proteins or peptides are immobilized, involves the use of antibody competition in the detection. In this ELISA, labeled antibodies are added to the wells, allowed to bind to the prostate disease, bladder cancer or breast cancer marker protein, and detected by means of their label. The amount of marker antigen in an unknown sample is then determined by mixing the sample with the labeled antibodies before or during incubation with coated wells. The presence of marker antigen in the sample acts to reduce the amount of antibody available for binding to the well and thus reduces the ultimate signal. This is appropriate for detecting antibodies in an unknown sample, where the unlabeled antibodies bind to the antigen-coated wells and also reduces the amount of antigen available to bind the labeled antibodies.

[0180] Irrespective of the format employed, ELISAs have certain features in common, such as coating, incubating or binding, washing to remove non-specifically bound species, and detecting the bound immunecomplexes. These are described as follows:

[0181] In coating a plate with either antigen or antibody, one will generally incubate the wells of the plate with a solution of the antigen or antibody, either overnight or for a specified period of hours. The wells of the plate will then be washed to remove incompletely adsorbed material. Any remaining available surfaces of the wells are then "coated" with a nonspecific protein that is antigenically neutral with regard to the test antisera. These include bovine serum albumin (BSA), casein and solutions of milk powder. The coating allows for blocking of nonspecific adsorption sites on the immobilizing surface and thus reduces the background caused by nonspecific binding of antisera onto the surface.

[0182] In ELISAs, it is probably more customary to use a secondary or tertiary detection means rather than a direct procedure. Thus, after binding of a protein or antibody to the well, coating with a non-reactive material to reduce background, and washing to remove unbound material, the immobilizing surface is contacted with the control human prostate, bladder or breast cancer and/or clinical or biological sample to be tested under conditions effective to allow immunecomplex (antigen/antibody) formation. Detection of the immunecomplex then requires a labeled secondary binding ligand or antibody, or a secondary binding ligand or antibody in conjunction with a labeled tertiary antibody or third binding ligand.

[0183] "Under conditions effective to allow immunecomplex (antigen/antibody) formation" means that the conditions preferably include diluting the antigens and antibodies with solutions such as BSA, bovine gamma globulin (BGG) and phosphate buffered saline (PBS)/Tween. These added agents also tend to assist in the reduction of nonspecific background.

[0184] The "suitable" conditions also mean that the incubation is at a temperature and for a period of time sufficient to allow effective binding. Incubation steps are typically from about 1 to 2 to 4 hours, at temperatures preferably on the order of 25° to 27° C., or may be overnight at about 4° C. or so.

[0185] Following all incubation steps in an ELISA, the contacted surface is washed so as to remove non-complexed material. A preferred washing procedure includes washing with a solution such as PBS/Tween, or borate buffer. Following the formation of specific immunecomplexes between the test sample and the originally bound material, and subsequent washing, the occurrence of even minute amounts of immunecomplexes may be determined.

[0186] To provide a detecting means, the second or third antibody will have an associated label to allow detection. Preferably, this will be an enzyme that will generate color development upon incubating with an appropriate chromogenic substrate. Thus, for example, one will desire to contact and incubate the first or second immunecomplex with a urease, glucose oxidase, alkaline phosphatase or hydrogen peroxidase-conjugated antibody for a period of time and under conditions that favor the development of further immunecomplex formation (e.g., incubation for 2 hours at room temperature in a PBS-containing solution such as PBS-Tween).

[0187] After incubation with the labeled antibody, and subsequent to washing to remove unbound material, the amount of label is quantified, e.g., by incubation with a chromogenic substrate such as urea and bromocresol purple or 2,2'-azido-di-(3-ethyl-benzthiazoline-6-sulfonic acid [ABTS] and H_2O_2, in the case of peroxidase as the enzyme label. Quantitation is then achieved by measuring the degree of color generation, e.g., using a visible spectra spectrophotometer.

[0188] 4 Use of Antibodies for Radioimaging

[0189] The antibodies of this invention will be used to quantify and localize the expression of the encoded marker proteins. The antibody, for example, will be labeled by any one of a variety of methods and used to visualize the localized concentration of the cells producing the encoded protein.

[0190] The invention also relates to an in vivo method of imaging a pathological prostate, bladder or breast cancer condition using the above described monoclonal antibodies. Specifically, this method involves administering to a subject an imaging-effective amount of a detectably-labeled prostate, bladder or breast cancer-specific monoclonal antibody or fragment thereof and a pharmaceutically effective carrier and detecting the binding of the labeled monoclonal antibody to the diseased tissue. The term "in vivo imaging" refers to any method which permits the detection of a labeled monoclonal antibody of the present invention or fragment thereof that specifically binds to a diseased tissue located in the subject's body. A "subject" is a mammal, preferably a human. An "imaging effective amount" means that the

amount of the detectably-labeled monoclonal antibody, or fragment thereof, administered is sufficient to enable detection of binding of the monoclonal antibody or fragment thereof to the diseased tissue.

[0191] A factor to consider in selecting a radionuclide for in vivo diagnosis is that the half-life of a nuclide be long enough so that it is still detectable at the time of maximum uptake by the target, but short enough so that deleterious radiation upon the host, as well as background, is minimized. Ideally, a radionuclide used for in vivo imaging will lack a particulate emission, but produce a large number of photons in a 140-2000 keV range, which may be readily detected by conventional gamma cameras.

[0192] A radionuclide may be bound to an antibody either directly or indirectly by using an intermediary functional group. Intermediary functional groups which are often used to bind radioisotopes which exist as metallic ions to antibody are diethylenetriaminepentaacetic acid (DTPA) and ethylene diaminetetracetic acid (EDTA). Examples of metallic ions suitable for use in this invention are ^{99m}Tc, ^{123}I, ^{131}I, ^{111}In, ^{131}I, ^{97}Ru, ^{67}Cu, ^{67}Ga, ^{125}I, ^{68}Ga, ^{72}As, ^{89}Zr, and ^{201}Tl.

[0193] In accordance with this invention, the monoclonal antibody or fragment thereof may be labeled by any of several techniques known to the art. The methods of the present invention may also use paramagnetic isotopes for purposes of in vivo detection. Elements particularly useful in Magnetic Resonance Imaging ("MRI") include ^{157}Gd, ^{55}Mn, ^{162}Dy, ^{52}Cr, and ^{56}Fe.

[0194] Administration of the labeled antibody may be local or systemic and accomplished intravenously, intraarterially, via the spinal fluid or the like. Administration may also be intradermal or intracavitary, depending upon the body site under examination. After a sufficient time has lapsed for the monoclonal antibody or fragment thereof to bind with the diseased tissue, for example 30 minutes to 48 hours, the area of the subject under investigation is examined by routine imaging techniques such as MRI, SPECT, planar scintillation imaging and emerging imaging techniques, as well. The exact protocol will necessarily vary depending upon factors specific to the patient, as noted above, and depending upon the body site under examination, method of administration and type of label used; the determination of specific procedures would be routine to the skilled artisan. The distribution of the bound radioactive isotope and its increase or decrease with time is then monitored and recorded. By comparing the results with data obtained from studies of clinically normal individuals, the presence and extent of the diseased tissue may be determined.

[0195] It will be apparent to those of skill in the art that a similar approach may be used to radio-image the production of the encoded prostate disease, bladder cancer or breast cancer marker proteins in human patients. The present invention provides methods for the in vivo diagnosis of prostate, bladder or breast cancer in a patient. Such methods generally comprise administering to a patient an effective amount of a prostate, bladder or breast cancer specific antibody, which antibody is conjugated to a marker, such as a radioactive isotope or a spin-labeled molecule, that is detectable by non-invasive methods. The antibody-marker conjugate is allowed sufficient time to come into contact with reactive antigens that be present within the tissues of the patient, and the patient is then exposed to a detection device to identify the detectable marker.

[0196] 5. Kits

[0197] In still further embodiments, the present invention concerns immunodetection kits for use with the immunodetection methods described above. As the encoded proteins or peptides may be employed to detect antibodies and the corresponding antibodies may be employed to detect encoded proteins or peptides, either or both of such components may be provided in the kit. The immunodetection kits will thus comprise, in suitable container means, an encoded protein or peptide, or a first antibody that binds to an encoded protein or peptide, and an immunodetection reagent.

[0198] In certain embodiments, the encoded protein or peptide, or the first antibody that binds to the encoded protein or peptide, may be bound to a solid support, such as a column matrix or well of a microtiter plate.

[0199] The immunodetection reagents of the kit may take any one of a variety of forms, including those detectable labels that are associated with or linked to the given antibody or antigen, and detectable labels that are associated with or attached to a secondary binding ligand. Exemplary secondary ligands are those secondary antibodies that have binding affinity for the first antibody or antigen, and secondary antibodies that have binding affinity for a human antibody.

[0200] Further suitable immunodetection reagents for use in the present kits include the two-component reagent that comprises a secondary antibody that has binding affinity for the first antibody or antigen, along with a third antibody that has binding affinity for the second antibody, the third antibody being linked to a detectable label.

[0201] The kits may further comprise a suitably aliquoted composition of the encoded protein or polypeptide antigen, whether labeled or unlabeled, as may be used to prepare a standard curve for a detection assay.

[0202] The kits may contain antibody-label conjugates either in fully conjugated form, in the form of intermediates, or as separate moieties to be conjugated by the user of the kit. The components of the kits may be packaged either in aqueous media or in lyophilized form.

[0203] The container means of the kits will generally include at least one vial, test tube, flask, bottle, syringe or other container means, into which the antibody or antigen may be placed, and preferably, suitably aliquoted. Where a second or third binding ligand or additional component is provided, the kit will also generally contain a second, third or other additional container into which this ligand or component may be placed. The kits of the present invention will also typically include a means for containing the antibody, antigen, and any other reagent containers in close confinement for commercial sale. Such containers may include injection or blow-molded plastic containers into which the desired vials are retained.

[0204] E. Detection and Quantitation of RNA Species

[0205] One embodiment of the instant invention comprises a method for identification of prostate, bladder or breast cancer cells in a biological sample by amplifying and

detecting nucleic acids corresponding to prostate, bladder or breast cancer cell markers. The biological sample may be any tissue or fluid in which prostate, bladder or breast cancer cells might be present. Various embodiments include bone marrow aspirate, bone marrow biopsy, lymph node aspirate, lymph node biopsy, spleen tissue, fine needle aspirate, skin biopsy or organ tissue biopsy. Other embodiments include samples where the body fluid is peripheral blood, lymph fluid, ascites, serous fluid, pleural effusion, sputum, cerebrospinal fluid, lacrimal fluid, stool or urine.

[0206] Nucleic acid used as a template for amplification is isolated from cells contained in the biological sample, according to standard methodologies. (Sambrook et al., 1989) The nucleic acid may be genomic DNA or fractionated or whole cell RNA. Where RNA is used, it may be desired to convert the RNA to a complementary cDNA. In one embodiment, the RNA is whole cell RNA and is used directly as the template for amplification.

[0207] Pairs of primers that selectively hybridize to nucleic acids corresponding to prostate, bladder or breast cancer specific markers are contacted with the isolated nucleic acid under conditions that permit selective hybridization. Once hybridized, the nucleic acid:primer complex is contacted with one or more enzymes that facilitate template-dependent nucleic acid synthesis. Multiple rounds of amplification, also referred to as "cycles," are conducted until a sufficient amount of amplification product is produced.

[0208] Next, the amplification product is detected. In certain applications, the detection may be performed by visual means. Alternatively, the detection may involve indirect identification of the product via chemiluminescence, radioactive scintigraphy of incorporated radiolabel or fluorescent label or even via a system using electrical or thermal impulse signals (Affymax technology; Bellus, 1994).

[0209] Following detection, one may compare the results seen in a given patient with a statistically significant reference group of normal patients and prostate, bladder or breast cancer patients. In this way, it is possible to correlate the amount of marker detected with various clinical states.

[0210] 1. Primers

[0211] The term primer, as defined herein, is meant to encompass any nucleic acid that is capable of priming the synthesis of a nascent nucleic acid in a template-dependent process. Typically, primers are oligonucleotides from ten to twenty base pairs in length, but longer sequences may be employed. Primers may be provided in double-stranded or single-stranded form, although the single-stranded form is preferred.

[0212] 2. Template Dependent Amplification Methods

[0213] A number of template dependent processes are available to amplify the marker sequences present in a given template sample. One of the best known amplification methods is the polymerase chain reaction (referred to as PCR) which is described in detail in U.S. Pat. Nos. 4,683,195, 4,683,202 and 4,800,159, and in Innis et al., 1990, each of which is incorporated herein by reference in its entirety.

[0214] Briefly, in PCR, two primer sequences are prepared which are complementary to regions on opposite complementary strands of the marker sequence. An excess of deoxynucleoside triphosphates are added to a reaction mixture along with a DNA polymerase, e.g., Taq polymerase. If the marker sequence is present in a sample, the primers will bind to the marker and the polymerase will cause the primers to be extended along the marker sequence by adding on nucleotides. By raising and lowering the temperature of the reaction mixture, the extended primers will dissociate from the marker to form reaction products, excess primers will bind to the marker and to the reaction products and the process is repeated.

[0215] A reverse transcriptase PCR amplification procedure may be performed in order to quantify the amount of mRNA amplified. Methods of reverse transcribing RNA into cDNA are well known and described in Sambrook et al., 1989. Alternative methods for reverse transcription utilize thermostable DNA polymerases. These methods are described in WO 90/07641 filed Dec. 21, 1990. Polymerase chain reaction methodologies are well known in the art. The most preferred methods of RT-PCR are as described herein in Example 1.

[0216] Another method for amplification is the ligase chain reaction ("LCR"), disclosed in European Application No. 320 308, incorporated herein by reference in its entirely. In LCR, two complementary probe pairs are prepared, and in the presence of the target sequence, each pair will bind to opposite complementary strands of the target such that they abut. In the presence of a ligase, the two probe pairs will link to form a single unit. By temperature cycling, as in PCR, bound ligated units dissociate from the target and then serve as "target sequences" for ligation of excess probe pairs. U.S. Pat. No. 4,883,750 describes a method similar to LCR for binding probe pairs to a target sequence.

[0217] Qbeta Replicase, described in PCT Application No. PCT/US87/00880, may also be used as still another amplification method in the present invention. In this method, a replicative sequence of RNA which has a region complementary to that of a target is added to a sample in the presence of an RNA polymerase. The polymerase will copy the replicative sequence which may then be detected.

[0218] An isothermal amplification method, in which restriction endonucleases and ligases are used to achieve the amplification of target molecules that contain nucleotide 5'-[alpha-thio]-triphosphatesin one strand of a restriction site may also be useful in the amplification of nucleic acids in the present invention. Walker et al., *Proc. Nat'l Acad. Sci. USA* 89:392-396 (1992), incorporated herein by reference in its entirety.

[0219] Strand Displacement Amplification (SDA) is another method of carrying out isothermal amplification of nucleic acids which involves multiple rounds of strand displacement and synthesis, i.e., nick translation. A similar method, called Repair Chain Reaction (RCR), involves annealing several probes throughout a region targeted for amplification, followed by a repair reaction in which only two of the four bases are present. The other two bases may be added as biotinylated derivatives for easy detection. A similar approach is used in SDA. Target specific sequences may also be detected using a cyclic probe reaction (CPR). In CPR, a probe having 3' and 5' sequences of non-specific DNA and a middle sequence of specific RNA is hybridized to DNA which is present in a sample. Upon hybridization, the reaction is treated with RNase H, and the products of the probe identified as distinctive products which are released

after digestion. The original template is annealed to another cycling probe and the reaction is repeated.

[0220] Still other amplification methods described in GB Application No. 2 202 328, and in PCT Application No. PCT/US89/01025, each of which is incorporated herein by reference in its entirety, may be used in accordance with the present invention. In the former application, "modified" primers are used in a PCR like, template and enzyme dependent synthesis. The primers may be modified by labeling with a capture moiety (e.g., biotin) and/or a detector moiety (e.g., enzyme). In the latter application, an excess of labeled probes are added to a sample. In the presence of the target sequence, the probe binds and is cleaved catalytically. After cleavage, the target sequence is released intact to be bound by excess probe. Cleavage of the labeled probe signals the presence of the target sequence.

[0221] Other nucleic acid amplification procedures include transcription-based amplification systems (TAS), including nucleic acid sequence based amplification (NASBA) and 3SR. Kwoh et al., *Proc. Nat'l Acad. Sci. USA* 86:1173 (1989); Gingeras et al., PCT Application WO 88/10315, incorporated herein by reference in their entirety. In NASBA, the nucleic acids may be prepared for amplification by standard phenol/chloroform extraction, heat denaturation of a clinical sample, treatment with lysis buffer and minispin columns for isolation of DNA and RNA or guanidinium chloride extraction of RNA. These amplification techniques involve annealing a primer which has target specific sequences. Following polymerization, DNA/RNA hybrids are digested with RNase H while double stranded DNA molecules are heat denatured again. In either case the single stranded DNA is made fully double stranded by addition of second target specific primer, followed by polymerization. The double-stranded DNA molecules are then multiply transcribed by a polymerase such as T7 or SP6. In an isothermal cyclic reaction, the RNA's are reverse transcribed into double stranded DNA, and transcribed once against with a polymerase such as T7 or SP6. The resulting products, whether truncated or complete, indicate target specific sequences.

[0222] Davey et al., European Application No. 329 822 (incorporated herein by reference in its entirely) disclose a nucleic acid amplification process involving cyclically synthesizing single-stranded RNA ("ssRNA"), ssDNA, and double-stranded DNA (dsDNA), which may be used in accordance with the present invention. The ssRNA is a first template for a first primer oligonucleotide, which is elongated by reverse transcriptase (RNA-dependent DNA polymerase). The RNA is then removed from the resulting DNA:RNA duplex by the action of ribonuclease H (RNase H, an RNase specific for RNA in duplex with either DNA or RNA). The resultant ssDNA is a second template for a second primer, which also includes the sequences of an RNA polymerase promoter (exemplified by T7 RNA polymerase) 5' to its homology to the template. This primer is then extended by DNA polymerase (exemplified by the large "Klenow" fragment of *E. coli* DNA polymerase I), resulting in a double-stranded DNA ("dsDNA") molecule, having a sequence identical to that of the original RNA between the primers and having additionally, at one end, a promoter sequence. This promoter sequence may be used by the appropriate RNA polymerase to make many RNA copies of the DNA. These copies may then re-enter the cycle leading to very swift amplification. With proper choice of enzymes, this amplification may be done isothermally without addition of enzymes at each cycle. Because of the cyclical nature of this process, the starting sequence may be chosen to be in the form of either DNA or RNA.

[0223] Miller et al., PCT Application WO 89/06700 (incorporated herein by reference in its entirety) disclose a nucleic acid sequence amplification scheme based on the hybridization of a promoter/primer sequence to a target single-stranded DNA ("ssDNA") followed by transcription of many RNA copies of the sequence. This scheme is not cyclic, i.e., new templates are not produced from the resultant RNA transcripts. Other amplification methods include "race" and "one-sided PCR." Frohman, M. A., In: *PCR PROTOCOLS: A GUIDE TO METHODS AND APPLICATIONS*, Academic Press, N.Y. (1990) and Ohara et al., *Proc. Nat'l Acad. Sci. USA,* 86:5673-5677 (1989), each herein incorporated by reference in their entirety.

[0224] Methods based on ligation of two (or more) oligonucleotides in the presence of nucleic acid having the sequence of the resulting "di-oligonucleotide", thereby amplifying the di-oligonucleotide, may also be used in the amplification step of the present invention. Wu et al., Genomics 4:560 (1989), incorporated herein by reference in its entirety.

[0225] 3. Separation Methods

[0226] Following amplification, it may be desirable to separate the amplification product from the template and the excess primer for the purpose of determining whether specific amplification has occurred. In one embodiment, amplification products are separated by agarose, agarose-acrylamide or polyacrylamide gel electrophoresis using standard methods. See Sambrook et al., 1989.

[0227] Alternatively, chromatographic techniques may be employed to effect separation. There are many kinds of chromatography which may be used in the present invention: adsorption, partition, ion-exchange and molecular sieve, and many specialized techniques for using them including column, paper, thin-layer and gas chromatography (Freifelder, 1982).

[0228] 4. Identification Methods

[0229] Amplification products must be visualized in order to confirm amplification of the marker sequences. One typical visualization method involves staining of a gel with ethidium bromide and visualization under UV light. Alternatively, if the amplification products are integrally labeled with radio- or fluorometrically-labeled nucleotides, the amplification products may then be exposed to x-ray film or visualized under the appropriate stimulating spectra, following separation.

[0230] In one embodiment, visualization is achieved indirectly. Following separation of amplification products, a labeled, nucleic acid probe is brought into contact with the amplified marker sequence. The probe preferably is conjugated to a chromophore but may be radiolabeled. In another embodiment, the probe is conjugated to a binding partner, such as an antibody or biotin, where the other member of the binding pair carries a detectable moiety.

[0231] In one embodiment, detection is by Southern blotting and hybridization with a labeled probe. The techniques

involved in Southern blotting are well known to those of skill in the art and may be found in many standard books on molecular protocols. See Sambrook et al., 1989. Briefly, amplification products are separated by gel electrophoresis. The gel is then contacted with a membrane, such as nitrocellulose, permitting transfer of the nucleic acid and non-covalent binding. Subsequently, the membrane is incubated with a chromophore-conjugated probe that is capable of hybridizing with a target amplification product. Detection is by exposure of the membrane to x-ray film or ion-emitting detection devices.

[0232] One example of the foregoing is described in U.S. Pat. No. 5,279,721, incorporated by reference herein, which discloses an apparatus and method for the automated electrophoresis and transfer of nucleic acids. The apparatus permits electrophoresis and blotting without external manipulation of the gel and is ideally suited to carrying out methods according to the present invention.

[0233] 5. Kit Components

[0234] All the essential materials and reagents required for detecting prostate disease, bladder cancer or breast cancer markers in a biological sample may be assembled together in a kit. The kit generally will comprise preselected primer pairs for one or more specific markers. For example a kit may include primers to detect RNA markers of normal tissue, BPH tissue, confined tumor tissue or metastically progressive tumor tissue, or any combination of these. Also included may be enzymes suitable for amplifying nucleic acids including various polymerases (RT, Taq, etc.), deoxynucleotides and buffers to provide the necessary reaction mixture for amplification. Preferred kits may also comprise primers for the detection of a control, non-differentially expressed RNA such as β-actin, for example.

[0235] The kits generally will comprise, in suitable means, distinct containers for each individual reagent and enzyme as well as for each marker primer pair. Preferred pairs of primers for amplifying nucleic acids are selected to amplify the sequences designated herein as SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:83 or SEQ ID NO:85.

[0236] In certain embodiments, kits will comprise hybridization probes specific for differentially expressed markers. The probes are designed to hybridize to a sequence or a complement of a sequence designated herein as SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:83 or SEQ ID NO:85. Such kits generally will comprise, in suitable means for close confinement, distinct containers for each individual reagent and enzyme as well as for each marker hybridization probe.

[0237] F. Use of RNA Fingerprinting to Identify Markers of Prostate Disease, Bladder Cancer or Breast Cancer

[0238] RNA fingerprinting is a means by which RNAs isolated from many different tissues, cell types or treatment groups may be sampled simultaneously to identify RNAs whose relative abundances vary. Two forms of this technology were developed simultaneously and reported in 1992 as RNA fingerprinting by differential display (Liang and Pardee, 1992; Welsh et al., 1992). (See also Liang and Pardee, U.S. Pat. No. 5,262,311, incorporated herein by reference in its entirety.) Both techniques were utilized in the studies described below. Some of the studies described herein were performed similarly to Donahue et al., *J. Biol. Chem.* 269: 8604-8609,1994.

[0239] All forms of RNA fingerprinting by PCR are theoretically similar but differ in their primer design and application. The most striking difference between differential display and other methods of RNA fingerprinting is that differential display utilizes anchoring primers that hybridize to the poly A tails of mRNAs. As a consequence, the PCR products amplified in differential display are biased towards the 3' untranslated regions of mRNAs.

[0240] The basic technique of differential display has been described in detail (Liang and Pardee, 1992). Total cell RNA is primed for first strand reverse transcription with an anchoring primer composed of oligo dT. The oligo dT primer is extended using a reverse transcriptase, for example, Moloney Murine Leukemia Virus (MMLV) reverse transcriptase. The synthesis of the second strand is primed with an arbitrarily chosen oligonucleotide, using reduced stringency conditions. Once the double-stranded cDNA has been synthesized, amplification proceeds by standard PCR techniques, utilizing the same primers. The resulting DNA fingerprint is analyzed by gel electrophoresis and ethidium bromide staining or autoradiography. A side by side comparison of fingerprints obtained from different cell derived RNAs using the same oligonucleotide primers identifies mRNAs that are differentially expressed.

[0241] RNA fingerprinting technology has been demonstrated as being effective in identifying genes that are differentially expressed in cancer (Liang et al., 1992; Wong et al., 1993; Sager et al., 1993; Mok et al., 1994; Watson et al., 1994; Chen et al., 1995; An et al., 1995). The present invention utilizes the RNA fingerprinting technique to identify genes that are differentially expressed in prostate, bladder or breast cancer. These studies utilized RNAs isolated from tumor tissues and tumor-derived cell lines that behave as tumors cells with different metastatic potential.

[0242] The underlying concept of these studies was that genes that are differentially expressed in cells with different metastatic potentials may be used as indicators of metastatic potential. Since metastasis is a prerequisite for prostate, bladder or breast cancer progression to life threatening pathologies, indicators of metastatic potential are likely to be indicators of pathological potential.

[0243] G. Design and Theoretical Considerations for Relative Quantitative RT-PCR

[0244] Reverse transcription (RT) of RNA to cDNA followed by relative quantitative PCR (RT-PCR) may be used to determine the relative concentrations of specific mRNA species in a series of total cell RNAs isolated from normal,

benign and cancerous prostate, bladder or breast tissues. By determining that the concentration of a specific mRNA species varies, it is shown that the gene encoding the specific mRNA species is differentially expressed. This technique may be used to confirm that mRNA transcripts shown to be differentially regulated by RNA fingerprinting are differentially expressed in prostate, bladder or breast cancer progression.

[0245] In PCR, the number of molecules of the amplified target DNA increase by a factor approaching two with every cycle of the reaction until some reagent becomes limiting. Thereafter, the rate of amplification becomes increasingly diminished until there is not an increase in the amplified target between cycles. If one plots a graph on which the cycle number is on the X axis and the log of the concentration of the amplified target DNA is on the Y axis, one observes that a curved line of characteristic shape is formed by connecting the plotted points. Beginning with the first cycle, the slope of the line is positive and constant. This is said to be the linear portion of the curve. After some reagent becomes limiting, the slope of the line begins to decrease and eventually becomes zero. At this point the concentration of the amplified target DNA becomes asymptotic to some fixed value. This is said to be the plateau portion of the curve.

[0246] The concentration of the target DNA in the linear portion of the PCR is directly proportional to the starting concentration of the target before the PCR was begun. By determining the concentration of the PCR products of the target DNA in PCR reactions that have completed the same number of cycles and are in their linear ranges, it is possible to determine the relative concentrations of the specific target sequence in the original DNA mixture. If the DNA mixtures are cDNAs synthesized from RNAs isolated from different tissues or cells, the relative abundances of the specific mRNA from which the target sequence was derived may be determined for the respective tissues or cells. This direct proportionality between the concentration of the PCR products and the relative mRNA abundances is only true in the linear range portion of the PCR reaction.

[0247] The final concentration of the target DNA in the plateau portion of the curve is determined by the availability of reagents in the reaction mix and is independent of the original concentration of target DNA. Therefore, the one condition that must be met before the relative abundances of an mRNA species may be determined by RT-PCR for a collection of RNA populations is that the concentrations of the amplified PCR products must be sampled when the PCR reactions are in the linear portion of their curves.

[0248] A second condition that must be met for an RT-PCR study to successfully determine the relative abundances of a particular mRNA species is that relative concentrations of the amplifiable cDNAs must be normalized to some independent standard. The goal of an RT-PCR study is to determine the abundance of a particular mRNA species relative to the average abundance of all mRNA species in the sample. In the studies described below, mRNAs for B-actin, asparagine synthetase and lipocortin II were used as external and internal standards to which the relative abundance of other mRNAs are compared.

[0249] Most protocols for competitive PCR utilize internal PCR standards that are approximately as abundant as the target. These strategies are effective if the products of the PCR amplifications are sampled during their linear phases. If the products are sampled when the reactions are approaching the plateau phase, then the less abundant product becomes relatively over represented. Comparisons of relative abundances made for many different RNA samples, such as when examining RNA samples for differential expression, become distorted in such a way as to make differences in relative abundances of RNAs appear less than they actually are. This is not a significant problem if the internal standard is much more abundant than the target. If the internal standard is more abundant than the target, then direct linear comparisons may be made between RNA samples.

[0250] The discussion above describes the theoretical considerations for an RT-PCR assay for clinically derived materials. The problems inherent in clinical samples are that they are of variable quantity (making normalization problematic), and that they are of variable quality (necessitating the co-amplification of a reliable internal control, preferably of larger size than the target). Both of these problems are overcome if the RT-PCR is performed as a relative quantitative RT-PCR with an internal standard in which the internal standard is an amplifiable cDNA fragment that is larger than the target cDNA fragment and in which the abundance of the mRNA encoding the internal standard is roughly 5-100 fold higher than the mRNA encoding the target. This assay measures relative abundance, not absolute abundance of the respective mRNA species.

[0251] Other studies described below were performed using a more conventional relative quantitative RT-PCR with an external standard protocol. These assays sample the PCR products in the linear portion of their amplification curves. The number of PCR cycles that are optimal for sampling must be empirically determined for each target cDNA fragment. In addition, the reverse transcriptase products of each RNA population isolated from the various tissue samples must be carefully normalized for equal concentrations of amplifiable cDNAs. This is very important since this assay measures absolute mRNA abundance. Absolute mRNA abundance may be used as a measure of differential gene expression only in normalized samples. While empirical determination of the linear range of the amplification curve and normalization of cDNA preparations are tedious and time consuming processes, the resulting RT-PCR assays may be superior to those derived from the relative quantitative RT-PCR with an internal standard.

[0252] One reason for this is that without the internal standard/competitor, all of the reagents may be converted into a single PCR product in the linear range of the amplification curve, increasing the sensitivity of the assay. Another reason is that with only one PCR product, display of the product on an electrophoretic gel or some other display method becomes less complex, has less background and is easier to interpret.

[0253] H. Diagnosis and Prognosis of Human Cancer

[0254] In certain embodiments, the present invention allows the diagnosis and prognosis of human prostate, bladder or breast cancer by screening for marker nucleic acids. The field of cancer diagnosis and prognosis is still uncertain. Various markers have been proposed to be correlated with metastasis and malignancy. They may be classified generally as cytologic, protein or nucleic acid markers.

[0255] Cytologic markers include such things as "nuclear roundedness" (Diamond et al., 1982) and cell ploidy. Protein markers include prostate specific antigen (PSA) and CA125. Nucleic acid markers have included amplification of Her2/neu, point mutations in the p53 or ras genes, and changes in the sizes of triplet repeat segments of particular chromosomes.

[0256] All of these markers exhibit certain drawbacks, associated with false positives and false negatives. A false positive result occurs when an individual without malignant cancer exhibits the presence of a "cancer marker". For example, elevated serum PSA has been associated with prostate carcinoma. However, it also occurs in some individuals with non-malignant, benign hyperplasia of the prostate. A false negative result occurs when an individual actually has cancer, but the test fails to show the presence of a specific marker. The incidence of false negatives varies for each marker, and frequently also by tissue type. For example, ras point mutations have been reported to range from a high of 95 percent in pancreatic cancer to a low of zero percent in some gynecologic cancers.

[0257] Additional problems arise when a marker is present only within the transformed cell itself. Ras point mutations may only be detected within the mutant cell, and are apparently not present in, for example, the blood serum or urine of individuals with ras-activated carcinomas. This means that, in order to detect a malignant tumor, one must take a sample of the tumor itself, or its metastatic cells. Since the object of cancer detection is to identify and treat tumors before they metastasize, essentially one must first identify and sample a tumor before the presence of the cancer marker can be detected.

[0258] Finally, specific problems occur with markers that are present in normal cells but absent in cancer cells. Most tumor samples will contain mixed populations of both normal and transformed cells. If one is searching for a marker that is present in normal cells, but occurs at reduced levels in transformed cells, the "background" signal from the normal cells in the sample may mask the presence of transformed cells.

[0259] The ideal cancer marker would be one that is present in malignant cancers, and either missing or else expressed at significantly lower levels in benign tumors and normal cells. Further, since any single marker would typically be present only in some proportion of malignant cancers, it is better to have a number of such markers for each cancer type. The present invention addresses this need for prostate, bladder and breast cancer markers by identifying several new nucleic acid markers that are expressed at much higher levels in malignant prostate carcinoma than in benign or normal prostate, as well as identifying a novel gene, UC Band #28 (SEQ ID NO:3, SEQ ID NO:83 and SEQ ID NO:85) whose mRNA transcripts are expressed at much higher levels in breast and bladder cancer than in their normal tissues of origin. In particular, the results for markers UC Band #28 (SEQ ID NO:3, SEQ ID NO:83 and SEQ ID NO:85) and UC Band #33 (SEQ ID NO:5), discussed in Examples 2 and 4 below, are quite promising in that these markers are apparently only overexpressed in malignant tumors and are present at very low levels in benign or normal prostate. Further, these markers are significantly elevated in a high percentage of human prostate cancers examined to date.

[0260] It is anticipated that in clinical applications, human tissue samples will be screened for the presence of the markers of prostate disease, bladder cancer or breast cancer identified herein. Such samples could consist of needle biopsy cores, surgical resection samples, lymph node tissue, or serum. In certain embodiments, nucleic acids would be extracted from these samples and amplified as described above. Some embodiments would utilize kits containing pre-selected primer pairs or hybridization probes. The amplified nucleic acids would be tested for the markers by, for example, gel electrophoresis and ethidium bromide staining, or Southern blotting, or a solid-phase detection means as described above. These methods are well known within the art. The levels of selected markers detected would be compared with statistically valid groups of metastatic, non-metastatic malignant, benign or normal prostate, bladder or breast samples. The diagnosis and prognosis of the individual patient would be determined by comparison with such groups.

[0261] Another embodiment of the present invention involves application of RT-PCR techniques to detect circulating prostate, bladder or breast cancer cells (ie., those that have already metastasized), using probes and primers selected from sequences or their complements designated herein as SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:83 or SEQ ID NO:85. Similar techniques have been described in PCT Patent Application No. WO 94/10343, incorporated herein by reference.

[0262] In this embodiment, metastatic prostate, bladder or breast cancer cells are detected in hematopoietic samples by amplification of prostate, bladder or breast cancer-specific nucleic acid sequences. Samples taken from blood or lymph nodes are treated as described below to purify total cell RNA. The isolated RNA is reverse transcribed using a reverse transcriptase and primers selected to bind under high stringency conditions to a nucleic acid sequence to the sequence of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:83 or SEQ ID NO:85. Following reverse transcription, the resulting cDNAs are amplified using standard PCR techniques (described below) and a thermostable DNA polymerase.

[0263] The presence of amplification products corresponding to prostate, bladder or breast cancer-marker nucleic acids may be detected by several alternative means. In one embodiment, the amplification product may be detected by gel electrophoresis and ethidium bromide staining. Alternatively, following the gel electrophoresis step the amplification product may be detected by standard Southern blotting techniques, using an hybridization probe selected to bind specifically to a prostate, bladder or breast cancer-marker nucleic acid sequence. Probe hybridization may in

turn be detected by a standard labeling means, for example, by incorporation of [^{32}P]-nucleotides followed by autoradiography. The amplification products may alternatively be detected using a solid phase detection system as described above, utilizing a prostate, bladder or breast cancer-marker specific hybridization probe and an appropriate labeling means. The presence of prostate, bladder or breast cancer-marker nucleic acids in blood or lymph node samples may be taken as indicative of a patient with metastatic prostate, bladder or breast cancer.

[0264] I. Targeted Inhibition of Prostate, Bladder and Breast Cancer Markers

[0265] In principal, the prostate, bladder or breast cancer-markers identified in the present invention may serve as targets for therapeutic intervention in prostate, bladder or breast cancer. One of the identified genes, cyclin A, has been described as a target for a number of agents that inhibit tumor cell growth by promoting differentiation or inhibiting cell division. For example, L-tyrosine has been reported to promote increased melanogenesis and replicative senescence in the B 16 melanoma cell line, correlated with a decrease in cyclin A activity. (Rieber & Rieber, 1994) Suramin is an antitumor agent that reduces the expression of cyclin A in the DU-145 prostate carcinoma cell line. (Qiao et al., 1994) Rapamycin inhibits cell proliferation in the YAC-1 T cell lymphoma and also inhibits cyclin A mRNA production. (Dumont et al., 1994) It is not clear if these inhibitors are acting directly on cyclin A, or somewhere upstream in a signal transduction/phosphorylation cascade pathway. However, inhibitors of cyclin A should inhibit cell proliferation and decrease tumor growth. Such inhibitors may have utility as therapeutic agents for the treatment of prostate cancer.

[0266] Inhibitors could also potentially be designed for the previously unreported prostate, bladder or breast cancer-markers identified in the present invention. This is complicated by the fact that no specific function has been identified for most of these gene products, and no data is available on their three-dimensional structures.

[0267] Identification of protein function may be extrapolated, in some cases, from the primary sequence data, provided that sequence homology exists between the unknown protein and a protein of similar sequence and known function. Proteins tend to occur in large families of relatively similar sequence and function. For example, a number of the serine proteases, like trypsin and chymotrypsin, have extensive sequence homologies and relatively similar three-dimensional structures. Other general categories of homologous proteins include different classes of transcriptional factors, membrane receptor proteins, tyrosine kinases, GTP-binding proteins, etc. The putative amino acid sequences encoded by the prostate, bladder or breast cancer-marker nucleic acids of the present invention may be cross-checked for sequence homologies versus the protein sequence database of the National Biomedical Research Fund. Homology searches are standard techniques for the skilled practitioner.

[0268] Even three-dimensional structure may be inferred from the primary sequence data of the encoded proteins. Again, if homologies exist between the encoded amino acid sequences and other proteins of known structure, then a model for the structure of the encoded protein may be designed, based upon the structure of the known protein. An example of this type of approach was reported by Ribas de Pouplana and Fothergill-Gilmore (Biochemistry 33: 7047-7055, 1994). These authors developed a detailed three-dimensional model for the structure of Drosophila alcohol dehydrogenase, based in part upon sequence homology with the known structure of 3-a, 20-β-hydroxysteroid dehydrogenase. Once a three-dimensional model is available, inhibitors may be designed by standard computer modeling techniques. This area has been recently reviewed by Sun and Cohen (Gene 137: 127-132,1993), herein incorporated by reference.

[0269] 1. Antisense Constructs

[0270] The term "antisense" is intended to refer to polynucleotide molecules complementary to a portion of a RNA marker of prostate disease, or a marker of bladder or breast cancer as defined herein. "Complementary" polynucleotides are those which are capable of base-pairing according to the standard Watson-Crick complementarity rules. That is, the larger purines will base pair with the smaller pyrimidines to form combinations of guanine paired with cytosine (G:C) and adenine paired with either thymine (A:T) in the case of DNA, or adenine paired with uracil (A:U) in the case of RNA. Inclusion of less common bases such as inosine, 5-methylcytosine, 6-methyladenine, hypoxanthine and others in hybridizing sequences does not interfere with pairing.

[0271] Antisense polynucleotides, when introduced into a target cell, specifically bind to their target polynucleotide and interfere with transcription, RNA processing, transport, translation and/or stability. Antisense RNA constructs, or DNA encoding such antisense RNA's, may be employed to inhibit gene transcription or translation or both within a host cell, either in vitro or in vivo, such as within a host animal, including a human subject.

[0272] The intracellular concentration of monovalent cation is approximately 160 mM (10 mM Na^+; 150 mM K^+). The intracellular concentration of divalent cation is approximately 20 mM (18 mM Mg^+; 2 mM Ca^{++}). The intracellular protein concentration, which would serve to decrease the volume of hybridization and, therefore, increase the effective concentration of nucleic acid species, is 150 mg/ml. Constructs can be tested in vitro under conditions that mimic these in vivo conditions.

[0273] Antisense constructs may be designed to bind to the promoter and other control regions, exons, introns or even exon-intron boundaries of a gene. It is contemplated that the most effective antisense constructs for the present invention will include regions complementary to the mRNA start site, or to those sequences identified herein as prostate disease or bladder or breast cancer markers. One can readily test such constructs simply by testing the constructs in vitro to determine whether levels of the target protein are affected. Similarly, detrimental non-specific inhibition of protein synthesis also can be measured by determining target cell viability in vitro.

[0274] As used herein, the terms "complementary" or "antisense" mean polynucleotides that are substantially complementary over their entire length and have very few base mismatches. For example, sequences of fifteen bases in length may be termed complementary when they have a complementary nucleotide at thirteen or fourteen nucle-

otides out of fifteen. Naturally, sequences which are "completely complementary" will be sequences which are entirely complementary throughout their entire length and have no base mismatches.

[0275] Other sequences with lower degrees of homology also are contemplated. For example, an antisense construct which has limited regions of high homology, but also contains a non-homologous region (e.g., a ribozyme) could be designed. These molecules, though having less than 50% homology, would bind to target sequences under appropriate conditions.

[0276] As stated above, although the antisense sequences may be full length cDNA copies, or large fragments thereof, they also may be shorter fragments, or "oligonucleotides," defined herein as polynucleotides of 50 or less bases. Although shorter oligomers (8-20) are easier to make and increase in vivo accessibility, numerous other factors are involved in determining the specificity of base-pairing. For example, both binding affinity and sequence specificity of an oligonucleotide to its complementary target increase with increasing length. It is contemplated that oligonucleotides of 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19,20,25, 30, 35, 40,45, 50 or 100 base pairs will be used. While all or part of the gene sequence may be employed in the context of antisense construction, statistically, any sequence of 14 bases long should occur only once in the human genome and, therefore, suffice to specify a unique target sequence.

[0277] In certain embodiments, one may wish to employ antisense constructs which include other elements, for example, those which include C-5 propyne pyrimidines. Oligonucleotides which contain C-5 propyne analogues of uridine and cytidine have been shown to bind RNA with high affinity and to be potent antisense inhibitors of gene expression (Wagner et al., 1993).

[0278] As an alternative to targeted antisense delivery, targeted ribozymes may be used. The term "ribozyme" is refers to an RNA-based enzyme capable of targeting and cleaving particular base sequences in both DNA and RNA. Ribozymes can either be targeted directly to cells, in the form of RNA oligonucleotides incorporating ribozyme sequences, or introduced into the cell as an expression vector encoding the desired ribozymal RNA. Ribozymes may be used and applied in much the same way as described for antisense polynucleotide. Ribozyme sequences also may be modified in much the same way as described for antisense polynucleotide. For example, one could incorporate non-Watson-Crick bases, or make mixed RNA/DNA oligonucleotides, or modify the phosphodiester backbone, or modify the 2'-hydroxy in the ribose sugar group of the RNA.

[0279] Alternatively, the antisense oligo- and polynucleotides according to the present invention may be provided as RNA via transcription from expression constructs that carry nucleic acids encoding the oligo- or polynucleotides. Throughout this application, the term "expression construct" is meant to include any type of genetic construct containing a nucleic acid encoding an antisense product in which part or all of the nucleic acid sequence is capable of being transcribed. Typical expression vectors include bacterial plasmids or phage, such as any of the pUC or Bluescript™ plasmid series or, as discussed further below, viral vectors adapted for use in eukaryotic cells.

[0280] In preferred embodiments, the nucleic acid encodes an antisense oligo- or polynucleotide under transcriptional control of a promoter. A "promoter" refers to a DNA sequence recognized by the synthetic machinery of the cell, or introduced synthetic machinery, required to initiate the specific transcription of a gene. The phrase "under transcriptional control" means that the promoter is in the correct location and orientation in relation to the nucleic acid to control RNA polymerase initiation.

[0281] The term promoter will be used here to refer to a group of transcriptional control modules that are clustered around the initiation site for RNA polymerase II. Much of the thinking about how promoters are organized derives from analyses of several viral promoters, including those for the HSV thymidine kinase (tk) and SV40 early transcription units. These studies, augmented by more recent work, have shown that promoters are composed of discrete functional modules, each consisting of approximately 7-20 bp of DNA, and containing one or more recognition sites for transcriptional activator or repressor proteins.

[0282] At least one module in each promoter functions to position the start site for RNA synthesis. The best known example of this is the TATA box, but in some promoters lacking a TATA box, such as the promoter for the mammalian terminal deoxynucleotidyl transferase gene and the promoter for the SV40 late genes, a discrete element overlying the start site itself helps to fix the place of initiation.

[0283] Additional promoter elements regulate the frequency of transcriptional initiation. Typically, these are located in the region 30-110 bp upstream of the start site, although a number of promoters have recently been shown to contain functional elements downstream of the start site as well. The spacing between promoter elements frequently is flexible, so that promoter function is preserved when elements are inverted or moved relative to one another. In the tk promoter, the spacing between promoter elements can be increased to 50 bp apart before activity begins to decline. Depending on the promoter, it appears that individual elements can function either co-operatively or independently to activate transcription.

[0284] The particular promoter that is employed to control the expression of a nucleic acid encoding the inhibitory peptide is not believed to be important, so long as it is capable of expressing the peptide in the targeted cell. Thus, where a human cell is targeted, it is preferable to position the nucleic acid coding the inhibitory peptide adjacent to and under the control of a promoter that is active in the human cell. Generally speaking, such a promoter might include either a human or viral promoter.

[0285] In various embodiments, the human cytomegalovirus (CMV) immediate early gene promoter, the SV40 early promoter and the Rous sarcoma virus long terminal repeat can be used to obtain high-level expression of various proteins. The use of other viral or mammalian cellular or bacterial phage promoters which are well-known in the art to achieve expression of peptides according to the present invention is contemplated as well, provided that the levels of expression are sufficient for a given purpose.

[0286] By employing a promoter with well-known properties, the level and pattern of expression of an antisense oligo- or polynucleotide can be optimized. Further, selection of a promoter that is regulated in response to specific physiologic signals can permit inducible expression of an

inhibitory protein. For example, a nucleic acid under control of the human PAI-1 promoter results in expression inducible by tumor necrosis factor. Additionally any promoter/enhancer combination (as per the Eukaryotic Promoter Data Base EPDB) also could be used to drive expression of a nucleic acid according to the present invention. Use of a T3, T7 or SP6 cytoplasmic expression system is another possible embodiment. Eukaryotic cells can support cytoplasmic transcription from certain bacterial promoters if the appropriate bacterial polymerase is provided, either as part of the delivery complex or as an additional genetic expression construct.

[0287] In certain embodiments of the invention, the delivery of a nucleic acid in a cell may be identified in vitro or in vivo by including a marker in the expression construct. The marker would result in an identifiable change to the transfected cell permitting easy identification of expression. Enzymes such as herpes simplex virus thymidine kinase (tk) (eukaryotic) or chloramphenicol acetyltransferase(CAT) (prokaryotic) may be employed.

[0288] One also may include a polyadenylation signal to effect proper polyadenylation of the transcript. The nature of the polyadenylation signal is not believed to be crucial to the successful practice of the invention, and any such sequence may be employed. For example, the SV40, β-globin or adenovirus polyadenylation signal may be employed. Also contemplated as an element of the expression cassette is a terminator. These elements can serve to enhance message levels and to minimize read through from the cassette into other sequences.

[0289] 2. Liposomal Formulations

[0290] In certain broad embodiments of the invention, the antisense oligo- or polynucleotides and/or expression vectors may be entrapped in a liposome. Liposomes are vesicular structures characterized by a phospholipid bilayer membrane and an inner aqueous medium. Multilamellar liposomes have multiple lipid layers separated by aqueous medium. They form spontaneously when phospholipids are suspended in an excess of aqueous solution. The lipid components undergo self-rearrangement before the formation of closed structures and entrap water and dissolved solutes between the lipid bilayers (Ghosh and Bachhawat, 1991). Also contemplated are cationic lipid-nucleic acid complexes, such as lipofectamine-nucleic acid complexes.

[0291] In certain embodiments of the invention, the liposome may be complexed with a hemagglutinating virus (HVJ). This has been shown to facilitate fusion with the cell membrane and promote cell entry of liposome-encapsulated DNA (Kaneda et al., 1989). In other embodiments, the liposome may be complexed or employed in conjunction with nuclear non-histone chromosomal proteins (HMG-1) (Kato et al., 1991). In yet further embodiments, the liposome may be complexed or employed in conjunction with both HVJ and HMG-1. In that such expression vectors have been successfully employed in transfer and expression of a polynucleotide in vitro and in vivo, then they are applicable for the present invention. Where a bacterial promoter is employed in the DNA construct, it also will be desirable to include within the liposome an appropriate bacterial polymerase.

[0292] "Liposome" is a generic term encompassing a variety of single and multilamellar lipid vehicles formed by the generation of enclosed lipid bilayers. Phospholipids are used for preparing the liposomes according to the present invention and can carry a net positive charge, a net negative charge or are neutral. Dicetyl phosphate can be employed to confer a negative charge on the liposomes, and stearylamine can be used to confer a positive charge on the liposomes.

[0293] Lipids suitable for use according to the present invention can be obtained from commercial sources. For example, dimyristyl phosphatidylcholine ("DMPC") can be obtained from Sigma Chemical Co., dicetyl phosphate ("DCP") is obtained from K & K Laboratories (Plainview, N.Y.); cholesterol ("Chol") is obtained from Calbiochem-Behring; dimyristyl phosphatidylglycerol ("DMPG") and other lipids may be obtained from Avanti Polar Lipids, Inc. (Birmingham, Ala.). Stock solutions of lipids in chloroform, chlorofom/methanol or t-butanol can be stored at about −20° C. Preferably, chloroform is used as the only solvent since it is more readily evaporated than methanol.

[0294] Phospholipids from natural sources, such as egg or soybean phosphatidylcholine, brain phosphatidic acid, brain or plant phosphatidylinositol, heart cardiolipin and plant or bacterial phosphatidylethanolamineare preferably not used as the primary phosphatide, i.e., constituting 50% or more of the total phosphatide composition, because of the instability and leakiness of the resulting liposomes.

[0295] Liposomes used according to the present invention can be made by different methods. The size of the liposomes varies depending on the method of synthesis. A liposome suspended in an aqueous solution is generally in the shape of a spherical vesicle, having one or more concentric layers of lipid bilayer molecules. Each layer consists of a parallel array of molecules represented by the formula XY, wherein X is a hydrophilic moiety and Y is a hydrophobic moiety. In aqueous suspension, the concentric layers are arranged such that the hydrophilic moieties tend to remain in contact with an aqueous phase and the hydrophobic regions tend to self-associate. For example, when aqueous phases are present both within and without the liposome, the lipid molecules will form a bilayer, known as a lamella, of the arrangement XY-YX.

[0296] Liposomes within the scope of the present invention can be prepared in accordance with known laboratory techniques. In one preferred embodiment, liposomes are prepared by mixing liposomal lipids, in a solvent in a container, e.g., a glass, pear-shaped flask. The container should have a volume ten-times greater than the volume of the expected suspension of liposomes. Using a rotary evaporator, the solvent is removed at approximately 40° C. under negative pressure. The solvent normally is removed within about 5 min to 2 hours, depending on the desired volume of the liposomes. The composition can be dried further in a desiccator under vacuum. The dried lipids generally are discarded after about 1 week because of a tendency to deteriorate with time.

[0297] Dried lipids can be hydrated at approximately 25-50 mM phospholipid in sterile, pyrogen-free water by shaking until all the lipid film is resuspended. The aqueous liposomes can be then separated into aliquots, each placed in a vial, lyophilized and sealed under vacuum.

[0298] In the alternative, liposomes can be prepared in accordance with other known laboratory procedures: the

method of Bangham et al. (1965), the contents of which are incorporated herein by reference; the method of Gregoriadis, as described in *DRUG CARRIERS IN BIOLOGY AND MEDICINE*, G. Gregoriadis ed. (1979) pp. 287-341, the contents of which are incorporated herein by reference; the method of Deamer and Uster (1983), the contents of which are incorporated by reference; and the reverse-phase evaporation method as described by Szoka and Papahadjopoulos (1978). The aforementioned methods differ in their respective abilities to entrap aqueous material and their respective aqueous space-to-lipid ratios.

[**0299**] The dried lipids or lyophilized liposomes prepared as described above may be reconstituted in a solution of nucleic acid and diluted to an appropriate concentration with an suitable solvent, e.g, DPBS. The mixture is then vigorously shaken in a vortex mixer. Unencapsulated nucleic acid is removed by centrifugation at 29,000× g and the liposomal pellets washed. The washed liposomes are resuspended at an appropriate total phospholipid concentration, e.g., about 50-200 mM. The amount of nucleic acid encapsulated can be determined in accordance with standard methods. After determination of the amount of nucleic acid encapsulated in the liposome preparation, the liposomes may be diluted to appropriate concentration and stored at 4° C. until use.

[**0300**] In a preferred embodiment, the lipid dioleoylphosphatidylcholine is employed. Nuclease-resistant oligonucleotides were mixed with lipids in the presence of excess t-butanol. The mixture was vortexed before being frozen in an acetone/dry ice bath. The frozen mixture was lyophilized and hydrated with Hepes-buffered saline (1 mM Hepes, 10 mM NaCl, pH 7.5) overnight, and then the liposomes were sonicated in a bath type sonicator for 10 to 15 min. The size of the liposomal-oligonucleotides typically ranged between 200-300 nm in diameter as determined by the submicron particle sizer autodilute model 370 (Nicomp, Santa Barbara, CA).

[**0301**] 3. Alternative Delivery Systems

[**0302**] Adenoviruses: Human adenoviruses are double-stranded DNA tumor viruses with genome sizes of approximate 36 kB (Tooze, 1981). As a model system for eukaryotic gene expression, adenoviruses have been widely studied and well characterized, which makes them an attractive system for development of adenovirus as a gene transfer system. This group of viruses is easy to grow and manipulate, and they exhibit a broad host range in vitro and in vivo. In lytically infected cells, adenoviruses are capable of shutting off host protein synthesis, directing cellular machineries to synthesize large quantities of viral proteins, and producing copious amounts of virus.

[**0303**] The El region of the genome includes E1A and E1B which encode proteins responsible for transcription regulation of the viral genome, as well as a few cellular genes. E2 expression, including E2A and E2B, allows synthesis of viral replicative functions, e.g. DNA-binding protein, DNA polymerase, and a terminal protein that primes replication. E3 gene products prevent cytolysis by cytotoxic T cells and tumor necrosis factor and appear to be important for viral propagation. Functions associated with the E4 proteins include DNA replication, late gene expression, and host cell shutoff. The late gene products include most of the virion capsid proteins, and these are expressed only after most of the processing of a single primary transcript from the major late promoter has occurred. The major late promoter (MLP) exhibits high efficiency during the late phase of the infection (Stratford-Perricaudet and Perricaudet, 1991).

[**0304**] As only a small portion of the viral genome appears to be required in cis (Tooze, 1981), adenovirus-derived vectors offer excellent potential for the substitution of large DNA fragments when used in connection with cell lines such as 293 cells. Ad5-transformed human embryonic kidney cell lines (Graham, et al., 1977) have been developed to provide the essential viral proteins in trans.

[**0305**] Particular advantages of an adenovirus system for delivering foreign proteins to a cell include (i) the ability to substitute relatively large pieces of viral DNA by foreign DNA; (ii) the structural stability of recombinant adenoviruses; (iii) the safety of adenoviral administration to humans; and (iv) lack of any known association of adenoviral infection with cancer or malignancies; (v) the ability to obtain high titers of the recombinant virus; and (vi) the high infectivity of adenovirus.

[**0306**] Further advantages of adenovirus vectors over retroviruses include the higher levels of gene expression. Additionally, adenovirus replication is independent of host gene replication, unlike retroviral sequences. Because adenovirus transforming genes in the E1 region can be readily deleted and still provide efficient expression vectors, oncogenic risk from adenovirus vectors is thought to be negligible (Grunhaus & Horwitz, 1992).

[**0307**] In general, adenovirus gene transfer systems are based upon recombinant, engineered adenovirus which is rendered replication-incompetent by deletion of a portion of its genome, such as El, and yet still retains its competency for infection. Sequences encoding relatively large foreign proteins can be expressed when additional deletions are made in the adenovirus genome. For example, adenoviruses deleted in both E1 and E3 regions are capable of carrying up to 10 kB of foreign DNA and can be grown to high titers in 293 cells (Stratford-Perricaudet and Perricaudet, 1991). Surprisingly persistent expression of transgenes following adenoviral infection has also been reported.

[**0308**] Other Viral Vectors as Expression Constructs. Other viral vectors may be employed as expression constructs in the present invention. Vectors derived from viruses such as vaccinia virus (Ridgeway, 1988; Baichwal and Sugden, 1986; Coupar et al., 1988) adeno-associated virus (AAV) (Ridgeway, 1988; Baichwal and Sugden, 1986; Hermonat and Muzycska, 1984) and herpes viruses may be employed. They offer several attractive features for various mammalian cells (Friedmann, 1989; Ridgeway, 1988; Baichwal and Sugden, 1986; Coupar et al., 1988; Horwich et al., 1990).

[**0309**] With the recent recognition of defective hepatitis B viruses, new insight was gained into the structure-function relationship of different viral sequences. In vitro studies showed that the virus could retain the ability for helper-dependent packaging and reverse transcription despite the deletion of up to 80% of its genome (Horwich et al., 1990). This suggested that large portions of the genome could be replaced with foreign genetic material. The hepatotropism and persistence (integration) were particularly attractive properties for liver-directed gene transfer. Chang et al.

recently introduced the chloramphenicol acetyltransferase (CAT) gene into duck hepatitis B virus genome in the place of the polymerase, surface, and pre-surface coding sequences. It was cotransfected with wild-type virus into an avian hepatoma cell line. Culture media containing high titers of the recombinant virus were used to infect primary duckling hepatocytes. Stable CAT gene expression was detected for at least 24 days after transfection (Chang et al., 1991).

[0310] Non-viral Methods. Several non-viral methods for the transfer of expression vectors into cultured mammalian cells also are contemplated by the present invention. These include calcium phosphate precipitation (Graham and Van Der Eb, 1973; Chen and Okayama, 1987; Rippe et al., 1990) DEAE-dextran (Gopal, 1985), lipofectamine-DNA complexes, and receptor-mediated transfection (Wu and Wu, 1987; Wu and Wu, 1988). Some of these techniquesmay be successfully adapted for in vivo or ex vivo use.

[0311] In one embodiment of the invention, the expression construct may simply consist of naked recombinant vector. Transfer of the construct may be performed by any of the methods mentioned above which physically or chemically permeabilize the cell membrane. For example, Dubensky et al. (1984) successfully injected polyomavirus DNA in the form of $CaPO_4$ precipitates into liver and spleen of adult and newborn mice demonstrating active viral replication and acute infection. Benvenisty and Neshif (1986) also demonstrated that direct intraperitoneal injection of $CaPO_4$ precipitated plasmids results in expression of the transfected genes. It is envisioned that DNA encoding an antisense prostate, bladder or breast disease marker construct may also be transferred in a similar manner in vivo.

[0312] Pharmaceutical Compositions and Routes of Administration

[0313] Where clinical application of liposomes containing antisense oligo- or polynucleotides or expression vectors is undertaken, it will be necessary to prepare the liposome complex as a pharmaceutical composition appropriate for the intended application. Generally, this will entail preparing a pharmaceutical composition that is essentially free of pyrogens, as well as any other impurities that could be harmful to humans or animals. One also will generally desire to employ appropriate buffers to render the complex stable and allow for uptake by target cells.

[0314] Aqueous compositions of the present invention comprise an effective amount of the antisense expression vector encapsulated in a liposome as discussed above, further dispersed in pharmaceutically acceptable carrier or aqueous medium. Such compositions also are referred to as inocula. The phrases "pharmaceutically or pharmacologically acceptable" refer to compositions that do not produce an adverse, allergic or other untoward reaction when administered to an animal, or a human, as appropriate.

[0315] As used herein, "pharmaceutically acceptable carrier" includes any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents and the like. The use of such media and agents for pharmaceutical active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active ingredient, its use in the therapeutic compositions is contemplated. Supplementary active ingredients also can be incorporated into the compositions.

[0316] Solutions of therapeutic compositions can be prepared in water suitably mixed with a surfactant, such as hydroxypropylcellulose. Dispersions also can be prepared in glycerol, liquid polyethylene glycols, mixtures thereof and in oils. Under ordinary conditions of storage and use, these preparations contain a preservative to prevent the growth of microorganisms.

[0317] The therapeutic compositions of the present invention are advantageously administered in the form of injectable compositions either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid prior to injection may also be prepared. These preparations also may be emulsified. A typical composition for such purpose comprises a pharmaceutically acceptable carrier. For instance, the composition may contain 10 mg, 25 mg, 50 mg or up to about 100 mg of human serum albumin per milliliter of phosphate buffered saline. Other pharmaceutically acceptable carriers include aqueous solutions, non-toxic excipients, including salts, preservatives, buffers and the like.

[0318] Examples of non-aqueous solvents are propylene glycol, polyethylene glycol, vegetable oil and injectable organic esters such as ethyloleate. Aqueous carriers include water, alcoholic/aqueous solutions, saline solutions, parenteral vehicles such as sodium chloride, Ringer's dextrose, etc. Intravenous vehicles include fluid and nutrient replenishers. Preservatives include antimicrobial agents, anti-oxidants, chelating agents and inert gases. The pH and exact concentration of the various components the pharmaceutical composition are adjusted according to well known parameters.

[0319] An effective amount of the therapeutic composition is determined based on the intended goal. The term "unit dose" or "dosage" refers to physically discrete units suitable for use in a subject, each unit containing a predetermined-quantity of the therapeutic composition calculated to produce the desired responses, discussed above, in association with its administration, i.e., the appropriate route and treatment regimen. The quantity to be administered, both according to number of treatments and unit dose, depends on the protection desired.

[0320] Precise amounts of the therapeutic composition also depend on the judgment of the practitioner and are peculiar to each individual. Factors affecting dose include physical and clinical state of the patient, the route of administration and the potency, stability and toxicity of the particular therapeutic substance. For the instant application, it is envisioned that the amount of therapeutic composition comprising a unit dose will range from about 5-30 mg of polynucleotide.

[0321] The following examples are included to demonstrate preferred embodiments of the invention. It should be appreciated by those of skill in the art that the techniques disclosed in the examples which follow represent techniques discovered by the inventors to function well in the practice of the invention, and thus may be considered to constitute preferred modes for its practice. However, those of skill in the art should, in light of the present disclosure, appreciate that many changes may be made in the particular embodiments which are disclosed and still obtain a like or similar result without departing from the spirit and scope of the invention.

[0322] J. Materials and Methods

[0323] 1. Application of RNA Fingerprinting to Discover Biomarkers for Prostate Cancers

[0324] RNA fingerprinting (according to Liang and Pardee, 1992; Welsh et al., 1992; Liang and Pardee, 1993) was applied to nucleic acids isolated from primary human prostate tumors or from prostate tumor derived cell lines that behave as tumor cells with different metastatic potential. The human prostate cancer cell lines examined in these studies were LnCaP, PC-3(pf), PC-3(mf), and DU-145. These cell lines vary in their metastatic potentials. LnCaP is only slightly metastatic while the other three cell lines are very aggressive and highly metastatic. The primary human prostate tumors used were of varying degrees of malignancy.

[0325] The cell lines were propagated in RPMI-1640 (GIBCO-BRL, Inc.) supplemented with 10% fetal bovine serum, 5 units/ml penicillin G, 5 pg/ml streptomycin, and Fungizone according to the supplier's directions. All antibiotics were purchased from GIBCO-BRL, Inc. Cells were harvested in late log phase of growth. RNA was isolated by the guanidinium thiocyanate method (Chomczynski and Sacchi, 1987). RNA was also isolated from solid prostate tumors by guanidinium thiocyanate extraction (Chomczynski and Sacchi, 1987), after the tumors were frozen and ground to a powder in liquid nitrogen.

[0326] After RNA isolation, the nucleic acids were precipitated with ethanol. The precipitates were pelleted by centrifugation and redissolved in water. The redissolved nucleic acids were then digested with RNase-free DNase I (Boehringer Mannheim, Inc.) following the manufacturer's instructions, followed by organic extraction with phenol:chloroform:isoamylalcohol (25:24:1) and reprecipitation with ethanol.

[0327] The DNase I treated RNA was then pelleted by centrifugation and redissolved in water. The purity and concentration of the RNA in solution was estimated by determining optical density at wave lengths of 260 nm and 280 nm (Sambrook et al., 1989). A small aliquot of the RNA was also separated by gel electrophoresis in a 3% formaldehyde gel with MOPS buffer (Sambrook et al., 1989) to confirm the estimation of concentration and to determine if the ribosomal RNAs were intact. This RNA, hereafter referred to as total cell RNA, was used in the studies described below.

[0328] 2. Methods Utilized in the Differential Display Technique

[0329] There were two kinds of RNA fingerprinting studies performed with the total cell RNA. The first of these kinds of studies followed the differential display protocol of Liang and Pardee (1992) except that it was modified by using 5' biotinylated primers for nonisotopic PCR product detection.

[0330] In these studies, 0.2 μg of total cell RNA was primed for reverse transcription with an anchoring primer composed of oligo dT, then two arbitrarily chosen nucleotides. The anchoring primers used in these studies were further modified to be biotinylated at the 5' end.

[0331] Reverse transcription was performed with 200 units of MMLV (Moloney Murine Leukemia Virus) reverse transcriptase (GIBCO/BRL) in the presence of 50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM $MgCl_2$, 10 mM DTT, 500 pM dNTP, 1 μM biotinylated anchored primer and 1 U/μl RNase inhibitor. The reaction mixture was incubated at room temperature for 10 minutes, then at 37° C. for 50 minutes. After reverse transcription the enzyme was denatured by heating to 65° C. for 10 minutes.

[0332] One tenth of the resulting reverse transcription reactions were then amplified by PCR using the same anchoring primer as was used in the reverse transcription step and a second oligonucleotide of arbitrarily chosen sequences. The PCR reaction contained 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 20 μM dNTP, 1.5 μM $MgCl_2$, 200 nM arbitrary decamer, 1 μM biotinylated anchored primer, and 1 unit of Taq DNA polymerase (Boehringer Mannheim) in a 40 μl volume. The amplification was performed in a thermal cycler (MJ Research) for 30 cycles with denaturing at 94° C. for 30 sec, annealing at 40° C. for 2 min, and extending at 72° C. for 30 sec.

[0333] The PCR products were then separated on a 6% TBE-urea sequencing gel (Sambrook et al., 1989) and detected by chemiluminescent reaction using the Seq-Light™ detection system (Tropix, Inc). Differentially appearing PCR products were excised from the gels, reamplified using the same primers used in the original amplification, and cloned using the TA cloning strategy (Invitrogen, Inc. and Promega, Inc.).

[0334] 3. Methods Utilized in the RNA Fingerprinting Technique

[0335] The second type of RNA fingerprinting studies performed more closely resembled the protocol of Welsh et al. (1992). This approach used a variation of the above as modified by the use of agarose gels and non-isotopic detection of bands by ethidium bromide staining (An et al., 1995). Total RNAs were isolated from the frozen prostate tissues or cultured cells as described (Chomczynski & Sacchi, 1987). Ten micrograms of total cellular RNAs were treated with 5 units of RNAse-free DNAse I (GIBCO/BRL) in 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 2 mM $MgCl_2$, and 20 units of RNAse inhibitor (Boehringer Mannheim). After extraction with phenol/chloroform and ethanol precipitation, the RNAs were redissolved in DEPC-treated water.

[0336] Two μg of each total cell RNA sample was reverse transcribed into cDNA using randomly selected hexamer primers and MMLV reverse transcriptase (GIBCO/BRL). PCR was performed using one or two arbitrarily chosen oligonucleotide primers (10-12mers). PCR conditions were: 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 50 μM dNTPs, 0.2 μM of primer(s), 1 unit of Taq DNA polymerase (GIBCO/BRL) in a final volume of 20 μl. The amplification parameters included 35 cycles of reaction with 30 sec denaturing at 94° C., 90 sec annealing at 40° C., and 60 sec extension at 72° C. A final extension at 72° C. was performed for 15 min. The resulting PCR products were resolved into a fingerprint by size separation by electrophoresis through 2% agarose gels in TBE buffer (Sambrook et al., 1989). The fingerprints were visualized by staining with ethidium bromide. No reamplification was performed.

[0337] Differentially appearing PCR products, that might represent differentially expressed genes, were excised from the gel with a razor blade, purified from the agarose using the Geneclean kit (Bio 101, Inc.), eluted in water and cloned

directly into plasmid vectors using the TA cloning strategy (Invitrogen, Inc., and Promega, Inc.). These products were not reamplified after the initial PCR fingerprinting protocol.

[0338] 4. Confirmation of Differential Expression by Relative Quantitative RT-PCR: Protocols for RT-PCR

[0339] a. Reverse Transcription

[0340] Five μg of total cell RNA from each tissue sample was reverse transcribed into cDNA. Reverse transcription was performed with 400 units of MMLV reverse transcriptase (GIBCO/BRL) in the presence of 50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM $MgCl_2$, 10 mM DTT, 500 μM dNTP, 50 ng random hexamers per microgram of RNA, and 1 U/μl RNase inhibitor. The reaction volume was 60 μl. The reaction mixture was incubated at room temperature for 10 minutes, then at 37° C. for 50 minutes. After reverse transcription the enzyme was denatured by heating to 65° C. for 10 minutes. After heat denaturation the samples were diluted with water to a final volume of 300 μl.

[0341] RT-PCR was utilized to examine mRNAs for differential expression. The sequences of oligonucleotides used as primers to direct the amplification of the various cDNA fragments are presented in Table 4.

[0342] b. Relative Quantitative RT-PCR With an Internal Standard

[0343] The concentrations of the original total cell RNAs were determined by measurement of $OD_{260/280}$ (Sambrook et al., 1989) and confirmed by examination of ribosomal RNAs on ethidium bromide stained agarose gels. It is required that all quantitative PCR reactions be normalized for equal amounts of amplifiable cDNA after the reverse transcription is completed. One solution to this is to terminate the reactions by driving the PCR reactions into plateau phase. This approach was utilized in some studies because it is quick and efficient. Lipocortin II was used as the internal standard or competitor. These PCRs were set up as:

[0344] Reagents: 200 μM each dNTP, 200 nM each oligonucleotide primer, 1× PCR buffer (Boehringer Mannheim including 1.5 mM $MgCl_2$), 3 μl diluted cDNA, and 2.5 units of Taq DNA polymerase/100 μl of reaction volume.

[0345] Cycling parameters: 30 cycles of 94° C. for 1 min; 55° C. for 1 min; and 72° C. for two min. Thermocyclers were either the MJ research thermocycler or the Stratagene Robocycler.

[0346] c. Relative Quantitative RT-PCR with an External Standard

[0347] There are three potential difficulties with the relative quantitative RT-PCR strategy described above. First, the internal standard must be roughly 4-10 times more abundant that the target for this strategy to normalize the samples. Second, because most of the PCR products are templated from the more abundant internal standard, the assay is less than optimally sensitive. Third, the internal standard must be truly unvarying. The result is that while the strategy described above is fast, convenient and applicable to samples of varying quality, it lacks sensitivity to modest changes in abundances.

[0348] To address these issues, a normalization was performed using both the 3-actin and asparagine synthetase mRNAs as external standards. These PCR reactions were performed with sufficient cycles to observe the products in the linear range of their amplification curves. Photographic negatives of gels of ethidium bromide stained PCR products were produced for each study. These negatives were scanned and quantified using a BioRad densitometer. The quantified data was then normalized for variations in the starting concentrations of amplifiable cDNA by comparing the quantified data from each study with that derived from a similar study which amplified a cDNA fragment copied from the 3-actin mRNA. Quantified data that had been normalized to beta actin were converted into bar graph representations.

K. EXAMPLES

Example 1

Relative Quantitative Reverse Transcriptase-polymerase Chain Reaction-A Method to Evaluate Novel Genes (ESTs) as Diagnostic Biomarkers

[0349] The reverse transcription-polymerase chain reaction (RT-PCR) protocols described in the following example were developed as a means to determine the relative abundances of mRNA species that are expressed in various tissues, organs and cells. The protocols used to meet this need must be robust, reproducible, relatively quantitative, sensitive, conservative in its use of resources, rapid and have a high throughput rate. Relative quantitative RT-PCR has the technical features that, in theory, meet all of these criteria. In practice there are six important barriers to implementing an RT-PCR based assay that compares the relative abundances of mRNA species. The protocol described herein addresses each of these six barriers and has permitted the realization of the potential of RT-PCR for this application. Although the present example is drawn to the identification and confirmation of differential expression in various physiological states in prostate tissue, the methods described herein may be applied to any type of tissue to provide a sensitive method of identifying differential expression.

[0350] The large majority of the candidate genes examined by this method are partial cDNA fragments that have been identified by RNA fingerprinting methodologies. This necessitated development of a relatively quantitative approach to independently confirm the differential expression of the mRNAs from which these partial cDNA fragments were derived. The key objective of the described screening protocol is the assessment of changes in the relative abundances of mRNA.

[0351] The gene discovery program described in the present disclosure is focused on analysis of human tissue and confirmation must be performed on the same biological material. Access to human tissue for isolation of RNA is limited. This limitation is especially problematic in Northern blots, the traditional means to determine differential gene expression. Northern blots typically consume roughly 20 μg of RNA per examined tissue per gene identified. This means that for the average size of tissue sample available, only 1-5 Northern blots can be performed before all of the RNA from a tissue sample is completely consumed. Clearly Northern blots are seriously limited for primary confirmation of discovered genes and consume extremely valuable biological resources required for gene discovery and characterization.

[0352] Because of such limitations on the amount of available tissue, and because of the need for high throughput and rapid turnaround of results, a two tiered assay protocol has been developed that is technologically grounded on reverse transcription (RT) of RNA into cDNA followed by amplification of specific cDNA sequences by polymerase chain reaction (PCR). This coupling of techniques is frequently referred to as RT-PCR.

[0353] One advantage of RT-PCR is that it consumes relatively small quantities of RNA. With 20 μg of RNA per examined sample, the amount of RNA required to perform a single Northern blot experiment, 50-200 RT-PCR assays can be performed with up to four data points per assay. Another advantage is a high throughput, eight independent experiments which examine eight different mRNA species for differential expression can be performed simultaneously in a single PCR machine with 96 wells. A single individual skilled in this technique can thereby examine and evaluate eight genes per day without significant time constraints. By comparison, even if RNA of sufficient quality and quantity were available to do this number of Northern blots, a similarly skilled individual performing Northern blots would be hard pressed to examine and evaluate eight genes per week. In addition to the lower throughput rate of Northern blots, eight Northern blots per week would require the consumption of about 400 μCi of ^{32}P per week. While not dangerous to use in the hands of a skilled individual, ^{32}P is certainly inconvenient to use. RT-PCR avoids the use of radioactive materials.

[0354] An additional advantage of RT-PCR over Northern blots as a technological platform for evaluating the relative expression of mRNA species is that RT-PCR is much less sensitive to differences in quality of the RNA being examined. The human tissues described herein were removed from patients for treatment purposes and were only incidentally saved for further studies. Hence the RNA, an extremely labile molecule, is expected to be at least partially degraded. Because the RNA is separated by size on a gel in the Northern blot assay, partially degraded RNA appears as a smear, rather than discrete bands. By contrast, RT-PCR amplifies only a section or domain of an RNA molecule, and as long as that portion is intact, the size or degradation state of the entire molecule is irrelevant. As a result, RNAs that are identical except that they vary by degree of partial degradation will give much more variable signals in a Northern blot than they will in an RT-PCR. When samples are of variable quality, as is often the case in human studies, the relative sensitivities of the techniques to variation in sample quality is an important consideration.

[0355] In the practice of this method, total cell RNA is first converted into cDNA using reverse transcriptase primed with random hexamers. This protocol results in a cDNA population in which each RNA has contributed according to its relative proportion in original total cell RNA. If two RNA species differ by ten fold in their original relative abundances in the total cell RNA, then the cDNA derived from these two RNAs will also differ by ten fold in their relative abundances in the resulting population of cDNA. This is a conservation of relative proportionality in the conversion of RNA to cDNA.

[0356] Another consideration is the relative rates of amplification of a targeted cDNA by PCR. In theory, the amount of an amplified product synthesized by PCR will be equal to M(EC). Where M is the mass of the targeted cDNA molecules before the beginning of PCR and C is the number of PCR cycle performed. E is an efficiency of amplification factor. This factor is complex and varies between 1 and 2. The important consideration in this assay is that over most of a PCR amplification, E will be nearly constant and nearly equal to 2. In PCR reactions that are identical in every way except the cDNAs being used as templates are derived from different total cell RNAs, then E will have the same value in each reaction. If a cDNA target has an initial mass of Ml in one PCR reaction and a mass of M_2 in another PCR reaction and if E has the same value in each reaction, then after C cycles of PCR there will be a mass of $M_1(E^C)$ of the amplified target in the first reaction and a mass of $M_2(E^C)$ of the amplified target in the second reaction. The ratios of these masses is unaltered by PCR amplification. That is $M1/M2=[M_1(E^C)]/M_2(E^C)$. Hence, there is a conservation of relative proportionality of amplified products during PCR.

[0357] Since both reverse transcription and PCR can be performed in such a way as to conserve proportionality, it is possible to compare the relative abundance of an mRNA species in two or more total cell RNA populations by first converting the RNA to cDNA and then amplifying a fragment of the cDNA derived from the specific mRNA by PCR. The ratio of the amplified masses of the targeted cDNA is very close to or identical to the ratios of the mRNAs in the original total cell RNA populations.

[0358] Six major challenges or barriers to be overcome in order to best use RT-PCR to quantitate the relative abundances of RNA are as follows:

[0359] 1.) Degradation of RNA must be minimized during RNA preparation.

[0360] 2.) Genomic DNA must be eliminated.

[0361] 3.) RNA must be free of contaminants that might interfere with reverse transcription.

[0362] 4.) The efficiency of RT is variable. cDNAs, not RNA, must be normalized for equal concentrations of amplifiable cDNA.

[0363] 5.) Limited linear range requires multiple sampling points in any amplification curve.

[0364] 6.) Tube to tube variability in PCR

[0365] It is the development of techniques to overcome these barriers and to provide a sensitive and accurate method of quantitative RT-PCR that is applicable to any tissue type or physiological state that is a part of the present invention.

[0366] The first three barriers to successful RT-PCR are all related to the quality of the RNA used in this assay. The protocols described in this section address the first two barriers as described in the last section. These are the requirements that degradation of RNA must be minimized during RNA preparation and that genomic DNA must be eliminated from the RNA.

[0367] Two preferred methods for RNA isolation are the guanididium thiocyanate method, which is well known in the art, and kits for RNA isolation manufactured by Qiagen, Inc. (Chatworth, Calif.), with the kits being the most pre-

ferred for convenience. Four protocols are performed on the RNA isolated by either method (or any method) before the RNA is be used in RT-PCR.

[0368] The first of these four protocols is digestion of the RNAs with DNaseI to remove all genomic DNA that was co-isolated with the total cell RNA. Prior to DNaseI digestion, the RNA is in a particulate suspension in 70% ethanol. Approximately 50 μg of RNA (as determined by OD260/280) is removed from the suspension and precipitated. This RNA is resuspended in DEPC treated sterile water. To this is added 10× DNaseI buffer (200 mM Tris-HCl; pH 8.4, 20 mM $MgCl_2$, 500 mM KCl), 10 units of RNase Inhibitor (GIBCO-BRL Cat#15518-012) and 20 units of DNaseI (GIBCO-BRL Cat# 18068-015). The volume is adjusted to 50 μl with additional DEPC treated water. The reaction is incubated at 37° C. for 30 minutes. After DNaseI digestion the RNAs are organic solvent-extracted with phenol and chloroform followed by ethanol precipitation. This represents the second ethanol precipitation of the isolated RNA. Empirical observations suggest that this repeated precipitation improves RNA performance in the RT reaction to follow.

[0369] Following DNaseI digestion, an aliquot of the RNA suspension in ethanol is removed and divided into thirds. A different procedure is performed on each one of the aliquot thirds. These three procedures are: (1). An $OD_{260/280}$ is obtained using a standard protocol and is used to estimate the amount of RNA present and its likely quality. (2). An aliquot is run out on an agarose gel, and the RNA is stained with ethidium bromide. Observation that both the 28S and 18S RNAs are visible as discreet bands and that there is little staining above the point at which the 28S rRNA migrates indicate that the RNA is relatively intact. While it is not critical to assay performance that the examined RNAs be completely free of partial degradation, it is important to determine that the RNA is not so degraded as to significantly effect the appearance of the 28S rRNA. (3). The total cell RNAs are run using a PCR-based test that confirms that the DNaseI treatment actually digested the contaminating genomic DNA to completion. It is very important to confirm complete digestion of genomic DNA because genomic DNA may act as a template in PCR reactions resulting in false positive signals in the relative quantitative RT-PCR assay described below. The assay for contaminating genomic DNA utilizes gene specific oligonucleotides that flank a 145 nucleotide long intron (intron #3) in the gene encoding Prostate Specific Antigen (PSA). This is a single copy gene with no pseudogenes. It is a member of the kallikrien gene family of serine proteases, but the oligonucleotides used in this assay are specific to PSA. The sequences of these oligonucleotides are:

5'CGCCTCAGGCTGGGGCAGCATT 3' (SEQ ID NO:79) and

5'ACAGTGGAAGAGTCTCATTCGAGAT 3' (SEQ ID NO:80).

[0370] In the assay for contaminating genomic DNA, 500 ng to 1.0 μg of each of the DNaseI treated RNAs are used as templates in a standard PCR (3540 cycles under conditions describe below) in which the oligonucleotides described above are used as primers. Human genomic DNA is used as the appropriate positive control. This DNA may be purchased from a commercial vender. A positive signal in this assay is the amplification of a 242 nucleotide genomic DNA specific PCR product from the RNA sample being tested as visualized on an ethidium bromide stained electrophoretic gel. There should be no evidence of genomic DNA as indicated by this assay in the RNAs used in the RT-PCR assay described below. Evidence of contaminating genomic DNA results in redigestion of the RNA with DNaseI and reevaluation of the DNase treated RNA by determining its $OD_{260/280}$ ratio, examination on electrophoretic gel and retesting for genomic DNA contamination using the described PCR assay.

[0371] The standard conditions used for PCR (as mentioned in the last paragraph) are:

[0372] 1× GIBCO-BRL PCR reaction buffer [20 mM Tris-Cl (pH 8.4), 50 mM KCl]

[0373] 1.5 mM $MgCl_2$

[0374] 200 μM each of the four dNTPs

[0375] 200 nM each oligonucleotide primer

[0376] concentration of template as appropriate

[0377] 2.5 units of Taq polymerase per 100 μl of reaction volume.

[0378] Using these conditions, PCR is performed with 35-40 cycles of:

[0379] 94° C. for 45 sec

[0380] 55°-60° C. for 45 sec

[0381] 72° C. for 1:00 minute.

[0382] The protocols described in the above section permit isolation of total cellular RNA that overcomes two of the six barriers to successful RT-PCR, ie. the RNA is acceptably intact and is free from contaminating genomic DNA.

[0383] Reverse transcriptases, also called RNA dependent DNA polymerases, as applied in currently used molecular biology protocols, are known to be less processive than other commonly used nucleic acid polymerases. It has been observed that not only is the efficiency of conversion of RNA to cDNA relatively inefficient, there is also several fold variation in the efficiency of cDNA synthesis between reactions that use RNAs as templates that otherwise appear indistinguishable. The sources of this variation are not well characterized, but empirically, it has been observed that the efficiencies of some reverse transcription (RT) reactions may be improved by repeated organic extractions and ethanol precipitations. This implies that some of the variation in RT is due to contaminates in the RNA templates. In this case, the DNaseI treatment described above may be aiding the efficiency of RT by subjecting the RNA to an additional cycle of extraction with phenol and chloroform and ethanol precipitation. Contamination of the template RNA with inhibitors of RT is an important barrier to successful RT that is partially overcome by careful RNA preparation and repeated organic extractions and ethanol precipitations.

[0384] Reverse transcription reactions are performed using the Superscript™ Preamplification System for First Strand cDNA Synthesis kit which is manufactured by GIBCO-BRL Life Technologies (Gaithersburg, Md.). Superscript™ is a cloned form of M-MLV reverse transcriptase that has been deleted for its endogenous RNaseH activity in

order to enhance its processivity. In the present example, the published protocols of the manufacturer are used for cDNA synthesis primed with random hexamers. cDNA synthesis may also be primed with a mixture of random hexamers (or other small oligonucleotides of random sequence) and oligo dT. The addition of oligo dT increases the efficiency of conversion of RNA to cDNA proximal to the polyA tail. As template, either 5 or 10 micrograms of RNA is used (depending on availability). After the RT reaction has been completed according to the protocol provided by GIBCO-BRL, the RT reaction is diluted with water to a final volume of 100 μl.

[0385] Even with the best prepared RNA and the most processive enzyme, there may be significant variation in the efficiency of RT. This variation would be sufficiently great that cDNA made in different RTs could not be reliably compared. To overcome this possible variation, cDNA populations made from different RT reactions may be normalized to contain equal concentrations of amplifiable cDNA synthesized from mRNAs that are known not to vary between the physiological states being examined. In the present examples, cDNAs made from total cell RNAs are normalized to contain equal concentrations of amplifiable β-actin cDNA.

[0386] One μl of each diluted RT reaction is subjected to PCR using oligonucleotides specific to β-actin as primers. These primers are designed to cross introns, permitting the differentiation of cDNA and genomic DNA. These β-actin specific oligonucleotides have the sequences:

5'CGAGCTGCCTGACGGCCAGGTCATC 3' (SEQ ID NO:81) and

5'GAAGCATTTGCGGTGGACGATGGAG 3' (SEQ ID NO:82)

[0387] PCR is performed under standard conditions as described previously for either 19 or 20 cycles. The resulting PCR product is 415 nucleotides in length. The product is examined by PCR using agarose gel electrophoresis followed by staining with ethidium bromide. The amplified cDNA fragment is then visualized by irradiation with ultra violet light using a transilluminator. A white light image of the illuminated gel is captured by an IS-1000 Digital Imaging System manufactured by Alpha Innotech Corporation. The captured image is analyzed using either version 2.0 or 2.01 of the software package supplied by the manufacturer to determine the relative amounts of amplified β-actin cDNA in each RT reaction.

[0388] To normalize the various cDNAs, water is added to the most concentrated cDNAs as determined by the assay described in the last paragraph. PCR using 1 μl of the newly rediluted and adjusted cDNA is repeated using the β-actin oligonucleotides as primers. The number of cycles of PCR must be increased to 21 or 22 cycles in order to compensate for the decreased concentrations of the newly diluted cDNAs. With this empirical method the cDNAs can be adjusted by dilution to contain roughly equal concentrations of amplifiable cDNA. Sometimes this process must be repeated to give acceptable final normalization. By dividing the average optical density of all observed bands by that of a particular band, a normalization statistic can be created that will permit more accurate comparisons of the relative abundances of RNAs examined in the normalized panel of cDNAs. A representative gel is shown if **FIG. 12**. An analysis of the data is shown in Table 2.

[0389] Once the normalization statistics are derived, PCR may be performed using different gene specific oligonucleotides as primers to determine the relative abundances of other mRNAs as represented as cDNAs in the normalized panel of diluted RT reaction products.

[0390] Most mRNA species are not differentially expressed. An example of a differentially expressed message is the mRNA encoding the transmembrane tyrosine kinase receptor, Hek, that is significantly up regulated in BPH as compared to normal prostates.

[0391] In an examination of the relative abundance levels of Hek mRNA, the normal and tumor specimens were examined as pools. Low level expression was observed in the pool of normal prostate tissues relative to that observed in BPH. By normalizing these values to the β-actin standard using the normalization statistics, it is possible to quantify this difference in the relative abundances of Hek mRNA. These normalized data are displayed graphically in the bar graph shown in **FIG. 13**. Most but not all of the BPH specimens showed elevated abundances of Hek mRNA relative to a pool of normal prostates. On average, the abundance of Hek mRNA was observed to be 2.9 fold higher in the BPH specimens than in an average normal prostate gland as represented by the pool of normal glands.

[0392] While these observations are consistent with many similar studies that examined Hek expression using other tissue samples and cDNAs, they vary from observations described in the next section in which an RT-PCR assay is discussed that uses pooled cDNAs and is more likely to capture data from PCRs while in the linear portions of their amplification curves. It was fairly obvious from the data obtained in the Hek study that at least some of the RT-PCR reactions were not in the linear portions of their amplification curves when the data was captured. This was concluded from observation that the intensity of the bands from BPH9 slightly decreased from a sample taken at 35 to a sample taken at 40 cycles. To a lesser extent this was true for other samples as well. This is a strong indication that the PCRs had left the linear portions of their amplification curves. While this observation limits the qualitative value of this experiment, it does not necessarily limit the ability of the assay to determine qualitative differences in mRNA abundances. The error caused by observing PCRs after the linear portion of PCR is in the direction of quantitatively underestimating mRNA abundance differences. It is still valid to conclude that Hek is up regulated in many prostate glands with BPH even if the absolute fold increase in abundance can not be determined. By looking at individuals, it is possible to examine questions as to what portion of individuals of a particular physiologic class, i.e. individuals with BPH, similarly regulate the mRNA being examined. To determine quantitative differences in mRNA expression, it is necessary that the data is collected in the linear portion of the respective PCR amplification curves. This requirement is met in the assay described in following paragraphs.

[0393] The last two barriers to RT-PCR are addressed in the sections that follow involving the use of pooled cDNAs as templates in RT-PCR. In practice, the protocols using pooled templates are usually performed before the protocol described above.

[0394] There are two additional barriers to relative mRNA quantitation with RT-PCR that frequently compromise interpretations of results obtained by this method. The first of these involves the need to quantify the amplification products while the PCR is still in the linear portion of the process where "E" behaves as a constant and is nearly equal to two. In the "linear" portion of the amplification curve, the log of the mass of the amplified product is directly proportional to the cycle number. At the end of the PCR process, "E" is not constant. Late in PCR, "E" declines with each additional cycle until there is no increase in PCR product mass with additional cycles. The most important reason why the efficiency of amplification decreases at high PCR cycle number, may be that the concentration of the PCR products becomes high enough that the two strands of the product begin to anneal to each other with a greater efficiency than that at which the oligonucleotide primers anneal to the individual product strands. This competition between the PCR product strands and the oligonucleotide primers creates a decrease in PCR efficiency. This part of the PCR where the efficiency of amplification is decreased is called the "plateau" phase of the amplification curve. When "E" ceases to behave as a constant and the PCR begins to move towards the plateau phase, the conservation of relative proportionality of amplified products during PCR is lost. This creates an error in estimating the differences in relative abundance of an mRNA species occurring in different total cell RNA populations. This error is always in the same direction, in that it causes differences in relative mRNA abundances to appear less than they actually are. In the extreme case, where all PCRs have entered the plateau phase, this effect will cause differentially expressed mRNAs to appear as if they are not differentially expressed at all.

[0395] To control for this type of error, it is important that the PCR products be quantified in the linear portion of the amplification curve. This is technically difficult because currently used means of DNA quantitation are only sensitive enough to quantify the PCR products when they are approaching concentrations at which the product strands begin to compete with the primers for annealing. This means that the PCR products can only be detected at the very end of the linear range of the amplification curve. Predicting in advance at what cycle number the PCR products should be quantified is technically difficult.

[0396] Practically speaking, it is necessary to sample the PCR products at a variety of cycle numbers that are believed to span the optimum detection range in which the products are abundant enough to detect, but still in the linear range of the amplification curve. It is impractical to do this in a study that involves large numbers of samples because the number of different PCR reactions and/or number of different electrophoretic gels that must be run becomes prohibitively large.

[0397] To overcome these limitations, a two tiered approach has been designed to relatively quantitate mRNA abundance levels using RT-PCR. In the first tier, pools of cDNAs produced by combining equal amounts of normalized cDNA are examined to determine how mRNA abundances vary in the average individual with a particular physiological state. This reduces the number of compared samples to a very small number such as two to four. In the studies described herein, three pools are examined. These are pools of normal prostates, those with BPH and a variety of prostate tumors. Each pool may contain a large number of individuals. While this approach does not discriminate differences between individuals, it can easily discern broad patterns of differential expression. The great advantage of examining pooled cDNAs is that it permits many duplicate PCR reactions to be simultaneously set up.

[0398] The individual duplicates can be harvested and examined at different cycle numbers of PCR. In studies described below, four duplicate PCR reactions were set up. One duplicate was terminated at 31, 34, 37, and 40 PCR cycles. Occasionally, PCR reactions were also terminated at 28 cycles. Examining the PCRs at different cycle numbers yielded the following benefits. It is very likely that at least one of the RT-PCRs will be in the optimum portion of the amplification curves to reliably compare relative mRNA abundances. In addition, the optimum cycle number will be known, so that studies with much larger sample sizes, such as the studies Hek described above, are much more likely to succeed. This is the second tier of a two tiered approach that has been taken to relatively quantitate mRNA abundance levels using RT-PCR. Doing the RT-PCR with the pooled samples permits much more efficient application of RT-PCR to the samples derived from individuals. A further benefit, also as discussed below, tube to tube variability in PCR can be discounted and controlled because most studies yield multiple data points due to duplication.

[0399] Like the previously described protocol involving individuals, the first step in this protocol is to normalize the pooled samples to contain equal amounts of amplifiable cDNA. This is done using oligonucleotides that direct the amplification of β-actin. In this example, a PCR amplification of a cDNA fragment derived from the β-actin mRNA from pools of normal prostates, glands with BPH and prostate tumors was performed. This study was set up as four identical PCR reactions. The products of these PCRs were collected and electrophoresed after 22, 25, 28 and 31 PCR cycles. Quantitation of these bands using the IS 1000 system shows that the PCRs are still in the linear ranges of their amplification curves at 22, 25 and 28 cycles but that they have left linearity at 31 cycles. This is known because the ratios of the band intensities remain constant and internally consistent for the data obtained from 22, 25 and 28 cycles, but these ratios become distorted at 31 cycles. This quantitation will also permit the derivation of normalizing statistics for the three pools relative to each other in exactly the same manner as was done previously for individuals (Table 2).

[0400] This study is then repeated using gene specific primers for a gene other than β-actin. For purposes of comparison, the mRNAs examined were the same as were previously shown, Hek. As was done previously for the samples derived from individuals, the intensities of the relevant bands were quantitated using the IS 1000 and normalized to the β-actin signals.

[0401] For Hek, the data deserves more interpretation. While the Hek derived PCR product was observable at 34 cycles of PCR, at 40 cycles, the Hek derived PCR product was present as a bold band in the PCRs using either the pooled BPH samples or pooled prostate tumor samples as templates. The Hek band obtained when a pool of normal prostates is examined is barely visible. It is clear that Hek is more abundantly expressed in BPH and prostate tumors than

it is in normal glands. Quantitation and normalization of this data as described previously was performed and shown in the bar graph in **FIG. 14**.

[**0402**] The central question to be answered in analyzing this data is whether the PCRs have been examined in the linear portions of their amplification curves. A test for this can be devised by determining if the proportionality of the PCR products has been conserved as PCR cycle number has increased. At 34 cycles, the Hek product is observed at 5.77 and 4.375 relative abundance units respectively for the pooled BPH and cancer samples as shown in **FIG. 14**. The ratio of these values is 1.32. Similarly, at 37 cycles the values for BPH and cancer are 23.1 and 17.5. The ratio of these values is also 1.32. This is strong evidence that the PCRs were in the linear portions of their amplification curves when these observations were made. This is a better conservation of proportionality than is frequently observed. In some studies, data was excepted when the rations were similar but not identical. This conservation of proportionality is lost at 40 cycles. The ratio of the BPH and cancer values has increased to 1.85. This indicates that these PCRs are nearing the plateau phases of their amplification curves. Further evidence that the plateau phase is nearing can be directly observed in the relative increases in the numerical data observed in this study. From 34 to 37 cycles of PCR the mass of the observed PCR products increased 4.0 fold in both the BPH and cancer reactions. Similar calculations of the increase in signals between 37 and 40 cycles indicate a 3.1 fold increase in the BPH reactions and only a 2.2 fold increase for the cancer reactions. In both cases, "E" is declining, and the reactions are nearing their plateau phases.

[**0403**] For the reactions that attempted to amplify Hek cDNA from a pool of normal prostates, a band was only observed at 40 cycles. Since the BPH and cancer reactions had left their linear phases, direct numerical quantitation of the fold increase in abundance between normal, BPH and cancer is not possible. It is, however, valid to conclude that Hek mRNA is more abundant in samples derived from BPH or prostate tumors than it is in normal prostate glands. It may also be true that Hek is more abundant in the average BPH specimen than it is in the average prostate tumor. This has been observed in many studies including the one shown here, but the difference in relative expression of Hek between BPH and prostate cancers is always small, as it is here. It is possible that the higher levels of expression in the tumor pool relative to normal prostates may be due to BPH tissue contaminating the tumor specimens. Alternatively, it may be due to higher Hek expression in the tumors themselves. Examination of tissue by in situ hybridization or by immunohistochemical methods may be required to distinguish between these possibilities.

[**0404**] The final major barrier to quantifying relative mRNA abundances with RT-PCR is tube to tube variability in PCR. This can result from many factors, including unequal heating and cooling in the thermocycler, imperfections in the PCR tubes and operator error. To control for this source of variation, the Cole-Parmer digital thermocouple Model # 8402-00 was used to calibrate the thermocyclers used in these studies. Only slight variations in temperature were observed. To rigorously demonstrate that PCR tube to tube variability was not a factor in the studies described above, 24 duplicate PCRs for β-actin using the same cDNA as template were performed. These PCR tubes were scattered over the surface of a 96 well thermocycler, including the corners of the block where it might be suspected the temperature might deviate from other areas. Tubes were collected at various cycle numbers. Nine tubes were collected at 21 cycles. Nine tubes were collected at 24 cycles, and six tubes were collected at 27 cycles. Quantitation of the intensities of the resulting bands with the IS 1000 system determined that the standard error of the mean of the PCR product abundances was ±13%. This is an acceptably small number to be discounted as a major source of variability in an RT-PCR assay.

[**0405**] The RT-PCR protocol examining pooled cDNAs is internally controlled for tube to tube variability that might arise from any source. By examining the abundance of the PCR products at several different cycle numbers, it can be determined that the mass of the expected PCR product is increasing appropriately with increasing PCR cycle number. Not only does this demonstrate that the PCRs are being examined in the linear phase of the PCR, where the data is most reliable, it demonstrates that each reaction with the same template is consistent with the data from the surrounding cycle numbers. If there was an unexplained source of variation, the expectation that PCR product mass would increase appropriately with increasing cycle number would not be met. This would indicate artifactual variation in results. Internal duplication and consistency of the data derived from different cycle numbers controls for system derived variation in tube to tube results.

[**0406**] As described in the preceding paragraphs, the RT-PCR protocol using pooled cDNA templates overcomes the last two barriers to effective relative quantitative RT-PCR. These barriers are the need examine the PCR products while the reactions are in the linear portions of their amplification curves and the need to control tube to tube variation in PCR. The described protocol examines PCR products at three to four different cycle numbers. This insures that the PCRs are quantitated in their linear ranges and, as discussed in the last paragraph, controls for possible tube to tube variation.

[**0407**] One final question is whether β-actin is an appropriate internal standard for mRNA quantitation. p-actin has been used by many investigators to normalize mRNA levels. Others have argued that β-actin is itself differentially regulated and therefore unsuitable as an internal normalization standard. In the protocols described herein differential regulation of β-actin is not a concern. More than fifty genes have been examined for differential expression using these protocols. Fewer than half were actually differentially expressed. The other half were regulated similarly to β-actin within the standard error of 13%. Either all of these genes are coordinately differentially regulated with β-actin, or none of them are differentially regulated. The possibility that all of these genes could be similarly and coordinately differentially regulated with β-actin seems highly unlikely. This possibility has been discounted.

[**0408**] β-actin has also been criticized by some as an internal standard in PCRs because of the large number of pseudogenes of β-actin that occur in mammalian genomes. This is not a consideration in the described assays because all of the RNAs used herein are demonstrated to be free of contaminating genomic DNA by a very sensitive PCR based assay. In addition, the cycle number of PCR needed to detect β-actin cDNA from the diluted RT reactions, usually

between 19 and 22 cycles, is sufficiently low to discount any contribution that genomic DNA might make to the abundance of amplifiable β-actin templates.

TABLE 2

Raw Numerical Data Captured on the IS1000 and Normalization by Comparison to B-Actin

Type of Tissue	Raw Data B-Actin	Raw Data corrected for background	Normalizing Statistic	Raw Data for Hek (UC205)	Normalized Data for Hek (UC205)
Normal Pool 1	16	11	1.42		
Normal Pool 2	35	30	0.52		
Total normal Pool	25.5	20.5	0.76	22	16.72
BPH1	13	8	1.96	37	72.52
BPH2	27	22	0.71	10	17.1
BPH3	36	31	0.5	44	22
BPH4			1	31	31
BPH5	18	13	1.2	24	28.8
BPH6	15	10	1.56	41	63.96
BPH7	17	12	1.3	51	66.3
BPH8	21	16	0.975	39	38
BPH9	11	6	2.6	50	130
BPH10	17	12	1.3	14	18.2
BPH Pool	19.4	14.4	1.08		
Cancer1	13	8	1.96		
Cancer2	18	13	1.2		
Cancer3	22	17	0.92		
Cancer4	25	20	0.78		
Cancer5	29	24	0.65		
Cancer6			1		
Cancer7	22	17	0.92		
Cancer8	22	17	0.92		
Cancer9	15	10	1.56		
Cancer10	16	11	1.42		
Cancer11	11	6	2.6		
Cancer (Met)12	34	29	0.54		
Cancer Pool	20.6	15.7	1	41	41
No template	5	0			
Background	5	0			
Total	497.9	377.9			
Average	20.6	15.6			

Example 2

Identification of Markers of Prostate Disease by RNA Fingerprinting

[0409] The technique of RNA fingerprinting was used to identify differentially expressed RNA species isolated from primary human prostate tumors or human prostate cancer cell lines grown in culture as described above. About 400 bands were observed in these studies. A number of these appeared to be differentially expressed, and were cloned as described above.

[0410] Slot blots of total cell RNA probed with riboprobes indicated that six of the clones were differentially expressed. These six cloned PCR products chosen for further analysis were named UC Band #25 (SEQ ID NO:1), UC Band #27 (SEQ ID NO:2), UC Band #28 (SEQ ID NO:3, SEQ ID NO:83 and SEQ ID NO:85), UC Band #31 (SEQ ID NO:4), UC Band #32 (SEQ ID NO:7) and UC Band #33 (SEQ ID NO:5).

[0411] Studies were performed using total cell RNA isolated from human prostate glands and primary human prostate tumor samples. The prostate disease markers observed to be differentially expressed in this series of studies include UC Band #25 (SEQ ID NO:1), UC Band #28 (SEQ ID NO:3, SEQ ID NO:83 and SEQ ID NO:85), UC Band #31 (SEQ ID NO:4), UC Band #32 (SEQ ID NO:7) and UC Band #33 (SEQ ID NO:5). Differential expression of these gene products in human prostate tumors compared with benign and normal prostate tissues was confirmed by quantitative RT-PCR, as described below.

[0412] DNA sequence determination indicated that UC Band #25 (SEQ ID NO:1), UC Band #27 (SEQ ID NO:2), UC Band #28 (SEQ ID NO:3, SEQ ID NO:83 and SEQ ID NO:85), UC Band #31 (SEQ ID NO:4) and UC Band #33 (SEQ ID NO:5) were previously unknown genes. UC Band #32 (SEQ ID NO:7) was derived from the mRNA of fibronectin. The result with the latter gene product is interesting because urinary fibronectin has been proposed as a potential biomarker for prostatic cancer (Webb & Lin, 1980.) The levels of expression for UC Band #25, UC Band #27, UC Band #28, UC Band #31, UC Band #33, fibronectin and lipocortin II were analyzed by the quantitative RT-PCR protocol in samples of normal, benign and malignant prostate glands. The results for UC Band #25 (SEQ ID NO:1), (**FIG. 1**), UC Band #27 (SEQ ID NO:2), (**FIG. 2**), UC Band #28 (SEQ ID NO:3, SEQ ID NO:83 and SEQ ID NO:85), (**FIG. 3**), UC Band #31 (SEQ ID NO:4), (**FIG. 4**), and UC Band #33 (SEQ ID NO:5), (**FIG. 6**), all show an increased level of expression in prostate carcinomas (NB, T and LM) compared with benign (B) and normal (N) prostate samples.

[0413] The results for UC Band #28 (**FIG. 3**) and UC Band #33 (**FIG. 6**) are particularly striking. These clones are expressed at very low levels in normal or benign prostate, and at significantly higher levels in metastatic and nonmetastatic prostate cancers. As such, they would provide excellent markers for the detection of malignant prostate tumors in biopsy samples containing a mixture of normal, benign and malignant prostate. The skilled practitioner will realize that all of these clones, particularly UC Band #28 and UC Band #33, have utility for the detection and diagnosis of prostate cancer, and such uses are included within the scope of the present invention.

[0414] The RT-PCR analysis for fibronectin (UC Band #32, **FIG. 5**) is also interesting. This marker appears to only be expressed in normal prostate samples, and is present at very low levels in either benign or malignant prostate (**FIG. 5**). The down regulation of fibronectin expression in BPH is a novel result. This observation is surprising in light of the previous report that fibronectin is a potential marker for prostate cancer. (Webb and Lin, 1980.) Those experienced in the art will realize that loss of fibronectin expression in BPH is of utility in diagnosing and detecting this condition in patients. The mRNA for lipocortin II, while differentially expressed in the cell lines was not differentially expressed in tumors.

[0415] Further RNA fingerprinting studies were done to identify genes that are differentially regulated at the level of mRNA transcription in normal prostate glands, glands with BPH, prostate tumors and metastases of prostate tumors. Differential expression was confirmed by relative quantitative RT-PCR. The oligonucleotides used are listed in Table

4. These studies resulted in the discovery of additional sequences that were differentially regulated. These sequences are designated herein as UC38, SEQ ID NO:10; UC40, SEQ ID NO: 1; UC41, SEQ ID NO:12; UC43, SEQ ID NO:19; UC45, UC46, UC47, matches GenBank Accession #M34840, prostatic acid phosphatase Nt 901-2095; UC201, SEQ ID NO: 13; UC202, UC203, UC204 (matches GB#Z28521 and GB# D42055), SEQ ID NO:20; UC205 (Humhek, GB# H8394, sense strand), SEQ ID NO:14; UC206 (antisense strand), UC207 (sense strand), SEQ ID NO:15; UC208 (sense strand), UC209, SEQ ID NO:16; UC210 (sense strand), SEQ ID NO:17; UC211 (antisense strand), SEQ ID NO:21; UC212 (sense strand), SEQ ID NO:22; and UC213 (sense strand, matches GB# T07736), SEQ ID NO:23. Of these UC38, UC41, UC47 and UC211 are more abundant in tumors and are potential tumor markers. UC40, UC205 and UC207 are more abundant in BPH. UC43 is more abundant in normal and BPH glands and is a potential tumor suppressor. UC201 and UC210 are more abundant in some tumors and are potential progression markers. UC212 is more abundant in BPH and perhaps in some tumors. UC209 is down regulated in some tumors and is a possible suppressor of progression, and UC213 is down regulated in tumors.

[0416] Those experienced in the art will recognize that the genes and gene products (RNAs and proteins) for the above described markers of prostate disease and normal prostate marker are included within the scope of the invention herein described. Those experienced in the art will also recognize that the diagnosis and prognosis of prostatic cancer by detection of the nucleic acid products of these genes are included within the scope of the present invention.

[0417] 3. Detection of Differentially Expressed RNA Species Using Primers Specific for TGF-β and Cyclin A

[0418] Relative quantitative RT-PCR with an external standard proved to be a powerful means to examine mRNAs for differential expression in prostate cancer. Other genes were examined for differential expression by these means. These were selected because they were either known to be up regulated as a consequence of transformation or could be hypothesized to be up regulated as a consequence of transformation.

[0419] The results of two of these assays are included here. They show that TGF-β1 (**FIG. 7**) and cyclin A (**FIG. 8**) are both up regulated in prostate cancer relative to normal and benign glands. The cyclin A result is particularly interesting because this protein is known to be a positive regulator of cell cycle progression. It has occasionally been shown to be up regulated in some cancers, but this is the first observation of cyclin A being up regulated in most or all tumors derived from a single organ source (prostate). The sequence of cyclin A is identified as SEQ ID NO:8. Those skilled in the art will recognize that the genes and gene products (RNAs and proteins), including the diagnosis and prognosis of prostatic cancer by detection of the RNA products for these two genes, are included within the scope of the invention herein described.

Example 4

Identification of Markers of Prostate Disease Using Probes Specific for a Truncated Form of Her2/neu

[0420] In the studies described below, a relative quantitative version of RT-PCR was performed. The oligonucleotides used as primers to direct the amplification by PCR of the various cDNA fragments are given in Table 5. Briefly, three oligonucleotide primers were designed, which are identified in Table 5 as Neu5', SEQ ID NO:44; Neu3', SEQ ID NO:71; and NeuT3', SEQ ID NO:72. Neu5' anneals to antisense sequence for both the full length and truncated form of the Her2/neu mRNAs at a position 5' of an alternate RNA processing site (see **FIG. 9**). Neu3' anneals to the sense strand of the full length Her2/neu mRNA at a position just 3' of the transmembrane domain (**FIG. 9**).

[0421] In an RT-PCR assay using Neu5' and Neu3' as primers, a 350 base pair long amplification product was generated using the full length mRNA as a template. Using these primers, a cDNA fragment can not be generated from the truncated mRNA because Neu3' will not anneal to this mRNA or its cDNA. The third oligonucleotide primer, NeuT3', anneals to the sense strand of the 3' untranslated region of the truncated form of the Her2/neu mRNA and cDNA (**FIG. 9**). In an RT-PCR assay using Neu5' and NeuT3' as primers, a 180 base pair long cDNA fragment was amplified using the truncated mRNA as a template. This primer pair can not direct the amplification of a fragment of the full length Her2/neu mRNA because NeuT3' will not anneal to the full length transcript.

[0422] The results of relative quantitative RT-PCR clearly showed that the relative abundance of the Her2/neu mRNA is increased in prostate cancers as compared to either normal prostate or benign prostatic hyperplasia (**FIG. 10**). These data were generated from a densitometry scan of a photographic negative of a photograph of an ethidium bromide stained gel. The raw densitometry scan data were then normalized to a similar scan of a PCR amplification from the same template of β-actin, a gene whose expression is not expected to vary as a function of transformation or tumor progression. The results are completely consistent with the increased abundance of Her2/neu protein in prostate tumors that was previously described in the literature reviewed above.

[0423] A relative quantitative RT-PCR assay examining the relative abundance of the truncated form of the Her2/neu mRNA (SEQ ID NO:9) in various prostate tissues was also performed. This assay was similar to that shown above for the full length Her2/neu transcript. The data from this study was quantified and normalized to 13-actin and is displayed in **FIG. 11**.

[0424] As shown in **FIG. 11**, the relative abundance of this truncated transcript was significantly increased in prostate cancers as compared to normal and benign prostate. As discussed in a previous section, this truncated form of the Her2/neu mRNA has been previously described in breast, ovarian and gastric tumors. This is the first report of differential expression of a truncated form of Her2/neu as a biomarker for prostate cancer.

[0425] As indicated in Scott et al. (1993), expression of this truncated Her2/neu mRNA may alter the cellular behavior of cancer cells to the detriment of patients. Those skilled in the art will recognize that therapeutic treatment of prostate cancer targeted towards the gene products (including mRNAs and proteins) of the truncated form of Her2/neu is included within the scope of this invention.

TABLE 3

Genes Whose mRNAs have Abundances that Vary in Prostate Cancer Relative to Normal and Benign Glands

Name of cDNA Fragment	Sequence Determined	Confirmed by RT-PCR	Previously Known	SEQ ID NO:
UC Band #25	Yes	Yes	No	1
UC Band #27	Yes	Yes	No	2
UC Band #28	Yes	Yes	No	3, 83, 84
UC Band #31	Yes	Yes	No	4
UC Band #32	Yes	Yes	fibronectin	7
US Band #33	Yes	Yes	No	5
Cyclin A	Yes	Yes	Cyclin A	8
Trunc. HER2/neu	Yes	Yes	Tru. HER2/ neu	9
UC Band #38	Yes	Yes	No	10
UC Band #40	Yes	Yes	No	11
UC Band #41	Yes	Yes	No	12
UC Band #43	Yes	Yes	No	19
UC Band #47	Yes	Yes	Prostatic Acid Phosphatase	47
UC Band #201	Yes	Yes	No	13
UC Band #204	Yes	Yes	GB #Z28521 and GB #D42055	20
UC Band #205	Yes	Yes	Humhek	14
UC Band #207	Yes	Yes	No	15
UC Band #209	Yes	Yes	No	16
UC Band #210	Yes	Yes	No	17
UC Band #211	Yes	Yes	No	21
UC Band #212	Yes	Yes	No	22
UC Band #213	Yes	Yes	GB #T07736	23
UC Band #214	Yes	Yes	No	45
UC Band #215	Yes	Yes	No	46

[0426] Table 4. Oligonucleotides used in the relative quantitative RT-PCR portion of these studies.

[0427] Oligonucleotides used to examine the expression of genes:

[0428] Cyclin A (SEQ ID NO:8)

5'TGCGTTCACCATTCATGTGGATGAAG-CAG3', SEQ ID NO:26

5'CTCCTACTTCAACTAACCAGTCCACGAG3', SEQ ID NO:27

[0429] UC Band #25 (SEQ ID NO:1)

5'GATGCTTTGAAGTTATCTCTCTTGG3', SEQ ID NO:28

5'ATCAGTGTGGCAGATATAATGGACC3', SEQ ID NO:29

[0430] UC Band #27 (SEQ ID NO:2)

5'GCCCCAAATGCCAGGCTGCACTGAT3', SEQ ID NO:30

5'GCCAGAAGACAAGAGTGTGAGCTT3', SEQ ID NO:31

[0431] UC Band #28 (SEQ ID NO:3)

5'GCTTCAGGGTGGTCCAATTAGAGTT3', SEQ ID NO:32

5'TCCAACAACGACACATTCAGGAGTT3', SEQ ID NO:33

[0432] UC Band #31 (SEQ ID NO:4)

5'GGACACAGAGTAAGATACCCACTGA3', SEQ ID NO:34

5'CCTCGGTCTTTGGTCTTTGCATATC3', SEQ ID NO:35

[0433] UC Band #32 (SEQ ID NO:7)

5'ACAAGGAAAGTGTCCCTATCTCTGA3', SEQ ID NO:36

5'CTCGAGGTCTCCCACTGAAGTGCTC3', SEQ ID NO:37

[0434] UC Band #33 (SEQ ID NO:5)

5'CACTGCACATTAAGATGGAGCCCGA3', SEQ ID NO:38

5'CCTGTAGAAGTTCTGCTGCGTGTGG3', SEQ ID NO:39

[0435] UC Band #38 (SEQ ID NO:10)

5' TCGCTCCACATTCATCCTTTCT3', SEQ ID NO:49

5' TGATCCCTGGGTGATATAGAGCATA3', SEQ ID NO:50

[0436] UCBand#40 (SEQ ID NO:11)

5' GCCCCACATCTGAACAAGCTAATAA3', SEQ ID NO:51

5' TGCGCCCTTCATACAGGCAGAGTTG3', SEQ ID NO:52

[0437] UC Band #41 (SEQ ID NO:12) j

5' CACGATGCCATTCTGCCATTTCTGT3', SEQ ID NO:53

5' GGAAGAGATGGAATAGAAACTGTAA3', SEQ ID NO:54

[0438] UC Band #43 (SEQ ID NO:19)

5' CACTGGAACCAACAGGCCTGCCTCAAC3', SEQ ID NO:57

5' CCGAGCCAATTGGTACAGGTCTGTTCTCCC3', SEQ ID NO:58

[0439] UC Band #47 (SEQ ID NO:47)

5' CCTCAAGACTGGTCCACGGAGTGTATGA3', SEQ ID NO:59

5' GGGTAATGGCCAAAGTATGTTCTCAAAGCA3', SEQ ID NO:60

[0440] UC Band #201 (SEQ ID NO:13)

5' AAACAAACGTCTTTGGGTAAA3', SEQ ID NO:61

5' CTGGACAAAGAGGAATATGA3', SEQ ID NO:62

[0441] UC Band #204 (SEQ ID NO:20)

5' GCCCTTTATAAATACGATTAGTATGGAG3', SEQ ID NO:63

5' TGTAGTTAGTGCAGCAAAAGGAAGA3', SEQ ID NO:64

[0442] UC Band #205 (Humhek) (SEQ ID NO:14)

5' GATGTAATTAAAGCTGTAGATGAGGG3', SEQ ID NO:65

5' GAATACTAACAATCTGCTCAAACTTGGG3', SEQ ID NO:66

[0443] UC Band #207 (SEQ ID NO:15)

5' GCCAAATGGGTAGCATTGTTGCTCGG3', SEQ ID NO:67

5' CAGAGTGGGGCAAGATACCCTTGAG3', SEQ ID NO:68

[0444] UC Band #209 (SEQ ID NO:16)

5' AATGGAATTTCTTATGCCCTC3', SEQ ID NO:69

5' CAATGCCAAGCACCCACTGATTC3', SEQ ID NO:70

[0445] UC Band #210 (SEQ ID NO:17)

5' ACACAGACACACACATGCACACCA3', SEQ ID NO:71

5' CCTACCTGTGCAGAAATCAA3', SEQ ID NO:72

[0446] UC Band #211 (SEQ ID NO:21)

5' AGCAGCATAGCCTCTCTGAAACTC3', SEQ ID NO:73

5' CCTTCTCATGTAGCCTGCAACCTGCTC3', SEQ ID NO:74

[0447] UC Band #212 (SEQ ID NO:22)

5' CATTGGTGCAGCAGGTTTAGATGG3', SEQ ID NO:75

5' GAGATATCAATTTATAAGCACCAAG3', SEQ ID NO:76

[0448] UC Band #213 (SEQ ID NO:23) j

5' ATCTCAATCATTGAGCCTGAAGG3', SEQ ID NO:77

5' CAGCAGGTTGAGTGAGGGATTTGG3', SEQ ID NO:78

[0449] UC Band #214 (SEQ ID NO:45)

5' CACAGATGTAGCTTCCTCACTGG3', SEQ ID NO:6

5' CTTCATGGCAGGACTCGGTTTGGG3', SEQ ID NO:18

[0450] UC Band #215 (SEQ ID NO:46)

5' CCTGTGGCGTAAGGCATCCCA3', SEQ ID NO:24

5' GCAAGCACTCCTTTGTAAAATGTCC3', SEQ ID NO:25

[0451] Controls used to normalize relative quantitative RT-PCR

[0452] β-actin

5' CGAGCTGCCTGACGGCCAGGTCATC3', SEQ ID NO:40

5' GAAGCATTTGCGGTGGACGATGGAG3', SEQ ID NO:41

[0453]

TABLE 5

Oligonucleotide for detection of the truncated Her2/new mRNA. NEU T3'
5' CCCCTTTTATAGTAAGAGCCCCAGA3', SEQ ID NO:44

Example 5

Identification of a Marker of Prostate, Bladder and Breast Cancer

[0454] The prostate disease marker UC Band #28 (SEQ ID NO:3), which was previously identified by RNA fingerprinting in the examples above, was chosen for further analysis. Using the original UC 28 EST as a probe, the full-length cDNA for UC 28 gene has been cloned by a combination of cDNA library screening and RACE (Rapid Cloning of cDNA Ends) methods (Frohman, In: *PCR PROTOCOLS: A GUIDE TO METHODS AND APPLICATIONS*, Academic Press, N.Y., 1990 incorporated by reference). Two alternative cDNA sequences for this gene corresponding to mRNA splice variants were isolated and sequenced, and are included as SEQ ID NO:83 and SEQ ID NO:85. Each sequence has the same open reading flame, and encodes a protein with 135 amino acids (SEQ ID NO:84 and SEQ ID NO:86).

[0455] Northern analysis using the UC 28 EST as the probe, confirmed there are two alternative splicing mRNA variants, with sizes 2.1 and 2.5 kb. The message for this gene was highly expressed in colon, small intestine, and prostate. The message was also seen in testes, spleen, thymus, and a modest amount of expression was seen in peripheral blood leukocytes. The two mRNA variants share an identical 5' untranslated region, an open reading frame and part of the 3' untranslated region; and they differ only at the end of the 3' untranslated region. Both mRNAs have poly A tails and predicted polyadenylation sites. Both mRNAs, with the 2.5 kb mRNA being more abundant than the 2.1 kb mRNA, were confirmed to be expressed in the prostate tissues by PCRTM using one common 5' primer and two different 3' primers.

[0456] The 5' end sense primer used for amplification of both mRNAs was:

5' TAGAAGACCAAATGCCCCGAGT3', SEQ ID NO:42

[0457] The 3' end anti-sense primer for the 2.1 kb mRNA was:

5' TGTATTTCTGTGGGATCGGTGG3', SEQ ID NO:43

[0458] The 3' end anti-sense primer for the 2.5 kb mRNA was:

5' CCCACCTCCCAAAGTGCTGGGA3', SEQ ID NO:87

[0459] The medium for PCRTM amplification contained 2 μl of prostate tissue cDNA, 10 mM Tris-HCl (pH 9.3), 50 mM KCl, 3 mM $MgCl_2$, 0.5 mM dNTP's, 1.25 U of Taq DNA polymerase (GIBCO/BRL) and 200 nM of sense and anti-sense primers in a total reaction volume of 30 μl. Amplification was performed in a thermal cycler (MJ Research), for 36 cycles of 1 min at 94° C., 1 min at 55° C. and 1 min at 72° C.

[0460] The expression of UC 28 has been reconfirmed to be up regulated more than five fold in organ-confined and metastatic prostate cancers using relative quantitative RT-PCRTM in an expanded panel of tissue samples (**FIG. 18**). Pools of mRNA from normal individuals and mRNA from a normal individual were compared to mRNA samples from BPH and prostate tumor bearing individuals (n>10). The expression of UC 28 has also been investigated by RT-PCR in breast, bladder lung, and colon cancers. UC 28 is significantly up regulated in breast cancer, but not in lung or colon cancers (**FIG. 15**). UC 28 expression was increased between two- and eight-fold in four out of five bladder cancer samples examined, compared with four normal bladder tissue controls. (**FIG. 16**).

[0461] Since hormones regulate both prostate and mammary glands, the possibility that expression of this gene is modified by androgen was examined. Results from the studies in LnCaP cells indicated that indeed, gene expression is stimulated by dihydrotestosterone (DHT) in a time and dosage-dependent manner (**FIG. 17**). Considering the fact that both prostate and mammary glands are hormonally regulated, and the gene is up regulated in prostate and breast cancers, this gene may be involved in the hormone-regulated cell growth or proliferation pathways in these glands.

[0462] In situ hybridization (ISH) studies were performed to investigate the expression of the gene and to localize UC 28 mRNA in formalin-fixed paraffin-embedded radical prostatectomy specimens. For ISH, a 25-mer antisense oligonucleotide, of the sequence listed below, was biotinylated at the 3' end and used as the probe with alkaline phosphatase or horseradish peroxidase used as the detection enzyme.

5' CTTAACTCGGGCATTTGGTCTTC 3' SEQ ID NO:55

[0463] Minimal levels of UC 28 mRNA were detected in the adjacent benign prostatic epithelial cells of prostate cancer tissue. A significantly higher level of UC 28 mRNA was localized in prostate adenocarcinoma tissue. The results further confirm the up-regulation of the gene in prostate cancer and localize the expression of the gene to the prostatic epithelia.

[0464] A first generation polyclonal antibody has been produced in rabbits using a KLH conjugated synthetic peptide (21 amino acids). The peptide, of sequence listed below, was chosen for antigenicity by a computer software program (DNASTAR, Madison, Wis.).

[0465] RKKEKVKRSQKATEFIDYSIE SEQ ID NO:56

[0466] The synthetic peptide was conjugated to KLH by standard techniques and injected into two rabbits, with bleeding started at ten weeks. The antibody was peptide affinity purified and then tested in prostate cancer cell lines, and breast and prostate cancer tissue, confirming the localization of the UC 28 protein to epithelial cells, mainly on the cell membrane.

[0467] EcoRI or HindIII digested human genomic DNA was also probed in a Southern analysis to demonstrate that UC 28 is encoded by a single copy of the gene in the human genome (**FIG. 19**). The gene has been mapped to chromosome 6q23-24 by FISH chromosome mapping (**FIG. 20**).

[0468] Computer analyses using bioinfomatics from public databases (MotifFinder program in the GenomeNet database, Japan, motif@genome.adjp) indicate that the UC 28 peptide has a possible 26 amino acid transmembrane domain from amino acid 34 to amino acid 50, and also contains three PKC phosphorylation sites beginning at amino acids 62 (SQK), 89 (TMK), and 94 (SMK) and one myristylation site beginning at amino acid 118 (GLECCL). In vitro translation studies using rabbit reticulocyte lysate methods were performed to evaluate the size of the translated product from the open reading frame. A single 17 kDa protein product was obtained, which is the correct predicted size from the open reading frame.

[0469] Those experienced in the art will recognize that the genes and gene products (RNAs and proteins) for the above described markers of prostate disease, normal prostate, bladder cancer and breast cancer are included within the scope of the invention herein described. Those experienced in the art will also recognize that the diagnosis and prognosis of prostatic, bladder or breast cancer by detection of the nucleic acid products of these genes are included within the scope of the present invention.

[0470] All of the compositions and methods disclosed and claimed herein can be made and executed without undue experimentation in light of the present disclosure. While the compositions and methods of this invention have been described in terms of preferred embodiments, it will be apparent to those of skill in the art that variations may be applied to the composition, methods and in the steps or in the sequence of steps of the method described herein without departing from the concept, spirit and scope of the invention.

[0471] More specifically, it will be apparent that certain agents which are both chemically and physiologically related may be substituted for the agents described herein while the same or similar results would be achieved. All such similar substitutes and modifications apparent to those skilled in the art are deemed to be within the spirit, scope and concept of the invention as defined by the appended claims.

REFERENCES

[0472] The following literature citations as well as those cited above are incorporated in pertinent part by reference herein for the reasons cited in the above text.

[0473] Alcaraz et al., *Cancer Res.,* 55:3998-4002, 1994.

[0474] Allhoff et al., *World J. Urol.,* 7:12-16, 1989.

[0475] American Cancer Society-Facts and Figures-1998, http://www.cancer.org/ statistics198cff/98prosta.html.

[0476] An et al., *Proc. Amer. Assn. Canc. Res.,* 36:82, 1995.

[0477] *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor Press, Cold Spring Harbor, N.Y., 1988.

[0478] Babian et al., *J. Urol.,* 156:432-437, 1996.

[0479] Badalament et al., *J. Urol.,* 156:1375-1380, 1996.

[0480] Baichwal and Sugden, In: *Gene Transfer*, Kucherlapati R, ed., New York, Plenum Press, pp. 117-148, 1986.

[0481] Bangham et al., *J. Mol. Biol.* 13: 238-252, 1965.

[0482] Barinaga, *Science,* 271: 1233, 1996.

[0483] Bellus, *J. Macromol. Sci. Pure Appl. Chem.*, A31(1):1355-1376, 1994.

[0484] Benvenisty and Neshif, *Proc. Nat. Acad. Sci. USA,* 83:9551-9555, 1986.

[0485] Bittner et al., *Methods in Enzymol.,* 153:516-544, 1987.

[0486] Bookstein et al., *Science,* 247:712-715, 1990a.

[0487] Bookstein et al., *Proc. Nat'l Acad. Sci. USA,* 87:7762-7767, 1990b.

[0488] Bova et al., *Cancer Res.,* 53:3869-3873, 1993.

[0489] Brawn et al., *The Prostate,* 28: 295-299, 1996.

[0490] Campbell, In: *Monoclonal Antibody Technology, Laboratory Techniques in Biochemistry and Molecular Biology*, Burden and Von Knippenberg, Eds., Vol. 13:75-83, Elsevier, Amsterdam, 1984.

[0491] Capaldi et al., *Biochem. Biophys. Res. Comm.,* 76:425, 1977.

[0492] Carter et al., *Proc. Nat'l Acad. Sci. USA,* 87:8751-8755, 1990.

[0493] Carter and Coffey, In: *Prostate Cancer: The Second Tokyo Symposium*, J. P. Karr and H. Yamanak (eds.), pp. 19-27, New York: Elsevier, 1989.

[0494] Carter and Coffey, *Prostate,* 16:39-48, 1990.

[0495] Carter et al., *Proc. Nat'l Acad. Sci. USA,* 87:8751-8755, 1990.

[0496] Carter et al., *Proc. Nat'l Acad. Sci. USA* 93: 749-753, 1996.

[0497] Carter et al., *J. Urol.,* 157:2206-2209, 1997.

[0498] Chang et al., *Hepatology,* 14:124A, 1991.

[0499] Chen and Okayama, *Mol. Cell Biol.,* 7:2745-2752, 1987.

[0500] Chen et al., *Clin. Chem.,* 41:273-282, 1995.

[0501] Chen et al., *Proc. Am. Urol. Assn.,* 153:267A, 1995.

[0502] Chinault and Carbon, *Gene,* 5:111-126, 1979.

[0503] Chomczynski and Sacchi, *Anal. Biochem.,* 162:156-159,1987.

[0504] Christensson et al., *J. Urol.,* 150:100-105, 1993.

[0505] Colberre-Garapinet al., *J. Mol. Biol.,* 150:1, 1981.

[0506] Cooner et al., *J. Urol.,* 143:1146-1154, 1990.

[0507] Coupar et al., *Gene,* 68:1-10, 1988.

[0508] Cox et al., *Breast Canc. Res. Treat.* 32: 19-38, 1994.

[0509] Davey et al., EPO No.329 822.

[0510] Deamer and Uster, "Liposome Preparation: Methods and Mechanisms," In: *Liposomes* M. Ostro, ed., 1983.

[0511] Diamond et al., *J. Urol.,* 128:729-734, 1982.

[0512] Diamond et al., *J Urol.,* 128: 729-734, 1982.

[0513] Donahue et al., *J. Biol. Chem.,* 269:8604-8609, 1994.

[0514] Dong et al., *Science,* 268: 884-886, 1995.

[0515] Dubensky et al., *Proc. Nat. Acad. Sci. USA,* 81:7529-7533, 1984.

[0516] Dumont et al., *J. Immunol.,* 152:992-1003, 1994.

[0517] Elledge et al., *Cancer Res.* 54: 3752-3757, 1994

[0518] Fearon et al., *Science* 247: 47-56, 1990.

[0519] Friedmann, *Science,* 244:1275-1281, 1989.

[0520] Freifelder, *Physical Biochemistry Applications to Biochemistry and Molecular Biology,* 2nd ed., Wm. Freeman and Co., New York, N.Y., 1982.

[0521] Frohman, *PCR PROTOCOLS: A GUIDE TO METHODS AND APPLICATIONS*, Academic Press, N.Y., 1990.

[0522] Gefter et al., *Somatic Cell Genet.,* 3:231-236, 1977.

[0523] Ghosh-Choudhury et al., *EMBO J,* 6:1733-1739, 1987.

[0524] Gingeras et al., PCT Application WO 88/10315.

[0525] Goding, In: *Monoclonal Antibodies: Principles and Practice,* 2d ed., Orlando, Fla., Academic Press, pp. 60-61, 65-66, 71-74, 1986.

[0526] Gopal, *Mol. Cell Biol.,* 5:1188-1190, 1985.

[0527] Graham et al., *J. Gen. Virol.,* 36:59-72, 1977.

[0528] Graham and van der Eb, *Virology,* 52:456-467, 1973.

[0529] Graham and Prevec, In: *Methods in Molecular Biology: Gene Transfer and Expression Protocols* 7, E. J. Murray (ed.), Clifton, N.J., Humana Press, pp. 205-225. 1991.

[0530] G. Gregoriadis (ed.) *DRUG CARRIERS IN BIOLOGY AND MEDICINE*, pp. 287-341, 1979.

[0531] Grunhaus and Horwitz, *Seminar in Virology,* 3:237-252, 1992.

[0532] Harris et al., *J. Urol.,* 157:1740-1743, 1997.

[0533] Hermonat and Muzycska, *Proc. Nat. Acad. Sci. USA,* 81:6466-6470, 1984.

[0534] Hess et al., *J. Adv. Enzyme Reg.,* 7:149, 1968.

[0535] Hitzeman et al., *J. Biol. Chem.,* 255:2073, 1980.

[0536] Holland et al., *Biochemistry,* 17:4900, 1978.

[0537] Horoszewicz, Kawinski and Murphy, *Anticancer Res.,* 7:927-936, 1987.

[0538] Horwich, et al., *J. Virol.,* 64:642-650, 1990.

[0539] Huang et al., *Prostate,* 23: 201-212, 1993.

[0540] Innis et al., *PCR Protocols*, Academic Press, Inc., San Diego Calif., 1990.

[0541] Inouye et al., *Nucleic Acids Res.,* 13:3101-3109, 1985.

[0542] Isaacs et al., *Cancer Res.,* 51:4716-4720, 1991.

[0543] Isaacs et al., *Seminars in Oncology,* 21: 1-18, 1994.

[0544] Israeli et al., *Cancer Research,* 54:1807-1811, 1994.

[0545] Israeli et al., *J. Urol.,* 153:573-577, 1995.

[0546] Jaakkola et al., *Clin. Chem.,* 41:182-186, 1995.

[0547] Jacobson et al., *JAMA,* 274:1445-1449, 1995.

[0548] Johnson et al., In: *BIOTECHNOLOGY AND PHARMACY,* Pezzuto et al., Eds., Chapman and Hall, New York, 1993.

[0549] Jones, *Genetics,* 85:12, 1977.

[0550] Kaneda et al., *Science,* 243:375-378, 1989.

[0551] Kato et al., *J. Biol. Chem.,* 266:3361-3364, 1991.

[0552] Kingsman et al., *Gene,* 7: 141, 1979.

[0553] Klein et al., *Nature,* 327:70-73, 1987.

[0554] Kohler and Milstein, *Nature,* 256:495-497, 1975.

[0555] Kohler and Milstein, *Eur. J. Immunol.,* 6:511-519, 1976.

[0556] Kwoh et al., *Proc. Nat. Acad. Sci. USA,* 86:1173, 1989.

[0557] Liang and Pardee, *Science,* 257:967-971, 1992.

[0558] Liang and Pardee, U.S. Pat. No. 5,262,311, 1993.

[0559] Liang et al., *Cancer Res.,* 52:6966-6968, 1992.

[0560] Lifton, *Science,* 272: 676, 1996.

[0561] Lilja et al., *Clin. Chem.,* 37:1618-1625, 1991.

[0562] Lithrup et al., *Cancer,* 74:3146-3150, 1994.

[0563] Lowy et al., *Cell,* 22:817, 1980.

[0564] Macoska et al., *Cancer Res.,* 54:3824-3830, 1994.

[0565] Marley et al., *Urology,* 48(6A):16-22, 1996.

[0566] McCormack et al., *Urology,* 45:729-744, 1995.

[0567] Miki et al., *Science* 266: 66-71, 1994.

[0568] Miller et al., PCT Application WO 89/06700.

[0569] Mok et al., *Gynecol. Oncol.,* 52:247-252, 1994.

[0570] Morahan et al., *Science* 272: 1811, 1996.

[0571] Mulligan et al., *Proc. Nat'l Acad Sci. USA,* 78:2072, 1981.

[0572] Mulligan, *Science,* 260:926-932, 1993.

[0573] Murphy et al., *Cancer,* 78: 809-818, 1996.

[0574] Murphy et al., *Prostate,* 26:164-168, 1995.

[0575] Myers, EP 0273085

[0576] Nagase et al., *DNA Res.,* 3(5):321-329, 1996.

[0577] Nakamura et al., In: *Handbook of Experimental Immunology* (4th Ed.), Weir, E., Herzenberg, L. A., Blackwell, C., Herzenberg, L. (eds), Vol. 1, Chapter 27, Blackwell Scientific Publ., Oxford, 1987.

[0578] O'Dowd et al., *J. Urol.,* 158: 687-698, 1997.

[0579] O'Hare et al., *Proc. Nat'l Acad. Sci. USA,* 78:1527, 1981.

[0580] Oesterling, *Contemporary Urology,* 8:76-92, 1996.

[0581] Oesterling et al., *J. Urol.,* 154:1090-1095, 1995.

[0582] Ohara et al., *Proc. Nat'l Acad. Sci. USA,* 86:5673-5677, 1989.

[0583] Orozco et al., *Urology,* 51: 186-195, 1998.

[0584] Parker et al., *CA Cancer J. Clin.,* 65:5-27, 1996.

[0585] Partin and Oesterling, *Urology,* 48 (6A):1-3, 1996.

[0586] Partin and Oesterling, *J. Urol.,* 152:1358-1368, 1994.

[0587] Partin and Oesterling, (Eds.) *Urology,* 48(6A) supplement: 1-87, 1996.

[0588] Partin et al., *Cancer Res.,* 53:744-746, 1993.

[0589] Pettersson et al., *Clin. Chem.,* 41(10):1480-1488, 1995.

[0590] Piironen et al., *Clin. Chem.* 42: 1034-1041, 1996.

[0591] Qiao et al., *Biochem. Biophys. Res. Comm.,* 201:581-588, 1994.

[0592] Ralph and Veltri, *Advance/Laboratory,* 6:51-56, 1997.

[0593] Ralph et al., *Proc. Natl. Acad. Sci. USA,* 90(22):10710-10714, 1993.

[0594] Ribas de Pouplana and Fothergill-Gilmore, *Biochemistry,* 33:7047-7055, 1994.

[0595] Ridgeway, In: *Vectors: A survey of molecular cloning vectors and their uses*, Rodriguez R L, Denhardt D T, ed., Stoneham:Butterworth, pp. 467-492, 1988.

[0596] Rieber, M. and Rieber, M. S., *Cell Growth Diff.,* 5:1339-1346, 1994.

[0597] Rippe et al., *Mol. Cell Biol.,* 10:689-695, 1990.

[0598] Sager et al., *FASEB J.,* 7:964-970, 1993.

[0599] Sambrook et al., (ed.), *MOLECULAR CLONING*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

[0600] Santerre et al., *Gene,* 30: 147-156, 1984.

[0601] Scott et al., *Molec. Cell. Biol.,* 13:2247-2257, 1993.

[0602] Sidransky et al., *Science,* 252: 706-709, 1991.

[0603] Silver et al., *Clin. Cancer Res.,* 3: 81-85, 1997.

[0604] Slamon et al., *Science,* 224:256-262, 1984.

[0605] Slamon et al., *Science* 235; 177-182, 1987.

[0606] Slamon et al., *Science,* 244: 707-712, 1989.

[0607] Smith, U.S. Pat. No. 4,215,051.

[0608] Soh et al., *J. Urol.,* 157:2212-2218, 1997.

[0609] Stenman et al., *Cancer Research,* 51:222-226, 1991.

[0610] Stinchcomb et al., *Nature,* 282:39, 1979.

[0611] Stratford-Perricaudet and Perricaudet, In: *Human Gene Transfer*, O. Cohen-Haguenauer et al., eds., John Libbey Eurotext, France, pp. 51-61, 1991.

[0612] Stratford-Perricaudet et al., *Hum. Gene. Ther.,* 1:241-256, 1990.

[0613] Stratton and Wooster, *Curr. Opin. Genet. Devel.* 6: 93-97, 1996.

[0614] Sun and Cohen, *Gene,* 137:127-132, 1993.

[0615] Szoka et al., *Proc. Natl. Acad. Sci. USA,* 75: 4194-4198, 1978.

[0616] Szybalska et al., *Proc. Nat'l Acad. Sci. USA,* 48:2026, 1962.

[0617] Takahashi et al., *Cancer Res.,* 54:3574-3579, 1994.

[0618] Takahashi et al., *Mol. Cell. Biol.,* 8:1853-1856, 1988.

[0619] Taparowsky et al., *Nature,* 300: 762-764, 1982.

[0620] Tooze, *Molecular Biology of DNA Tumor Viruses,* 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1991.

[0621] Tschemper et al., *Gene,* 10:157, 1980.

[0622] VanArsdale et al., *Proc. Natl. Acad. Sci. USA,* 94:2460-2465, 1997.

[0623] Veltri et al., *J. Cell Biochem.* 19(suppl): 249-258, 1994.

[0624] Veltri et al., *Urology,* 48: 685-691, 1996.

[0625] Visakorpi et al., *Am. J. Pathol.,* 145:1-7, 1994.

[0626] Wagner et al., *Science,* 260:1510-1513, 1990.

[0627] Walker et al., *Proc. Nat'l Acad. Sci. USA,* 89:392-396, 1992.

[0628] Watson et al., *Cancer Res.,* 54:4598-4602, 1994.

[0629] Webb and Lin, *Invest Urol.,* 17:401-404, 1980.

[0630] Welsh et al., *Nucleic Acids Res.,* 20:4965-4970, 1992.

[0631] Welsh and McClelland, *Nucl. Acids. Res.,* 18:7213-7218, 1990.

[0632] Wigler et al., *Cell,* 11:223, 1977.

[0633] Wigler et al., *Proc. Nat'l Acad. Sci. USA,* 77:3567, 1980.

[0634] Wingo et al., *CA Cancer J. Clin.,* 47: 239-242, 1997.

[0635] Wong et al., *Gene,* 10: 87-94,1980.

[0636] Wong et al., *Int. J. Oncol.,* 3:13-17, 1993.

[0637] Wu and Wu, *J. Biol. Chem.,* 262:4429-4432, 1987.

[0638] Wu and Wu, *Biochemistry,* 27: 887-892, 1988.

[0639] Wu and Wu, *Adv. Drug Delivery Rev.,* 12:159-167, 1993.

[0640] Wu et al., *Genomics,* 4:560, 1989.

[0641] Zlotta, et al., *J. Urol.,* 157:1315-1321, 1997.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(iii) NUMBER OF SEQUENCES: 87

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 391 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single

-continued

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

GTCCAGTCGC TCAGAAATTT CCTTTGATGC TTTGAAGTTA TCTCTCTTGG ATCTGCTTCC 60

TCCTTATCGT CTCTACATCC CAAGAACAGA GAGTGAGTCT TCTTTATTTT CTTATCTCTG 120

TTTTTAGCAC AGTATTTGAT ATATAGTGTA GATACTATAA ATGCTTGCTA AACTTTGTCA 180

AATTCCACAT TTTTAAAATA AAAATGAGAA TGAGCTTGTA GTCAACATGG CGTTTGTAAG 240

TTTGGAGTCT ATATATGGTA GATATACATA TTTTTAAATC TAAGTGCAAC TTTTCTCTTG 300

ATTATCTTGA AATGCCTTAT CATCTCCACA TTTGCTGTAG GCAGTAGTTT AGTGGGTCCA 360

TTATATCTGC CACACTGATT GTCTTAAATA A 391

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 614 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

CAGTAGTGGC CCCAAATGCC AGGCTGCACT GATATTTATT GGATATAAGA CAAAGGGGCA 60

GGGTAAGGAA TGTGAACCAT CTCCAATAAT AGGTAAGGTC ACATGGGTCA TGTGTCCACT 120

GGACAGGGGG CCCTTCCCTG CCTGGCAGCA GAGGCAGAGA GAGAGAGAAG AGAGAGAGAC 180

AGCTTATGCC ATTATTTCTG CATATCAGAC ATTTAGTACT TTCACTAATT TGCTCCTGCT 240

ATCTAAAAGG CAGAGCCAGG TATACAGGAT GGAACATGAA AGCGGACTAG GAGCGTGACC 300

ACTGAAGCAC AGCATCACAG GGAGACAGGC CTCTGGATAC TGGCCGGGGG GCCCTGACTG 360

ATGTCAAGGC CCTCCACAAG AGTGGAGGAG TTAGTCTTCC TCTAAACTCC CCCGGGGGAA 420

AGGGAGGCTC CTTTTCCCAG TCTGCTAAGT AGTGGGTGTT TTTCCTTGAC ACTGATGCTA 480

CTGCTAGACC ATGGTCCACT TTGCAACAGG CATCTTCCCA GACACTGGTG TTACTGCTAG 540

ACCAAGCCCT CTGGTGGCCC TGTCCGGGCA TAAGAGAAGG CTCACACTCT TGTCTTCTGG 600

CCACTTCGCA CTAT 614

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 757 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

ACAACGACAC ATTCAGGAGT TAAATATTTA TCATCAAACA TTGGATTTTT CCTTAACGCT 60

AGAGATTGCT ACAAATCTTC TGAAGGGTCT CAATGGCTTC AGGCTAAGAA GAGATTTCTC 120

CCTGTTATAA GCAGCAAGAC AAATTAGCCA TTTCACTCTC AAACTTCACT AATGATCACA 180

TTCTTTCCAA AAGGAACTCT AGAAGACCAA ATGCCCCGAG TTAAGAACAT CAAAACTAAC 240

CATCTGAAGA AACTTCCCAA GTGTAAGACT CTGCCATTAA AACATTACCG AGAGGGGACT 300

CAAACAGTCT TTTCTTCCCT TTGTCGTGTT TCTTTGCTCC CAGACCCAAG GCACTTGGCG 360

GACAGTACTT GATACAATAA TTTAAAAAGC ACCACTCCCT TCCCACTTTG TAAATACCCA 420

-continued

GAACTCTAAT TGGACCACCC TGAAGCTTAG GACCTACCAG CCATACAAAT AGTAAACTCT 480

GTCCACGATT CACTCATCTG TGTATTTTCT ATAGATGTTT ACTAGGCGTT TGTTATATAA 540

AAATACCCCG GCCAGGCACG GTGGCTCACG CCTGTAATCC CAGCACTTTG GGAGGTGGGT 600

GGATCACCTG AGGTCGGGAG TTCGAGACCA GCCTGACCAG CATGGTGGAA CCCCCATCTC 660

TACTAAAAAC ACAAAAAATT AGCCGGGCGT GGTGGCACAT GCCTGTAATC CCAGCTACTC 720

AGGAGGCTGA GGCGGAGAAT TGCTTGAACC CGGAAGG 757

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 673 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

CAGGACACAG AGTAAGATAC CCACTGACTT CTTGTGGTCT ACTTCCTGGG TGTTGTTTCA 60

ATGGGCTTTG TTATAACAGG ACTAGTCTTC TGTAAATACA ACTTGGTAAA TAGGATGAAA 120

CATAACTTTG CGACAATTCA GTAGAAATAG GCATACAAAC CTGGGCCTGA TGACACTCAC 180

CTCCCCTTGG CTATAAACAT TACCCTACCT GTTAAGTCAG TAATCCTTTG GGAGAGCGCT 240

TACTGAGTAT CTATGATATG CAAAGACCAA AGACCGAGGG GGATCCCTGG TGTAGAGCAA 300

GCACACACCT GGTTATTAGC TACCTGCCAC CCTGCTGGGC ATGCAACATA CATTGTCTCA 360

AATTCTAACC ACCCTGCAAG GCAAGCTTCC TTGTTCTTTT AAAGAAGAAA AGTAGACCAC 420

CAAGATTGAT TTGCTCAAGA TTACACAGCC TGGAATCTTG TCATGGGCAT GTCTGACTCT 480

GATAGCAATA CCCTCAAAGA AACTGTCAGA GAAGACTCAA TAAGAAGAAA GTTGAGATAC 540

AGAAACCAAC AGGAGAAGGT AATTCAGAAA TTCAAACAGA GTGGGTGTGA TGGGAAGAAT 600

TCATTAATAA GAAGGTACCT CTGTAGAAAA ATCTTACCAG ACAGTCTGGA AGTGAAGGAA 660

ACAGCCAATA GTC 673

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 358 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

GTCACTGCAC ATTAAGATGG AGCCCGAAGA GCCACACTCC GAGGGGGCAT CGCAGGAGGA 60

TGGGGCTCAA GGTGCCTGGG GCTGGGCACC CCTAAGTCAC GGCTCTAAGG AGAAAGCTCT 120

CTTCCTGCCC GGCGGAGCCC TCCCCTCCCC CCGGATCCCC GTGCTTTCCC GAGAGGGGAG 180

GACCAGAGAC CGGCAGATGG CTGCAGCGCT CCTCACTGCC TGGTCCCAGA TGCCAGTGAC 240

TTTCGAGGAT GTGGCCTTGT ACCTCTCCCG GGAGGAGTGG GGACGGCTGG ACCACACGCA 300

GCAGAACTTC TACAGGGAAT GTCCTGCAGA AGAAAAATGG GCTGTCACTG GGCTTTCC 358

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid

-continued

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

CACAGATGTA GCTTCCTCAC TGG 23

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 610 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

CTGGAGTACA ATGTCAGTGT TTACACTGTC AAGGATGACA AGGAAAGTGT CCCTATCTCT 60

GATACCATCA TCCCAGCTGT TCCTCCTCCC ACTGACCTGC GATTCACCAA CATTGGTCCA 120

GACACCATGC GTGTCACCTG GGCTCCACCC CCATCCATTG ATTTAACCAA CTTCCTGGTG 180

CGTTACTCAC CTGTGAAAAA TGAGGAAGAT GTTGCAGAGT TGTCAATTTC TCCTTCAGAC 240

AATGCAGTGG TCTTAACAAA TCTCCTGCCT GGTACAGAAT ATGTAGTGAG TGTCTCCAGT 300

GTCTACGAAC AACATGAGAG CACACCTCTT AGAGGAAGAC AGAAAACAGG TCTTGATTCC 360

CCAACTGGCA TTGACTTTTC TGATATTACT GCCAACTCTT TTACTGTGCA CTGGATTGCT 420

CCTCGAGCCA CCATCACTGG CTACAGGATC CGCCATCATC CCGAGCACTT CAGTGGGAGA 480

CCTCGAGAAG ATCGGGTGCC CCACTCTCGG AATTCCATCA CCCTCACCAA CCTCACTCCA 540

GGCACAGAGT ATGTGGTCAG CATCGTTGCT CTTAATGGCA GAGAGGAAAG TCCCTTATTG 600

ATTGGCCAAC 610

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1649 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

CGGCAGCCAG CCTATTCTTT GGCCGGGTCG GTGCGAGTGG TCGGCTGGGC AGAGTGCACG 60

CTGCTTGGCG CCGCAGGTGA TCCCGCCGTC CACTCCCGGG AGCAGTGATG TTGGGCAACT 120

CTGCGCCGGG GCCTGCGACC CGCGAGGCGG GCTCGGCGCT GCTAGCATTG CAGCAGACGG 180

CGCTCCAAGA GGACCAGGAG AATATCAACC CGGAAAAGGC AGCGCCCGTC CAACAACCGC 240

GGACCCGGGC CGCGCTGGCG GTACTGAAGT CCGGGAACCC GCGGGGTCTA GCGCAGCAGC 300

AGAGGCCGAA GACGAGACGG GTTGCACCCC TTAAGGATCT TCCTGTAAAT GATGAGCATG 360

TCACCGTTCC TCCTTGGAAA GCAAACAGTA AACAGCCTGC GTTCACCATT CATGTGGATG 420

AAGCAGAAAA AGAAGCTCAG AAGAAGCCAG CTGAATCTCA AAAAATAGAG CGTGAAGATG 480

CCCTGGCTTT TAATTCAGCC ATTAGTTTAC CTGGACCCAG AAAACCATTG GTCCCTCTTG 540

ATTATCCAAT GGATGGTAGT TTTGAGTCAC CACATACTAT GGACATGTCA ATTGTATTAG 600

AAGATGAAAA GCCAGTGAGT GTTAATGAAG TACCAGACTA CCATGAGGAT ATTCACACAT 660

ACCTTAGGGA AATGGAGGTT AAATGTAAAC CTAAAGTGGG TTACATGAAG AAACAGCCAG 720

ACATCACTAA CAGTATGAGA GCTATCCTCG TGGACTGGTT AGTTGAAGTA GGAGAAGAAT 780

-continued

ATAAACTACA GAATGAGACC CTGCATTTGG CTGTGAACTA CATTGATAGG TTCCTGTCTT 840

CCATGTCAGT GCTGAGAGGA AAACTTCAGC TTGTGGGCAC TGCTGCTATG CTGTTAGCCT 900

CAAAGTTTGA AGAAATATAC CCCCCAGAAG TAGCAGAGTT TGTGTACATT ACAGATGATA 960

CCTACACCAA GAAACAAGTT CTGAGAATGG AGCATCTAGT TTTGAAAGTC CTTACTTTTG 1020

ACTTAGCTGC TCCAACAGTA AATCAGTTTC TTACCCAATA CTTTCTGCAT CAGCAGCCTG 1080

CAAACTGCAA AGTTGAAAGT TTAGCAATGT TTTTGGGAGA ATTAAGTTTG ATAGATGCTG 1140

ACCCATACCT CAAGTATTTG CCATCAGTTA TTGCTGGAGC TGCCTTTCAT TTAGCACTCT 1200

ACACAGTCAC GGGACAAAGC TGGCCTGAAT CATTAATACG AAAGACTGGA TATACCCTGG 1260

AAAGTCTTAA GCCTTGTCTC ATGGACCTTC ACCAGACCTA CCTCAAAGCA CCACAGCATG 1320

CACAACAGTC AATAAGAGAA AAGTACAAAA ATTCAAAGTA TCATGGTGTT TCTCTCCTCA 1380

ACCCACCAGA GACACTAAAT CTGTAACAAT GAAAGACTGC CTTTGTTTTC TAAGATGTAA 1440

ATCACTCAAA GTATATGGTG TACAGTTTTT AACTTAGGTT TTTAATTTTA CAATCATTTC 1500

TGAATACAGA AGTTGTGGCC AAGTACAAAT TATGGTATCT ATTACTTTTT AAATGGTTTT 1560

AATTTGTATA TCTTTTGTAT ATGTATCTGT CTTAGATATT TGGCTAATTT TAAGTGGTTT 1620

TGTTAAAGTA TTAATGATGC CAGCTGCCG 1649

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 175 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

ACCCACTCGT GAGTCCAACG GTCTTTTCTG CAGAAAGGAG GACTTTCCTT TCAGGGGTCT 60

TTCTGGGGCT CTTACTATAA AAGGGGACCA ACTCTCCCTT TGTCATATCT TGTTTCTGAT 120

GACAAAAAAT AACACATTGT TAAAATTGTA AAATTAAAAC ATGAAATATA AATTA 175

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 166 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

GTTTCGCTCC ACATTCATCC TTTCTTACTG GGCACTGATG TTGAGAGCAT CAGGCAGGGT 60

ATAATGTTAT GTTGCAGTAA CAAACACCCT CAATATCTCA GTGGCTTAAA ATGACAACGA 120

TCTTTTTTTT GTTTGTTTGT TTATGCTCTA TATCACCCAG GGATCA 166

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 107 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

-continued

TGCTCTGCCC CACATCTGAA CAAGCTAATA AGAAAGCCCG ATGTTCTTTC CTTTGGTGCC 60

ATTGGGAAAT TCAAACCATG CACAACTCTG CCTGTATGAA GGGCGCA 107

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 183 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

CAACCTTAGC CCCTCTCCTC TTCTTCACGA TGCCATTCTG CCATTTCTGT TTTGTGGTAG 60

ACAGGTTGGC CCAGGCACTC TAAGGCCCAG GCTGGCACAG GTTGGCCCAG GCACTTCAAG 120

CCTAAGTCCA TTTACAGTTT CTATTCCATC TCTTCCTAAA GAAGAGGAGA GGGGCTAAGG 180

TTG 183

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 92 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

AAACAAACGT CTTTGGGTAA AATTCTATTT CTTTTAATGT TTTAAAATAT TTGTAGTCAC 60

TAATTGTAAG TCATATTCCT CTTTGTCCAG CT 92

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 182 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

GATGTAATTA AAGCTGTAGA TGAGGGCTAT CGACTGCCAC CCCCCATGGA CTGCCCAGCT 60

GCCTTGTATC AGCTGATGCT GGACTGCTGG CAGAAAGACA GGAACAACAG ACCCAAGTTT 120

GAGCAGATTG TTAGTATTCT GGACAAGCTT ATCCGGAATC CCGGCAGCCT GAAGGATCAT 180

CA 182

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 174 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

GCCAAATGGG TAGCATTGTT GCTCGGCCTT CTAGTCTGCC AGTAGGAAAG TCCAACCATT 60

AGGTCGGGGA AGAAGGGTCT GGATTTGGTT GACAATGGTT GGATGGGGGA TAGAAGCAGA 120

GAGAGAGAGG GAGGGCAGCT CAAGGGTATC TTGCCCCACT CTGTTTATGC TGAT 174

(2) INFORMATION FOR SEQ ID NO: 16:

-continued

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 132 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

CACCTAACAA TATATCAATT TTTTAAAAAT GGAATTTCTT ATGCCCTCTT TATTTATGGA 60

CATGTATGTC CATAATGGGA GACGTTTTCT TTGGACTGAT GCTTGAATCA GTGGGTGCTT 120

GGCATTGCTG AT 132

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 135 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

CAGACACACA CATGCACACC ATTCTAGAAT GCTTCCTTAA AAGAAGGAGG GTTGCCCTAG 60

TCTCAAAATC TTAAAAGCCA TATGTGCATT GATTTCTGCA CAGGTAGGCA ATTTGTGATG 120

TTATTTTTCC TTATG 135

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

CTTCATGGCA GGACTCGGTT TGGG 24

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 471 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

GCCCCAAATG CCAGGCTGCA CTGATCTCAT GTCTGTGTCA CTGGAACCAA CAGGCCTGCC 60

TCAACCACTG TCCACCTGCA CATCTGAGAG GCTGGCAGGT CACCAGGGCT AGCCGTGCAC 120

GTCAGTTCCT GGGAAGAAAG TAGAATGTGA ATCATCTTCT CTCAAACGCC TATCAAAAGC 180

CCAGCTGAGA TCAATAATTT GGTGGGAGAA CAGACCTGTA CCAATTGGCT CGGTGTTTGG 240

TGGGGTATTG TAAATTTGGA TCCTAAATCA AAGGGTATCC CTAGAAGGAC CCACATGGAA 300

TGGCCTCCTC CTAAACATCC CTCCATGTTG GTACTTCCTG ACTCTTTTCC AGCAATCTCA 360

AAGCACAAGA AGCAGTGGTG GGAACCCAGG CCTGGCATCT TGTTGGAGCC CATGGTTGGG 420

GGGTAGGAGC AACTTTACAG GCCATCAATT ATGCCCCTAT ACGCACCTCC C 471

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:

-continued

(A) LENGTH: 209 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

GCCCTTTATA AATACGATTA GTATGGAGAA TTGATACATT AACAGTTAGC TTTATAAATT 60

GACAGATTTC TAAATTAACC TATGGTCCAC AAATCAAGTT CTATCACTAT TTCCTGCCAC 120

CAAAATCAGT GATGAAGCCT CTCCCACACT AAATGAAGAG TGGCGAGGGA CAGAATTCCA 180

CTTGTCTTCC TTTTGCTGCA CTAACTACA 209

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 407 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

CAAGCAGCAT AGCCTCTCTG AAACTCAATT TCCTCACATT TATAAATGAG CTTTTATATT 60

ATTTACAAAC CTACCTCATA GAGCAGGTTG CAGGCTACAT GAGAAGGTGC AAGTTCAATG 120

CCAAGCAGGG TCCTAGTATT TAATAAAAGC TCAATAAATA TTCATTTTCT TCTTTCCTTC 180

TCTTACTTGA AGTATAACAT TTGATAATGA ATTTTCTCAT TGCAACAATA ACACCCCTTC 240

CACTGAGGGA TTTGTATCCC TGCTTAAGAA GCTATTAGTA TTCTACAGCA GGACTCACCC 300

CACACAATCT TGGCAGGAAT ACATCCCTCT ACCTCTCTGG TCAATAACCT GCCTGGCCTG 360

TGACCCCAGG CTTCCTGGAG AAGCACCAAG TCCTCCCAGT TTCCCCC 407

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 267 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

CATTGGTGCA GCAGGTTTAG ATGGCTATGT GCTAGAGTAT TGCTTTGAAG GAAGTAAGTA 60

CAACCAGTAG ATAAAATGAA TACTGTCATC AATAGGTGAG ATATGTCCCT CCCCTTTCTG 120

TTGTCTCTCT TTCTTGAGAA CGCATCACCT TCCTACGAAA ATAAGATCAA GCCAAACGTC 180

ATCCTTCTGA GATGTATATA AACTAAGCCC TTTTTTAGTA CTTGGTGCTT ATAAATTGAT 240

ATCTCAAAAG TATCTTGGCT AGGCTGC 267

(2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 333 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

CATAGTCCAG GAGCAGAGTT AGCCAGAATT GCCTCCTGCT GCCCCAGCTT AGAGAGCTCC 60

CATCTCAATC ATTGAGCCTG AAGGCTTCAA GCCCAAAATG CAACAAGACC CCCAGCCTAC 120

-continued

ATTTCTCAGC TCCCCTGGAG CCAGTGATCC TGTAACGCTG CTGGAGGTCA GTCTGAGCTA 180

CCAAGACTGT CCCTAGACAA AGGTGGGAGT CCCCCACACT GCCAAGACCA AATCCCTCAC 240

TCAACCTGCT GAGGTGTTGG ATGGGGAAAC AAGAGGCAAA ACTGAGGCAC CTGATGCATT 300

CAGCCCTGCT TGTGCAGAAG TGCATTGACT GCC 333

(2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

CCTGTGGCGT AAGGCATCCC A 21

(2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

GCAAGCACTC CTTTGTAAAA TGTCC 25

(2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

TGCGTTCACC ATTCATGTGG ATGAAGCAG 29

(2) INFORMATION FOR SEQ ID NO: 27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

CTCCTACTTC AACTAACCAG TCCACGAG 28

(2) INFORMATION FOR SEQ ID NO: 28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

GATGCTTTGA AGTTATCTCT CTTGG 25

(2) INFORMATION FOR SEQ ID NO: 29:

-continued

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

ATCAGTGTGG CAGATATAAT GGACC 25

(2) INFORMATION FOR SEQ ID NO: 30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

GCCCCAAATG CCAGGCTGCA CTGAT 25

(2) INFORMATION FOR SEQ ID NO: 31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

GCCAGAAGAC AAGAGTGTGA GCCTT 25

(2) INFORMATION FOR SEQ ID NO: 32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

GCTTCAGGGT GGTCCAATTA GAGTT 25

(2) INFORMATION FOR SEQ ID NO: 33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

TCCAACAACG ACACATTCAG GAGTT 25

(2) INFORMATION FOR SEQ ID NO: 34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

GGACACAGAG TAAGATACCC ACTGA 25

-continued

(2) INFORMATION FOR SEQ ID NO: 35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

CCTCGGTCTT TGGTCTTTGC ATATC 25

(2) INFORMATION FOR SEQ ID NO: 36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

ACAAGGAAAG TGTCCCTATC TCTGA 25

(2) INFORMATION FOR SEQ ID NO: 37:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

CTCGAGGTCT CCCACTGAAG TGCTC 25

(2) INFORMATION FOR SEQ ID NO: 38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

CACTGCACAT TAAGATGGAG CCCGA 25

(2) INFORMATION FOR SEQ ID NO: 39:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

CCTGTAGAAG TTCTGCTGCG TGTGG 25

(2) INFORMATION FOR SEQ ID NO: 40:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

CGAGCTGCCT GACGGCCAGG TCATC 25

-continued

(2) INFORMATION FOR SEQ ID NO: 41:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41:

GAAGCATTTG CGGTGGACGA TGGAG 25

(2) INFORMATION FOR SEQ ID NO: 42:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

TAGAAGACCA AATGCCCCGA GT 22

(2) INFORMATION FOR SEQ ID NO: 43:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 43:

TGTATTTCTG TGGGATCGGT GG 22

(2) INFORMATION FOR SEQ ID NO: 44:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 44:

CCCCTTTTAT AGTAAGAGCC CCAGA 25

(2) INFORMATION FOR SEQ ID NO: 45:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 369 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 45:

CCATAAGAGA AATGATTGGT AGGTTTGCAT GAAATTTTAA AATTTCCTGT GGCGTAAGGC 60

ATCCCATAAC GAAGCCAAAA GGTGAGTGAT AGACTGGGAG AAATAACTGC CAGACGTTGC 120

CAGACAAAGA TTTCATATTT CTAATATGCT AGAGTACCTT TAATTTGATA AGAAAAAGAT 180

AAGCAATCCT GTAATAAAAT GGACATTTTA CAAAGGAGTG CTTGCAAATG GCCAGTGAAT 240

TTATGCAAAT ATGTTCAGGG AAATAGGAAT GAAAACGAGA TTCCACTTTT TCATCATCCA 300

TTTGATTGGC AAGAAATTTT TAAAAGAGTA ATACCTAGTG AATCACTCAT GTAGGAAAAT 360

-continued

GGGTTGGTG 369

(2) INFORMATION FOR SEQ ID NO: 46:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 301 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 212
(D) OTHER INFORMATION: /note= "N = A, C, G, or T"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 46:

GCCCTTGAAG AGTGTAACCA AGAAGCATCT CTCAATCAAT GAACCTGAGA CAGCCTGTTC 60

ACTTCTGACC ATCATTCTTG TCCTTTAGAT CTCAGTTTCA AATTCATTTC TTCTAGACAT 120

TCATCTCTTC CCATGTTTAA TCTGGAACCA TCTACCCTTC CACCAGACCA ATTATCCTGG 180

CAAATTAATG TAATAGACCA GTATTAATTA TNTGGTTGTA TGTCTTAACA ACATTCTAGG 240

TGCTGTGCCA AAAACAAATG AATAGCAACA CAAGGTCTTC TTGGTTACAC TCTTCAAGGG 300

C 301

(2) INFORMATION FOR SEQ ID NO: 47:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3061 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 15..1172

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 47:

```
CGGCTCTCCT CAAC ATG AGA GCT GCA CCC CTC CTC CTG GCC AGG GCA GCA       50
                Met Arg Ala Ala Pro Leu Leu Leu Ala Arg Ala Ala
                  1               5                  10

AGC CTT AGC CTT GGC TTC TTG TTT CTG CTT TTT TTC TGG CTA GAC CGA       98
Ser Leu Ser Leu Gly Phe Leu Phe Leu Leu Phe Phe Trp Leu Asp Arg
         15                  20                  25

AGT GTA CTA GCC AAG GAG TTG AAG TTT GTG ACT TTG GTG TTT CGG CAT      146
Ser Val Leu Ala Lys Glu Leu Lys Phe Val Thr Leu Val Phe Arg His
     30                  35                  40

GGA GAC CGA AGT CCC ATT GAC ACC TTT CCC ACT GAC CCC ATA AAG GAA      194
Gly Asp Arg Ser Pro Ile Asp Thr Phe Pro Thr Asp Pro Ile Lys Glu
 45                  50                  55                  60

TCC TCA TGG CCA CAA GGA TTT GGC CAA CTC ACC CAG CTG GGC ATG GAG      242
Ser Ser Trp Pro Gln Gly Phe Gly Gln Leu Thr Gln Leu Gly Met Glu
                 65                  70                  75

CAG CAT TAT GAA CTT GGA GAG TAT ATA AGA AAG AGA TAT AGA AAA TTC      290
Gln His Tyr Glu Leu Gly Glu Tyr Ile Arg Lys Arg Tyr Arg Lys Phe
             80                  85                  90

TTG AAT GAG TCC TAT AAA CAT GAA CAG GTT TAT ATT CGA AGC ACA GAC      338
Leu Asn Glu Ser Tyr Lys His Glu Gln Val Tyr Ile Arg Ser Thr Asp
         95                  100                 105

GTT GAC CGG ACT TTG ATG AGT GCT ATG ACA AAC CTG GCA GCC CTG TTT      386
Val Asp Arg Thr Leu Met Ser Ala Met Thr Asn Leu Ala Ala Leu Phe
    110                 115                 120
```

-continued

```
CCC CCA GAA GGT GTC AGC ATC TGG AAT CCT ATC CTA CTC TGG CAG CCC      434
Pro Pro Glu Gly Val Ser Ile Trp Asn Pro Ile Leu Leu Trp Gln Pro
125                 130                 135                 140

ATC CCG GTG CAC ACA GTT CCT CTT TCT GAA GAT CAG TTG CTA TAC CTG      482
Ile Pro Val His Thr Val Pro Leu Ser Glu Asp Gln Leu Leu Tyr Leu
                145                 150                 155

CCT TTC AGG AAC TGC CCT CGT TTT CAA GAA CTT GAG AGT GAG ACT TTG      530
Pro Phe Arg Asn Cys Pro Arg Phe Gln Glu Leu Glu Ser Glu Thr Leu
            160                 165                 170

AAA TCA GAG GAA TTC CAG AAG AGG CTG CAC CCT TAT AAG GAT TTT ATA      578
Lys Ser Glu Glu Phe Gln Lys Arg Leu His Pro Tyr Lys Asp Phe Ile
        175                 180                 185

GCT ACC TTG GGA AAA CTT TCA GGA TTA CAT GGC CAG GAC CTT TTT GGA      626
Ala Thr Leu Gly Lys Leu Ser Gly Leu His Gly Gln Asp Leu Phe Gly
    190                 195                 200

ATT TGG AGT AAA GTC TAC GAC CCT TTA TAT TGT GAG AGT GTT CAC AAT      674
Ile Trp Ser Lys Val Tyr Asp Pro Leu Tyr Cys Glu Ser Val His Asn
205                 210                 215                 220

TTC ACT TTA CCC TCC TGG GCC ACT GAG GAC ACC ATG ACT AAG TTG AGA      722
Phe Thr Leu Pro Ser Trp Ala Thr Glu Asp Thr Met Thr Lys Leu Arg
                225                 230                 235

GAA TTG TCA GAA TTG TCC CTC CTG TCC CTC TAT GGA ATT CAC AAG CAG      770
Glu Leu Ser Glu Leu Ser Leu Leu Ser Leu Tyr Gly Ile His Lys Gln
            240                 245                 250

AAA GAG AAA TCT AGG CTC CAA GGG GGT GTC CTG GTC AAT GAA ATC CTC      818
Lys Glu Lys Ser Arg Leu Gln Gly Gly Val Leu Val Asn Glu Ile Leu
        255                 260                 265

AAT CAC ATG AAG AGA GCA ACT CAG ATA CCA AGC TAC AAA AAA CTT ATC      866
Asn His Met Lys Arg Ala Thr Gln Ile Pro Ser Tyr Lys Lys Leu Ile
    270                 275                 280

ATG TAT TCT GCG CAT GAC ACT ACT GTG AGT GGC CTA CAG ATG GCG CTA      914
Met Tyr Ser Ala His Asp Thr Thr Val Ser Gly Leu Gln Met Ala Leu
285                 290                 295                 300

GAT GTT TAC AAC GGA CTC CTT CCT CCC TAT GCT TCT TGC CAC TTG ACG      962
Asp Val Tyr Asn Gly Leu Leu Pro Pro Tyr Ala Ser Cys His Leu Thr
                305                 310                 315

GAA TTG TAC TTT GAG AAG GGG GAG TAC TTT GTG GAG ATG TAC TAT CGG     1010
Glu Leu Tyr Phe Glu Lys Gly Glu Tyr Phe Val Glu Met Tyr Tyr Arg
            320                 325                 330

AAT GAG ACG CAG CAC GAG CCG TAT CCC CTC ATG CTA CCT GGC TGC AGC     1058
Asn Glu Thr Gln His Glu Pro Tyr Pro Leu Met Leu Pro Gly Cys Ser
        335                 340                 345

CCT AGC TGT CCT CTG GAG AGG TTT GCT GAG CTG GTT GGC CCT GTG ATC     1106
Pro Ser Cys Pro Leu Glu Arg Phe Ala Glu Leu Val Gly Pro Val Ile
    350                 355                 360

CCT CAA GAC TGG TCC ACG GAG TGT ATG ACC ACA AAC AGC CAT CAA GGT     1154
Pro Gln Asp Trp Ser Thr Glu Cys Met Thr Thr Asn Ser His Gln Gly
365                 370                 375                 380

ACT GAG GAC AGT ACA GAT TAGTGTGCAC AGAGATCTCT GTAGAAAGAG            1202
Thr Glu Asp Ser Thr Asp
                385

TAGCTGCCCT TTCTCAGGGC AGATGATGCT TTGAGAACAT ACTTTGGCCA TTACCCCCCA   1262

GCTTTGAGGA AAATGGGCTT TGGATGATTA TTTTATGTTT TAGGGACCCC CAACCTCAGG   1322

CAATTCCTAC CTCTTCACCT GACCCTGCCC CCACTTGCCA TAAAACTTAG CTAAGTTTTG   1382

TTTTGTTTTT CAGCGTTAAT GTAAAGGGGC AGCAGTGCCA AAATATAATC AGAGATAAAG   1442
```

-continued

```
CTTAGGTCAA AGTTCATAGA GTTCCCATGA ACTATATGAC TGGCCACACA GGATCTTTTG   1502

TATTTAAGGA TTCTGAGATT TTGCTTGAGC AGGATTAGAT AAGTCTGTTC TTTAAATTTC   1562

TGAAATGGAA CAGATTTCAA AAAAAATTCC CACAATCTAG GGTGGGAACA AGGAAGGAAA   1622

GATGTGAATA GGCTGATGGG GAAAAAACCA ATTTACCCAT CAGTTCCAGC CTTCTCTCAA   1682

GGAGAGGCAA AGAAAGGAGA TACAGTGGAG ACATCTGGAA AGTTTTCTCC ACTGGAAAAC   1742

TGCTACTATC TGTTTTTATA TTTCTGTTAA AATATATGAG GCTACAGAAC TAAAAATTAA   1802

AACCTCTTTG TGTCCCTTGG TCCTGGAACA TTTATGTTCC TTTTAAAGAA ACAAAAATCA   1862

AACTTTACAG AAAGATTTGA TGTATGTAAT ACATATAGCA GCTCTTGAAG TATATATATC   1922

ATAGCAAATA AGTCATCTGA TGAGAACAAG CTATTTGGGC ACAACACATC AGGAAAGAGA   1982

GCACCACGTG ATGGAGTTTC TCCAGAAGCT CCAGTGATAA GAGATGTTGA CTCTAAAGTT   2042

GATTTAAGGC CAGGCATGGT GGTTTACGCC TATAATCCCA GCATTTTGGG ACTCCGAGGT   2102

GGGCAGATCA CTTGAGCTCA GGAGCTCAAG ATCAGCCTGG GCAACATGGT GAAACCTTGT   2162

CTCTACATAA AATACAAAAA CTTAGATGGG CATGGTGCTG TGTGCCTATA GTCCACTACT   2222

TGTGGGGCTA AGGCAGGAGG ATCACTTGAG CCCCGGAGGT CGAGGCTACA GTGACCCAAG   2282

AGTGCACTAC TGTACTCCAG CCAGGGCAAG AGAGCGAGAC CCTGTCTCAA TAAATAAATA   2342

AATAAATAAA TAAATAAATA AATAAAAACA AAGTTGATTA AGAAAGGAAG TATAGGCCAG   2402

GCACAGTGGC TCACACCTGT AATCCTTGCA TTTTGGAAGG CTGAGGCAGG AGGATCACTT   2462

TAGGCCTGGT GTGTTCAAGA CCAGCCTGGT CAACATAGTG AGACACTGTC TCTACCAAAA   2522

AAAGGAAGGA AGGGACACAT ATCAAACTGA AACAAAATTA GAAATGTAAT TATGTTATGT   2582

TCTAAGTGCC TCCAAGTTCA AAACTTATTG GAATGTTGAG AGTGTGGTTA CGAAATACGT   2642

TAGGAGGACA AAAGGAATGT GTAAGTCTTT AATGCCGATA TCTTCAGAAA ACCTAAGCAA   2702

ACTTACAGGT CCTGCTGAAA CTGCCCACTC TGCAAGAAGA AATCATGATA TAGCTTTCCA   2762

TGTGGCAGAT CTACATGTCT AGAGAACACT GTGCTCTATT ACCATTATGG ATAAAGATGA   2822

GATGGTTTCT AGAGATGGTT TCTACTGGCT GCCAGAATCT AGAGCAAAGC CATCCCCCCT   2882

CCTGGTTGGT CACAGAATGA CTGACAAAGA CATCGATTGA TATGCTTCTT TGTGTTATTT   2942

CCCTCCCAAG TAAATGTTTG TCCTTGGGTC CATTTTCTAT GCTTGTAACT GTCTTCTAGC   3002

AGTGAGCCAA ATGTAAAATA GTGAATAAAG TCATTATTAG GAAGTTCAAA AAAAAAAAA    3061
```

(2) INFORMATION FOR SEQ ID NO: 48:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 386 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 48:

```
Met Arg Ala Ala Pro Leu Leu Leu Ala Arg Ala Ala Ser Leu Ser Leu
  1               5                  10                  15

Gly Phe Leu Phe Leu Leu Phe Phe Trp Leu Asp Arg Ser Val Leu Ala
             20                  25                  30

Lys Glu Leu Lys Phe Val Thr Leu Val Phe Arg His Gly Asp Arg Ser
         35                  40                  45
```

-continued

```
Pro Ile Asp Thr Phe Pro Thr Asp Pro Ile Lys Glu Ser Ser Trp Pro
     50                  55                  60

Gln Gly Phe Gly Gln Leu Thr Gln Leu Gly Met Glu Gln His Tyr Glu
 65                  70                  75                  80

Leu Gly Glu Tyr Ile Arg Lys Arg Tyr Arg Lys Phe Leu Asn Glu Ser
                 85                  90                  95

Tyr Lys His Glu Gln Val Tyr Ile Arg Ser Thr Asp Val Asp Arg Thr
            100                 105                 110

Leu Met Ser Ala Met Thr Asn Leu Ala Ala Leu Phe Pro Pro Glu Gly
        115                 120                 125

Val Ser Ile Trp Asn Pro Ile Leu Leu Trp Gln Pro Ile Pro Val His
    130                 135                 140

Thr Val Pro Leu Ser Glu Asp Gln Leu Leu Tyr Leu Pro Phe Arg Asn
145                 150                 155                 160

Cys Pro Arg Phe Gln Glu Leu Glu Ser Glu Thr Leu Lys Ser Glu Glu
                165                 170                 175

Phe Gln Lys Arg Leu His Pro Tyr Lys Asp Phe Ile Ala Thr Leu Gly
            180                 185                 190

Lys Leu Ser Gly Leu His Gly Gln Asp Leu Phe Gly Ile Trp Ser Lys
        195                 200                 205

Val Tyr Asp Pro Leu Tyr Cys Glu Ser Val His Asn Phe Thr Leu Pro
    210                 215                 220

Ser Trp Ala Thr Glu Asp Thr Met Thr Lys Leu Arg Glu Leu Ser Glu
225                 230                 235                 240

Leu Ser Leu Leu Ser Leu Tyr Gly Ile His Lys Gln Lys Glu Lys Ser
                245                 250                 255

Arg Leu Gln Gly Gly Val Leu Val Asn Glu Ile Leu Asn His Met Lys
            260                 265                 270

Arg Ala Thr Gln Ile Pro Ser Tyr Lys Lys Leu Ile Met Tyr Ser Ala
        275                 280                 285

His Asp Thr Thr Val Ser Gly Leu Gln Met Ala Leu Asp Val Tyr Asn
    290                 295                 300

Gly Leu Leu Pro Pro Tyr Ala Ser Cys His Leu Thr Glu Leu Tyr Phe
305                 310                 315                 320

Glu Lys Gly Glu Tyr Phe Val Glu Met Tyr Tyr Arg Asn Glu Thr Gln
                325                 330                 335

His Glu Pro Tyr Pro Leu Met Leu Pro Gly Cys Ser Pro Ser Cys Pro
            340                 345                 350

Leu Glu Arg Phe Ala Glu Leu Val Gly Pro Val Ile Pro Gln Asp Trp
        355                 360                 365

Ser Thr Glu Cys Met Thr Thr Asn Ser His Gln Gly Thr Glu Asp Ser
    370                 375                 380

Thr Asp
385
```

(2) INFORMATION FOR SEQ ID NO: 49:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 49:

-continued

TCGCTCCACA TTCATCCTTT CT 22

(2) INFORMATION FOR SEQ ID NO: 50:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 50:

TGATCCCTGG GTGATATAGA GCATA 25

(2) INFORMATION FOR SEQ ID NO: 51:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 51:

GCCCCACATC TGAACAAGCT AATAA 25

(2) INFORMATION FOR SEQ ID NO: 52:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 52:

TGCGCCCTTC ATACAGGCAG AGTTG 25

(2) INFORMATION FOR SEQ ID NO: 53:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 53:

CACGATGCCA TTCTGCCATT TCTGT 25

(2) INFORMATION FOR SEQ ID NO: 54:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 54:

GGAAGAGATG GAATAGAAAC TGTAA 25

(2) INFORMATION FOR SEQ ID NO: 55:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

-continued

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 55:

CTTAACTCGG GCATTTGGTC TTC 23

(2) INFORMATION FOR SEQ ID NO: 56:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: <Unknown>
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 56:

Arg Lys Lys Glu Lys Val Lys Arg Ser Gln Lys Ala Thr Glu Phe Il
1 5 10 15

Asp Tyr Ser Ile Glu
20

(2) INFORMATION FOR SEQ ID NO: 57:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 57:

CACTGGAACC AACAGGCCTG CCTCAAC 27

(2) INFORMATION FOR SEQ ID NO: 58:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 58:

CCGAGCCAAT TGGTACAGGT CTGTTCTCCC 30

(2) INFORMATION FOR SEQ ID NO: 59:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 59:

CCTCAAGACT GGTCCACGGA GTGTATGA 28

(2) INFORMATION FOR SEQ ID NO: 60:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 60:

GGGTAATGGC CAAAGTATGT TCTCAAAGCA 30

(2) INFORMATION FOR SEQ ID NO: 61:

-continued

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 61:

AAACAAACGT CTTTGGGTAA A 21

(2) INFORMATION FOR SEQ ID NO: 62:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 62:

CTGGACAAAG AGGAATATGA 20

(2) INFORMATION FOR SEQ ID NO: 63:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 63:

GCCCTTTATA AATACGATTA GTATGGAG 28

(2) INFORMATION FOR SEQ ID NO: 64:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 64:

TGTAGTTAGT GCAGCAAAAG GAAGA 25

(2) INFORMATION FOR SEQ ID NO: 65:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 65:

GATGTAATTA AAGCTGTAGA TGAGGG 26

(2) INFORMATION FOR SEQ ID NO: 66:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 66:

GAATACTAAC AATCTGCTCA AACTTGGG 28

-continued

(2) INFORMATION FOR SEQ ID NO: 67:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 67:

GCCAAATGGG TAGCATTGTT GCTCGG 26

(2) INFORMATION FOR SEQ ID NO: 68:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 68:

CAGAGTGGGG CAAGATACCC TTGAG 25

(2) INFORMATION FOR SEQ ID NO: 69:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 69:

AATGGAATTT CTTATGCCCT C 21

(2) INFORMATION FOR SEQ ID NO: 70:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 70:

CAATGCCAAG CACCCACTGA TTC 23

(2) INFORMATION FOR SEQ ID NO: 71:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 71:

ACACAGACAC ACACATGCAC ACCA 24

(2) INFORMATION FOR SEQ ID NO: 72:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 72:

-continued

CCTACCTGTG CAGAAATCAA 20

(2) INFORMATION FOR SEQ ID NO: 73:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 73:

AGCAGCATAG CCTCTCTGAA ACTC 24

(2) INFORMATION FOR SEQ ID NO: 74:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 74:

CCTTCTCATG TAGCCTGCAA CCTGCTC 27

(2) INFORMATION FOR SEQ ID NO: 75:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 75:

CATTGGTGCA GCAGGTTTAG ATGG 24

(2) INFORMATION FOR SEQ ID NO: 76:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 76:

GAGATATCAA TTTATAAGCA CCAAG 25

(2) INFORMATION FOR SEQ ID NO: 77:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 77:

ATCTCAATCA TTGAGCCTGA AGG 23

(2) INFORMATION FOR SEQ ID NO: 78:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

-continued

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 78:

CAGCAGGTTG AGTGAGGGAT TTGG 24

(2) INFORMATION FOR SEQ ID NO: 79:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 79:

CGCCTCAGGC TGGGGCAGCA TT 22

(2) INFORMATION FOR SEQ ID NO: 80:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 80:

ACAGTGGAAG AGTCTCATTC GAGAT 25

(2) INFORMATION FOR SEQ ID NO: 81:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 81:

CGAGCTGCCT GACGGCCAGG TCATC 25

(2) INFORMATION FOR SEQ ID NO: 82:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 82:

GAAGCATTTG CGGTGGACGA TGGAG 25

(2) INFORMATION FOR SEQ ID NO: 83:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2087 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 99..503

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 83:

GACCTTAAAT ATATCGAGGT GGCTAATTGA TGTATAATAA TTTACAAAAT TATTCTTCTA 60

TTGCTACAGA GCTACAATTC AATTTACAGT AGGCCACC ATG AGG GCC TTC TTA 113
Met Arg Ala Phe Leu
1 5

-continued

```
AGG AAC CAG AAA TAT GAG GAT ATG CAC AAT ATT ATT CAC ATT TTA CAG      161
Arg Asn Gln Lys Tyr Glu Asp Met His Asn Ile Ile His Ile Leu Gln
                 10                  15                  20

ATC AGA AAA TTG AGG CAC AGA TTA AGT AAC TTC CCA AGG CTA CCA GGC      209
Ile Arg Lys Leu Arg His Arg Leu Ser Asn Phe Pro Arg Leu Pro Gly
             25                  30                  35

ATT CTA GCT CCA GAA ACT GTG CTC TTA CCA TTC TGC TAC AAG GTA TTT      257
Ile Leu Ala Pro Glu Thr Val Leu Leu Pro Phe Cys Tyr Lys Val Phe
         40                  45                  50

CGA AAA AAA GAA AAA GTA AAA AGA AGT CAA AAG GCA ACA GAG TTC ATT      305
Arg Lys Lys Glu Lys Val Lys Arg Ser Gln Lys Ala Thr Glu Phe Ile
     55                  60                  65

GAT TAT TCC ATA GAA CAG TCA CAC CAT GCA ATT CTC ACA CCC TTG CAG      353
Asp Tyr Ser Ile Glu Gln Ser His His Ala Ile Leu Thr Pro Leu Gln
 70                  75                  80                  85

ACA CAC TTG ACC ATG AAA GGT TCC TCA ATG AAA TGT TCC TCA TTA TCT      401
Thr His Leu Thr Met Lys Gly Ser Ser Met Lys Cys Ser Ser Leu Ser
                 90                  95                 100

TCA GAA GCC ATA TTA TTC ACA TTG ACT TTG CAG TTA ACT CAG ACC CTA      449
Ser Glu Ala Ile Leu Phe Thr Leu Thr Leu Gln Leu Thr Gln Thr Leu
             105                 110                 115

GGT CTG GAA TGC TGT CTT CTC TAC TTA TCC AAA ACT ATA CAT CCA CAG      497
Gly Leu Glu Cys Cys Leu Leu Tyr Leu Ser Lys Thr Ile His Pro Gln
         120                 125                 130

ATC ATA TAAACTCTCA GCCCTGCTGC AAAGCCTTTC CAGAAAAATA AAAATGGTTG       553
Ile Ile
    135

AAAAGGCAAT TCTGCTACCA ATGACTGTTT AAGCCCAGCC AAGTAACTGA ACCATTCCAA    613

CTTCAATTTA CTTATGAAAA GAATTTGATG ATGTAGGAGG TTATTTCAAT TCTAAAATAC    673

AAACCCATGT TGATCTTTCT CAATCTTGAA CTCATAGATT ATTATCTATT ATCTCAATTT    733

AGTTTGTTAT TTATCCTAGT GGGCCATTAA AAACTACCAC ATGTGTTTCT GTCTCTCCAT    793

TAGTCAATAA CTAAACTAAC GAGCAATTAG TAAGCCATGT GCCAGATGCT CCGCTAGGCA    853

CCAGAGGGAT AAAAACAATA CTTATAGTAT ACCACTAATT TTCGCTTAGT AACTAGTGAA    913

ATGTTCAAGT CATGCCTGAG TCAAGAGTTG AGGAGACATT ACAATGTGTA ATGGAAACCA    973

AGGAAAGTGA AACTTTGGAT AAGTGGGGAC TAGTGTATTT ATATATTTAA TTGATTTCCA   1033

ACTCTATCAT TGGCCTCCAA ACACAGATTG TGTTTTTCTT TGGTTTTGTT TTCTTCACTA   1093

TGGGATCTTC TGTGCCCAGC ACAGTGCCTG ACACATAGAA AACAATCAAT ATTTGCTGAA   1153

TAAATGATTA AAAAATCAGA GAACTTTCCC ATTCTGTTTG GATCTATAGA ACATCCAGAG   1213

TAAGTGATGA GGGCCTCTGC ATTTATATGC GCTTAAATTA AGATTATGTG AGAAAAGTTT   1273

AAAGACACTT AGTAGAGTGA TTTTGAAATA TAGTAAACAC TTGGAAATGG TGGTGCTTTA   1333

AAAAGATATT AATAGATAAT ATGAAAATCT CCATCTCAAA AATAATGCAT AAACTATTTA   1393

AAGGAAAATC ACATCTCCAG GCTTTCAATG TTTGTTCATT ACTTTTTCAT ATATTTTTAC   1453

CATCTGCTGA AGGCAGTCAT ATCAAAGGGT AAAGAAAGAT GGGAGGAAAA CTCAGTAAGA   1513

ATTATATTAG TCTGTTTGCA AAGTAGAAAA AGATTCTCAT CACTCAACCT TATGAGCAGG   1573

AAGAGGGAAG GCTGTTTGAG AACCATTTAC TTAGCAGAAC CACATATTTT AGACACTTCC   1633

CTGCATTAAC TGCACAAACA ATATGTTTGC AAACTTGTTG ATCAACCTCC AACAACGACA   1693

CATTCAGGAG TTAAATATTT TTCATCAAAC ATTGGATTTT TCCTTAACGC TAGAGATTGC   1753
```

-continued

```
TACAAATCTT CTGAAGGGTC TCAATGGCTT CAGGCTAAGA AGAGATTTCT CCCTGTTATA   1813

AGCAGCAAGA CAAATTAGCC ATTTCACTCT CAAACTTCAC TAATGATCAC ATTCTTTCCA   1873

AAAGGAACTC TAGAAGACCA AATGCCCCGA GTTAAGAACA TCAAAACTAA CCATCTGAAG   1933

AAACTTCCCA AGTGTAAGAC TCTGCCTGCA CGACAACACA TAAAAAAAGA GAGAAGAATC   1993

AAATAGACAC AATAAAAAAT GATAAAGGGG ATATCACCAC CGATCCCACA GAAATACAAA   2053

CTACCATCAG AGAATACTAC AAACACCTCT ACGC                               2087
```

(2) INFORMATION FOR SEQ ID NO: 84:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 135 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 84:

```
Met Arg Ala Phe Leu Arg Asn Gln Lys Tyr Glu Asp Met His Asn Ile
  1               5                  10                  15

Ile His Ile Leu Gln Ile Arg Lys Leu Arg His Arg Leu Ser Asn Phe
             20                  25                  30

Pro Arg Leu Pro Gly Ile Leu Ala Pro Glu Thr Val Leu Leu Pro Phe
         35                  40                  45

Cys Tyr Lys Val Phe Arg Lys Lys Glu Lys Val Lys Arg Ser Gln Lys
     50                  55                  60

Ala Thr Glu Phe Ile Asp Tyr Ser Ile Glu Gln Ser His His Ala Ile
 65                  70                  75                  80

Leu Thr Pro Leu Gln Thr His Leu Thr Met Lys Gly Ser Ser Met Lys
                85                  90                  95

Cys Ser Ser Leu Ser Ser Glu Ala Ile Leu Phe Thr Leu Thr Leu Gln
            100                 105                 110

Leu Thr Gln Thr Leu Gly Leu Glu Cys Cys Leu Leu Tyr Leu Ser Lys
        115                 120                 125

Thr Ile His Pro Gln Ile Ile
    130                 135
```

(2) INFORMATION FOR SEQ ID NO: 85:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2505 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 99..503

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 85:

```
GACCTTAAAT ATATCGAGGT GGCTAATTGA TGTATAATAA TTTACAAAAT TATTCTTCTA     60

TTGCTACAGA GCTACAATTC AATTTACAGT AGGCCACC ATG AGG GCC TTC TTA        113
                                          Met Arg Ala Phe Leu
                                            1               5

AGG AAC CAG AAA TAT GAG GAT ATG CAC AAT ATT ATT CAC ATT TTA CAG      161
Arg Asn Gln Lys Tyr Glu Asp Met His Asn Ile Ile His Ile Leu Gln
                10                  15                  20
```

-continued

```
ATC AGA AAA TTG AGG CAC AGA TTA AGT AAC TTC CCA AGG CTA CCA GGC      209
Ile Arg Lys Leu Arg His Arg Leu Ser Asn Phe Pro Arg Leu Pro Gly
             25                  30                  35

ATT CTA GCT CCA GAA ACT GTG CTC TTA CCA TTC TGC TAC AAG GTA TTT      257
Ile Leu Ala Pro Glu Thr Val Leu Leu Pro Phe Cys Tyr Lys Val Phe
         40                  45                  50

CGA AAA AAA GAA AAA GTA AAA AGA AGT CAA AAG GCA ACA GAG TTC ATT      305
Arg Lys Lys Glu Lys Val Lys Arg Ser Gln Lys Ala Thr Glu Phe Ile
     55                  60                  65

GAT TAT TCC ATA GAA CAG TCA CAC CAT GCA ATT CTC ACA CCC TTG CAG      353
Asp Tyr Ser Ile Glu Gln Ser His His Ala Ile Leu Thr Pro Leu Gln
 70                  75                  80                  85

ACA CAC TTG ACC ATG AAA GGT TCC TCA ATG AAA TGT TCC TCA TTA TCT      401
Thr His Leu Thr Met Lys Gly Ser Ser Met Lys Cys Ser Ser Leu Ser
                 90                  95                 100

TCA GAA GCC ATA TTA TTC ACA TTG ACT TTG CAG TTA ACT CAG ACC CTA      449
Ser Glu Ala Ile Leu Phe Thr Leu Thr Leu Gln Leu Thr Gln Thr Leu
            105                 110                 115

GGT CTG GAA TGC TGT CTT CTC TAC TTA TCC AAA ACT ATA CAT CCA CAG      497
Gly Leu Glu Cys Cys Leu Leu Tyr Leu Ser Lys Thr Ile His Pro Gln
        120                 125                 130

ATC ATA TAAACTCTCA GCCCTGCTGC AAAGCCTTTC CAGAAAAATA AAAATGGTTG       553
Ile Ile
    135

AAAAGGCAAT TCTGCTACCA ATGACTGTTT AAGCCCAGCC AAGTAACTGA ACCATTCCAA    613

CTTCAATTTA CTTATGAAAA GAATTTGATG ATGTAGGAGG TTATTTCAAT TCTAAAATAC    673

AAACCCATGT TGATCTTTCT CAATCTTGAA CTCATAGATT ATTATCTATT ATCTCAATTT    733

AGTTTGTTAT TTATCCTAGT GGGCCATTAA AAACTACCAC ATGTGTTTCT GTCTCTCCAT    793

TAGTCAATAA CTAAACTAAC GAGCAATTAG TAAGCCATGT GCCAGATGCT CCGCTAGGCA    853

CCAGAGGGAT AAAAACAATA CTTATAGTAT ACCACTAATT TTCGCTTAGT AACTAGTGAA    913

ATGTTCAAGT CATGCCTGAG TCAAGAGTTG AGGAGACATT ACAATGTGTA ATGGAAACCA    973

AGGAAAGTGA AACTTTGGAT AAGTGGGGAC TAGTGTATTT ATATATTTAA TTGATTTCTG   1033

ACTCTATCAT TGGCCTCCAA ACACAGATTG TGTTTTTCTT TGGTTTTGTT TTCTTCACTA   1093

TGGGATCTTC TGTGCCCAGC ACAGTGCCTG ACACATAGAA AACAATCAAT ATTTGCTGAA   1153

TAAATGATTA AAAAATCAGA GAACTTTCCC ATTCTGTTTG GATCTATAGA ACATCCAGAG   1213

TAAGTGATGA GGGCCTCTGC ATTTATATGC GCTTAAATTA AGATTATGTG AGAAAAGTTT   1273

AAAGACACTT AGTAGAGTGA TTTTGAAATA TAGTAAACAC TTGGAAATGG TGGTGCTTTA   1333

AAAAGATATT AATAGATAAT ATGAAAATCT CCATCTCAAA AATAATGCAT AAACTATTTA   1393

AAGGAAAATC ACATCTCCAG GCTTTCAATG TTTGTTCATT ACTTTTTCAT ATATTTTTAC   1453

CATCTGCTGA AGGCAGTCAT ATCAAAGGGT AAAGAAAGAT GGGAGGAAAA CTCAGTAAGA   1513

ATTATATTAG TCTGTTTGCA AAGTAGAAAA AGATTCTCAT CACTCAACCT TATGAGCAGG   1573

AAGAGGGAAG GCTGTTTGAG AACCATTTAC TTAGCAGAAC CACATATTTT AGACACTTCC   1633

CTGCATTAAC TGCACAAACA ATATGTTTGC AAACTTGTTG ATCAACCTCC AACAACGACA   1693

CATTCAGGAG TTAAATATTT TTCATCAAAC ATTGGATTTT TCCTTAACGC TAGAGATTGC   1753

TACAAATCTT CTGAAGGGTC TCAATGGCTT CAGGCTAAGA AGAGATTTCT CCCTGTTATA   1813

AGCAGCAAGA CAAATTAGCC ATTTCACTCT CAAACTTCAC TAATGATCAC ATTCTTTCCA   1873
```

-continued

AAAGGAACTC TAGAAGACCA AATGCCCCGA GTTAAGAACA TCAAAACTAA CCATCTGAAG 1933

AAACTTCCCA AGTGTAAGAC TCTGCCATTA AAACATTACC GAGAGGGGAC TCAAACAGTC 1993

TTTCTTCCTT TGTCGTGTTT CTTGCTCCCA GACCAAGGCA CTGACGACAG TACTGATACA 2053

TAATTTAAAA GCACACTCCC TTCCACTTTG GTAATACCAG AACTCTAATT GGACCACCCT 2113

GAAGCTTAGG ACTACCAGCC ATACAAATAG TAAACTCTGT CCACGATTCA CTCATCTGTG 2173

TATTTTCTAT AGATGTTTAC TAGGCGTTTG TTATATAAAA ATACCCCGGC CAGGCACGGT 2233

GGCTCACGCC TGTAATCCCA GCACTTTGGG AGGTGGGTGG ATCACCTGAG GTCGGGAGTT 2293

CGAGACCAGC CTGACCAGCA TGGTGGAACC CCCATCTCTA CTAAAAACAC AAAAAATTAG 2353

CCGGGCGTGG TGGCACATGC CTGTAATCCC AGCTACTCAG GAGGCTGAGG CGGAGAATTG 2413

CTTGAACCCG GAAGGTGGAG GTTGTTGCGG TGAGCTGAGA TTGCACTATT GCACTCCAGC 2473

CTGGGCAACA GGAGTAAAAC TCCCCCCCAC CC 2505

(2) INFORMATION FOR SEQ ID NO: 86:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 135 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 86:

Met Arg Ala Phe Leu Arg Asn Gln Lys Tyr Glu Asp Met His Asn Ile
1 5 10 15

Ile His Ile Leu Gln Ile Arg Lys Leu Arg His Arg Leu Ser Asn Phe
20 25 30

Pro Arg Leu Pro Gly Ile Leu Ala Pro Glu Thr Val Leu Leu Pro Phe
35 40 45

Cys Tyr Lys Val Phe Arg Lys Lys Glu Lys Val Lys Arg Ser Gln Lys
50 55 60

Ala Thr Glu Phe Ile Asp Tyr Ser Ile Glu Gln Ser His His Ala Ile
65 70 75 80

Leu Thr Pro Leu Gln Thr His Leu Thr Met Lys Gly Ser Ser Met Lys
85 90 95

Cys Ser Ser Leu Ser Ser Glu Ala Ile Leu Phe Thr Leu Thr Leu Gln
100 105 110

Leu Thr Gln Thr Leu Gly Leu Glu Cys Cys Leu Leu Tyr Leu Ser Lys
115 120 125

Thr Ile His Pro Gln Ile Ile
130 135

(2) INFORMATION FOR SEQ ID NO: 87:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 87:

CCCACCTCCC AAAGTGCTGG GA 22

1. An isolated nucleic acid segment comprising a full length sequence or the fill length complement of a sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:83 and SEQ ID NO:85.

2. An isolated nucleic acid molecule, of a size between about 14 and 100 bases in length, identical in sequence to a contiguous portion of at least 14 bases of a nucleic acid or its complement selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:83 and SEQ ID NO:85.

3. The isolated nucleic acid molecule of claim 2, of a size of between about 17 and 100 bases in length.

4. The isolated nucleic acid molecule of claim 2, of a size of between about 20 and 100 bases in length.

5. The isolated nucleic acid molecule of claim 2, of a size of between about 25 and 100 bases in length.

6. The isolated nucleic acid molecule of claim 2, of a size of between about 30 and 100 bases in length.

7. The isolated nucleic acid according to claim 1, wherein the sequence is SEQ ID NO:1.

8. The isolated nucleic acid according to claim 1, wherein the sequence is SEQ ID NO:2.

9. The isolated nucleic acid according to claim 1, wherein the sequence is SEQ ID NO:3.

10. The isolated nucleic acid according to claim 1 wherein the sequence is SEQ ID NO:4.

11. The isolated nucleic acid according to claim 1, wherein the sequence is SEQ ID NO:5.

12. The isolated nucleic acid according to claim 1, wherein the sequence is SEQ ID NO:10.

13. The isolated nucleic acid according to claim 1 wherein the sequence is SEQ ID NO:11.

14. The isolated nucleic acid according to claim 1, wherein the sequence is SEQ ID NO: 12.

15. The isolated nucleic acid according to claim 1, wherein the sequence is SEQ ID NO:13.

16. The isolated nucleic acid according to claim 1, wherein the sequence is SEQ ID NO:15.

17. The isolated nucleic acid according to claim 1, wherein the sequence is SEQ ID NO:16.

18. The isolated nucleic acid according to claim 1, wherein the sequence is SEQ ID NO:17.

19. The isolated nucleic acid according to claim 1, wherein the sequence is SEQ ID NO: 19.

20. The isolated nucleic acid according to claim 1, wherein the sequence is SEQ ID NO:20.

21. The isolated nucleic acid according to claim 1, wherein the sequence is SEQ ID NO:21.

22. The isolated nucleic acid according to claim 1, wherein the sequence is SEQ ID NO:22.

23. The isolated nucleic acid according to claim 1, wherein the sequence is SEQ ID NO:23.

24. The isolated nucleic acid according to claim 1 wherein the sequence is SEQ ID NO:45.

25. The isolated nucleic acid according to claim 1, wherein the sequence is SEQ ID NO:46.

26. The isolated nucleic acid according to claim 1 wherein the sequence is SEQ ID NO:83.

27. The isolated nucleic acid according to claim 1, wherein the sequence is SEQ ID NO:85.

28. An isolated polypeptide with an amino acid sequence encoded by a polynucleotide having a sequence selected from a group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:45 and SEQ ID NO:46.

29. An isolated peptide, of a size between 10 and 50 amino acids in length, with an amino acid sequence encoded within a polynucleotide having a sequence selected from a group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:45 and SEQ ID NO:46.

30. A method for identifying markers for human prostate cancer, comprising the following steps:

a) providing human prostate RNAs;

b) amplifying said RNAs to provide nucleic acid amplification products;

c) separating said nucleic acid amplification products; and

d) identifying those RNAs that are differentially expressed between human prostate cancers versus normal or benign human prostate.

31. The method according to claim 30, further comprising converting said RNAs into cDNAs using reverse transcriptase prior to amplification.

32. The method according to claim 31, further comprising amplifying the cDNAs by polymerase chain reaction (PCR) using arbitrarily chosen oligonucleotide primers under initially reduced stringency conditions.

33. The method according to claim 31, further comprising:

a) using one oligo dT anchoring primer and an arbitrarily chosen oligonucleotide primer for the reverse transcription step; and

b) using an oligo dT anchoring primer and an arbitrarily chosen oligonucleotide primer for the amplification step.

34. A method for detecting prostate cancer cells in a biological sample comprising the step of detecting a prostate cancer marker in said sample, wherein said prostate cancer marker is a nucleic acid having a sequence selected from a group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:83 and SEQ ID NO: 85.

35. The method of claim 34, further comprising the steps of

a) providing nucleic acids from said sample;

b) amplifying said nucleic acids to form nucleic acid amplification products;

c) contacting said nucleic acid amplification products with an oligonucleotide probe that will hybridize under stringent conditions with said prostate cancer marker;

d) detecting the nucleic acid amplification products which hybridize with said probe; and

e) measuring the amount of said nucleic acid amplification products that hybridize with said probe.

36. The method of claim 35, in which said oligonucleotide probe is selected to bind specifically an isolated nucleic acid having a sequence selected from a group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:83 and SEQ ID NO:85.

37. The method of claim 35, in which the sequence of said oligonucleotide probe is selected to bind specifically to a nucleic acid product of a known gene, said nucleic acid product selected from a group consisting of cyclin A (SEQ ID NO:8), fibronectin (SEQ ID NO:7), and a truncated form of Her2/neu (SEQ ID NO:9).

38. The method of claim 35, in which the sequence of said oligonucleotide probe is selected to bind specifically to a truncated nucleic acid product of the Her2/neu gene.

39. The method of claim 35, in which the sequence of said oligonucleotide probe is selected to bind specifically to a nucleic acid product of the cyclin A gene.

40. The method of claim 35, in which the sequence of said oligonucleotide probe is selected to bind specifically to a nucleic acid product of the fibronectin gene.

41. The method of claim 34, further comprising the steps of

a) providing nucleic acids from said sample;

b) providing primers that will selectively amplify said prostate cancer marker;

c) amplifying said nucleic acids with said primers to form nucleic acid amplification products;

d) detecting said nucleic acid amplification products; and

e) quantifying said nucleic acid amplification products.

42. The method of claim 41, wherein said primers are selected to amplify a nucleic acid having a sequence selected from a group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:83 and SEQ ID NO:85.

43. The method of claim 41, wherein said primers are selected to amplify a nucleic acid product of a known gene, said nucleic acid product selected from a group consisting of cyclin A, fibronectin, and a truncated form of Her2/neu.

44. The method of claim 41, wherein said primers are selected to amplify a truncated nucleic acid product of the Her2/neu gene.

45. The method of claim 41, wherein said primers are selected to amplify a nucleic acid product of the cyclin A gene.

46. The method of claim 41, wherein said primers are selected to amplify a nucleic acid product of the fibronectin gene.

47. The method of claim 41, further comprising determining the prognosis of prostate cancer patients by quantifying the nucleic acid amplification product binding to a probe specific for said prostate cancer marker.

48. The method of claim 41, further comprising determining the diagnosis of human prostate cancer by quantifying the nucleic acid amplification product binding to a probe specific for said prostate cancer marker.

49. The method of claim 41, further comprising determining the prognosis of prostate cancer patients by quantifying the nucleic acid amplification product.

50. The method of claim 41, further comprising determining the diagnosis of human prostate cancer by quantifying the nucleic acid amplification product.

51. A method of treating individuals with prostate cancer, comprising the steps of:

a) obtaining a sample of tissue from an individual with prostate cancer;

b) screening said sample for the expression of a polypeptide encoded by a polynucleotide having a sequence selected from a group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:83 and SEQ ID NO:85;

c) providing an antibody that reacts immunologically against said polypeptide; and

d) administering an effective amount of said antibody to an individual with prostate cancer.

52. A method of treating individuals with prostate cancer, comprising the steps of:

a) obtaining a sample of tissue from an individual with prostate cancer;

b) screening said sample for the expression of a polynucleotide having a sequence selected from a group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5,, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:83 and SEQ ID NO:85;

c) providing an antisense DNA molecule that encodes an RNA molecule that binds to said polynucleotide;

d) providing said antisense DNA molecule in the form of a human vector containing appropriate regulatory elements for the production of said RNA molecule; and

e) administering an effective amount of said vector to an individual with prostate cancer.

53. A kit for use in detecting prostate cancer cells in a biological sample, comprising:

(a) a primer pair for amplifying a nucleic acid having a sequence selected from a group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:83 and SEQ ID NO:85; and

(b) containers for each of said primers.

54. A kit for use in detecting prostate cancer cells in a biological sample, comprising:

(a) a primer pair for amplifying a nucleic acid product of a human gene, said nucleic acid product selected from a group consisting of cyclin A, fibronectin, and a truncated form of Her2/neu; and

(b) containers for each of said primers.

55. A kit for use in detecting prostate cancer cells in a biological sample, comprising:

(a) an oligonucleotide probe which binds under high stringency conditions to an isolated nucleic acid having a sequence selected from a group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:83 and SEQ ID NO:85; and

(b) a container for said probe.

56. A kit for use in detecting prostate cancer cells in a biological sample, comprising:

(a) an oligonucleotide probe which binds under high stringency conditions to a nucleic acid product of a human gene, said nucleic acid product selected from a group consisting of cyclin A, fibronectin gene, and a truncated form of Her2/neu; and

(b) a container for said probe.

57. A kit for use in detecting prostate cancer cells in a biological sample, comprising:

(a) an antibody which binds immunologically to a protein having an amino acid sequence encoded by a nucleic acid sequence selected from a group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:83 and SEQ ID NO:85; and

(b) a container for said antibody.

58. A method for detecting prostate cancer cells in biological samples, comprising the following steps:

(a) providing an antibody that binds immunologically to a peptide, said peptide encoded by an isolated nucleic acid selected from a group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:83 and SEQ ID NO:85, cyclin A, fibronectin, and a truncated form of Her2/neu;

(b) contacting a human tissue sample with said antibody;

(c) separating antibody bound to said tissue sample from unbound antibody; and

(d) detecting the bound antibody.

59. A kit for use in detecting prostate cancer cells in a biological sample, comprising:

(a) an antibody which binds immunologically to a polypeptide having an amino acid sequence encoded by a nucleic acid selected from a group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:83 and SEQ ID NO:85, cyclin A, fibronectin, and a truncated form of Her2/neu; and

(b) a container for said antibody.

60. A method for treating individuals with prostate cancer, comprising the following steps:

(a) selecting a prostate cancer marker from a group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:83, SEQ ID NO:85, cyclin A, fibronectin, and a truncated form of Her2/neu;

(b) providing an inhibitor designed to bind specifically to the protein product of said prostate cancer marker; and

(c) administering an effective dosage of said inhibitor to a prostate cancer patient.

61. An isolated nucleic acid segment useful as a marker of bladder cancer or breast cancer and having a sequence or the full length complement of a sequence selected from the group consisting of SEQ ID NO:3, SEQ ID NO:83 and SEQ ID NO:85.

62. An isolated nucleic acid molecule, of a size between about 14 and 100 bases in length, identical in sequence to a contiguous portion of at least 14 bases of a nucleic acid or its complement selected from the group consisting of SEQ ID NO:3, SEQ ID NO:83 and SEQ ID NO:85.

63. An isolated polypeptide with an amino acid sequence encoded by a nucleic acid having a sequence selected from a group consisting of SEQ ID NO:3, SEQ ID NO:83 and SEQ ID NO:85.

64. An isolated peptide, of a size between 10 and 50 amino acids in length, with an amino acid sequence encoded within a polynucleotide having a sequence selected from a group consisting of SEQ ID NO:3, SEQ ID NO:83 and SEQ ID NO:85.

65. A method for detecting bladder cancer or breast cancer cells in a biological sample comprising the step of detecting a bladder cancer or breast cancer marker in said sample, wherein said bladder cancer or breast cancer marker is a nucleic acid having a sequence selected from a group consisting of SEQ ID NO:3, SEQ ID NO:83 and SEQ ID NO:85.

66. The method of claim 65, further comprising the steps of

a) providing nucleic acids from said sample;

b) amplifying said nucleic acids to form nucleic acid amplification products;

c) contacting said nucleic acid amplification products with an oligonucleotide probe that will hybridize under stringent conditions with said bladder cancer or breast cancer marker;

d) detecting the nucleic acid amplification products which hybridize with said probe; and

e) quantifying the nucleic acid amplification products that hybridize with said probe.

67. The method of claim 65, further comprising the steps of

a) providing nucleic acids from said sample;

b) providing primers that will selectively amplify said bladder cancer or breast cancer marker;

c) amplifying said nucleic acids with said primers to form nucleic acid amplification products;

d) detecting said nucleic acid amplification products; and

e) quantifying said nucleic acid amplification products.

68. The method of claim 65, further comprising determining the prognosis of bladder cancer or breast cancer patients by quantifying the nucleic acid amplification product binding to a probe specific for said bladder cancer or breast cancer marker.

69. The method of claim 65, further comprising determining the diagnosis of human bladder cancer or breast cancer by quantifying the nucleic acid amplification product binding to a probe specific for said bladder cancer or breast cancer marker.

70. The method of claim 65, further comprising determining the prognosis of bladder cancer or breast cancer patients by quantifying the nucleic acid amplification product.

71. The method of claim 65, further comprising determining the diagnosis of human bladder cancer or breast cancer by quantifying the nucleic acid amplification product.

72. A method of treating individuals with bladder cancer or breast cancer, comprising the steps of:

a) obtaining a sample of tissue from an individual with bladder cancer or breast cancer;

b) screening said sample for the expression of a polypeptide encoded by a polynucleotide having a sequence selected from a group consisting of SEQ ID NO:3, SEQ ID NO:83 and SEQ ID NO:85;

c) providing an antibody that reacts immunologically against said polypeptide; and

d) administering an effective amount of said antibody to an individual with bladder cancer or breast cancer.

73. A method of treating a subject with bladder cancer or breast cancer, comprising the steps of:

a) obtaining a sample of tissue from an individual with bladder cancer or breast cancer;

b) screening said sample for the expression of a polynucleotide having a sequence selected from a group consisting SEQ ID NO:3, SEQ ID NO:83 and SEQ ID NO:85;

c) providing an antisense DNA molecule that encodes an RNA molecule that binds to said polynucleotide;

d) providing said antisense DNA molecule in the form of a human vector containing appropriate regulatory elements for the production of said RNA molecule; and

e) administering an effective amount of said vector to an individual with bladder cancer or breast cancer.

74. A kit for use in detecting bladder cancer cells or breast cancer cells in a biological sample, comprising:

a) a primer pair for amplifying a nucleic acid having a sequence selected from a group consisting of SEQ ID NO:3, SEQ ID NO:83 and SEQ ID NO:85; and

b) containers for each of said primers.

75. A kit for use in detecting bladder cancer cells or breast cancer cells in a biological sample, comprising:

a) an oligonucleotide probe which binds under high stringency conditions to an isolated nucleic acid having a sequence selected from a group consisting of SEQ ID NO:3, SEQ ID NO:83 and SEQ ID NO:85; and

b) a container for said probe.

76. A kit for use in detecting bladder cancer cells or breast cancer cells in a biological sample, comprising:

a) an antibody which binds immunologically to a protein having an amino acid sequence encoded by a polynucleotide having a sequence selected from a group consisting of SEQ ID NO:3, SEQ ID NO:83 and SEQ ID NO:85; and

b) a container for said antibody.

77. A method for detecting bladder cancer cells or breast cancer cells in biological samples, comprising the following steps:

a) providing an antibody that binds immunologically to a polypeptide encoded by an isolated nucleic acid selected from a group consisting of SEQ ID NO:3, SEQ ID NO:83 and SEQ ID NO:85;

b) contacting a human tissue sample with said antibody;

c) separating antibody bound to said tissue sample from unbound antibody; and

d) detecting the bound antibody.

* * * * *