(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
25 May 2001 (25.05.2001)

PCT

(10) International Publication Number
WO 01/36674 A2

(51) International Patent Classification[7]: C12Q 1/68, G01N 33/53, A61P 43/00

(21) International Application Number: PCT/GB00/04267

(22) International Filing Date:
8 November 2000 (08.11.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
9926805.4 13 November 1999 (13.11.1999) GB

(71) Applicant *(for all designated States except MG, US)*: **ASTRAZENECA AB** [SE/SE]; S-151 85 Sodertalje (SE).

(71) Applicant *(for MG only)*: **ASTRAZENECA UK LIMITED** [GB/GB]; 15 Stanhope Gate, London W1Y 6LN (GB).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **BULL, John, Henry** [GB/GB]; 25 Pownall Street, Macclesfield, Cheshire SK10 1DF (GB). **ELLISON, Gillian** [GB/GB]; Alderley Park, Macclesfield, Cheshire SK10 4TG (GB). **PASKINS, Lynn, Dora** [IE/GB]; 45 St. Anne's Avenue, Grappenhall, Warrington WA4 2PL (GB).

(74) Agent: **GAINEY, Laurence, David, Scott**; Astrazeneca, Global Intellectual Property, P.O. Box 272, Mereside, Alderley Park, Macclesfield, Cheshire SK10 4GR (GB).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *Without international search report and to be republished upon receipt of that report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: DIAGNOSTIC METHODS

(57) Abstract: The present invention derives from the identification of 26 gene transcripts (Markers) that exhibit aberrant expression levels in prostate disorder tissues. The invention therefore relates to diagnostic techniques for the detection of human prostate disorders, such as cancer, by detecting one or more of these Markers, intermediates, precursors or products (mRNA, cDNA, genomic DNA, or protein). The invention is also directed to methods for identifying modulators of prostate disorders, which modulators, such as chemical compounds, antisense molecules and antibodies interact with and modulate any one of the Markers identified.

WO 01/36674 A2

DIAGNOSTIC METHODS

The inventors have identified a number of expression Markers that are expressed at a higher level in prostate disorder tissue than normal tissue. The present invention therefore relates to diagnostic techniques for the detection of human prostate disorders, such as cancer, by detecting one or more of these Markers, intermediates, precursors or products (mRNA, cDNA, genomic DNA, or protein). The invention is also directed to methods for identifying modulators of prostate disorders, which modulators, such as chemical compounds, antisense molecules and antibodies interact with and modulate any one of the Markers identified.

Each of the Markers is known in the art but none has been shown to be up-regulated in, or associated with, prostate disorders.

Prostate cancer is the most common form of cancer in men, excluding skin cancer, and is the second leading cause of cancer death. Prostate cancers are adenocarcinomas arising from the epithelial component of the prostate gland. The earliest identified pre-malignant lesion is prostatic intraepithelial neoplasia (PIN). PIN is recognised histologically by abnormal cytology and is classed as low or high grade by nuclear pleiomorphism. Population postmortem studies show that PIN is present 10-20 years before prostate cancer. Evidence from morphology and genetic lesions strongly suggests that PIN is the precursor of prostate cancer. Individuals found to have high grade PIN are frequently found to have prostate cancer on subsequent biopsies, and lack of prostate cancer detection at the initial biopsy may be due to sampling limitations.

It has also been suggested that atypical adenomatous hyperplasia (AAH) might represent the precursor lesion of prostate cancer in the transitional and central zones of the prostate.

There are relatively few signs and symptoms of disease in the early stages of prostate cancer. Symptomatic disease results from metastasis, local invasion or when the size and location of the tumour affects the urinary tract. The most common symptom of prostate cancer arises from urinary tract obstruction due to the prostate gland being enlarged. When prostate cancer becomes advanced and distance metastases have occurred one of the most common symptoms experienced is bone pain. Acute onset of impotence also raises the possibility that prostate cancer is present. Mortality is principally a result of distant metastatic disease.

In early disease, where the cancer is confined to the prostate gland and has not spread beyond the gland, the tumour can potentially be removed by surgery. Surgery is a common treatment of prostate cancer. Transurethral resection is a procedure in which the cancer is cut from the prostate via the urethra. A radical prostatectomy is where the whole prostate and some of the tissue around the gland is removed. Radiation therapy is also used to treat early prostate cancer.

Once disease has become metastatic, its course is rapid and relatively predictable, leading to death within 2-5 years if untreated.

Prostate cancer commonly occurs in men above the age of 50 years. The incidence of prostate cancer rises sharply between the ages of 60 and 80 years, with more than 80% of all cases of the diseases being diagnosed in men over 65 and less than 1% in men under 50 years of age. The incidence and prevalence of prostate cancer have increased considerably throughout the world over the past two decades. The average age of detection is 72 in the UK and 66 in the US. Active screening programmes may lead to a lowering of the average age at first diagnosis.

Routine medical examinations for obstructive urinary symptoms may lead to an initial diagnosis of prostate cancer. When prostate cancer is asymptomatic, there are two tests that can be performed to detect the presence of cancer: digital rectal examination (DRE) and prostate specific antigen (PSA). PSA is an enzyme which is secreted almost exclusively by epithelial cells in the prostate; measurement of the quantity of PSA in the blood provides a more reliable indicator of the presence of prostate cancer than the DRE test. However, elevations in serum levels of PSA can also be found in patients with benign prostatic hyperplasia (BPH) and prostatitis. In some cases, where a patient has elevated levels of PSA in his blood, a third diagnostic technique is used called 'transrectal ultrasonograpy' (TRUS), which can further substantiate the presence of prostate cancer. In conjunction with taking a biopsy of the tissue, TRUS has been shown to increase the detection rate of prostate cancer two-fold, when compared to DRE alone. The combination of digital rectal examination, serum PSA level and transrectal ultrasound is currently the best available diagnostic tool (Gorgoulis VG. et al. Anticancer Res 19:2327-2348 (1999)).

It is common for the prostate gland to become enlarged with ageing, a condition called benign prostatic hyperplasia (BPH), or benign prostatic hypertrophy. Severe BPH can cause

serious problems over time. Urine retention and strain on the bladder can lead to urinary tract infections, bladder or kidney damage, bladder stones, and incontinence.

Although some of the signs of BPH and prostate cancer are the same, having BPH does not seem to increase the chances of getting prostate cancer. Nevertheless, a man who has BPH may have undetected prostate cancer at the same time or may develop prostate cancer in the future.

Prostate cancer is a very heterogeneous disease and many aspects of its management are controversial and anecdotal. Understanding of the disease is hampered by its slow development, inaccessibility and impact of age related death. The incidence of the disease is increasing, irrespective of PSA screening, and its significance will increase with decreasing age of incidence and increased average life span.

Major clinical diagnostic problems for prostate cancer include:

1. Detection is often too late for effective treatment.
2. Selection of the most appropriate management strategy is not clear cut.
3. Distinguishing patients with curable disease where a cure would prolong life.
4. Over-treatment by radical prostatectomy due to inadequate staging/prognostic information.
5. Ineffective monitoring of patients on disease management programmes.
6. Methods of differential diagnosis are generally invasive.
7. Laborious screening of TURP chips for cancer tissue, because 90% of enlarged prostates are caused by BPH.

There is therefore a need in the art for new diagnostic methods to identify prostate related disorders, particularly cancer, and especially methods capable of discriminating between benign (BPH) tissue, PIN tissue and tumour tissue. The present invention sets out to address this need.

The present inventors have identified a number of gene transcript Markers (mRNAs) that are up-regulated in diseased prostatic tissue. The gene, mRNA and protein sequences corresponding to these transcripts are therefore Markers of prostate disorder, and can be used to design specific probes, or to generate antibodies, capable of detecting the levels of the Marker, nucleic acid or encoded protein, present in a test sample such as a core biopsy sample, TURP sample, prostate segment, prostatic margin biopsy, lymph node biopsy or body fluid or cell sample. In addition these Markers are potential targets for therapeutic intervention in prostate disorders and disease, for instance in the development of antisense

nucleic acid targeted to the Marker mRNA; or more widely in the identification or development of chemical or hormonal therapeutic agents. The person skilled in the art is also capable of devising screening assays to identify compounds (chemical or biological) that modulate (activate or inhibit) one or other of the identified Marker, which compounds may prove useful as therapeutic agents in treating a prostate disorder. Monitoring the Markers could also be useful in identifying inhibitors, antagonists or biochemical signalling effects in high throughput screening.

According to a first aspect of the invention there is provided a method for diagnosing or prognosing or monitoring a prostate disorder comprising testing a biological sample for aberrant levels of one or more of the Markers selected from the group consisting of Marker 1 to Marker 26. Preferred Markers are numbers: 8, 13, 14, 15, 17, 18, 19, 20, 22, 24, 25 and 26.

The invention lies in the identification of these differentially expressed Markers in prostate disorders. Accordingly, the invention is directed to any diagnostic method capable of assessing the differential expression levels, relative to expression in control tissues, of one or more of the identified Markers, either alone or as a panel. In particular, such methods include assessment of mRNA transcript levels and/or protein levels. The presence of aberrant expression levels of one of the Markers indicating the presence of a prostate disorder.

In a preferred embodiment the diagnostic method involves testing for more than one of the Markers identified herein. As separate independent embodiments the diagnostic method may involve testing for 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 or all of the Markers identified herein, optionally as part of a panel test of other gene Markers.

It will be apparent to the person skilled in the art that there are a large number of analytical procedures which may be used to detect the amount of any of the Marker products present in a test sample.

The test sample comprising nucleic acid or protein is conveniently any prostatic material or biological sample, including TURP chip, biopsy, excised prostate or part thereof, a sample of bone marrow aspirate, bone marrow biopsy, lymph node aspirate, lymph node biopsy, spleen tissue, fine needle aspirate, skin biopsy or organ tissue biopsy, particularly prostate gland tissue, invaded surgical margin, invaded lymph node, invaded lung, invaded

bone or other bodily organ or material or cell sample, or other body fluid or tissue taken from an individual.

It will be appreciated that the test sample may equally be a nucleic acid sequence corresponding to the sequence in the test sample, that is to say that all or a part of the region in the sample nucleic acid may firstly be amplified using any convenient technique e.g. polymerase chain reaction (PCR), before analysis. The nucleic acid may be genomic DNA or fractionated or whole cell RNA. In one embodiment the RNA is whole cell RNA and is used directly as the template for labelling a first strand cDNA using random primers or poly A primers. The nucleic acid or protein in the test sample may be extracted from the sample according to standard methodologies (Sambrook et al. "Molecular Cloning- A Laboratory manual", second edition. Cold Spring Harbor, NY (1989)).

Prostate disorder includes PIN, BPH and prostate cancer.

Aberrant expression refers to expression levels that are outside the normal range. The normal range can be determined by testing many normal tissues or may be determined from a side by side comparison of the test sample with the normal or control sample. For the purposes of this application, aberrant expression refers to a 1.5 -fold difference or more in level of Marker nucleic acid in a disease sample compared to control normal or, when distinguishing cancer from BPH, in a cancer sample compared to control BPH. Control thus refers to a data set from typical normal prostate sample(s) (control normal), or a data set from typical BPH sample(s) (control BPH) as appropriate.

Nucleic acid as used herein refers to both RNA and DNA.

The inventors have found that Markers 1 - 26 exhibit a multi-fold increase (often in the range of 1.5 to 6-fold) in mRNA expression levels in prostate cancer tissue (see Fig. 1). Marker 19 was found to be expressed per se in cancer tissue but not in normal tissue making this Marker highly specific for prostate cancer and particularly useful in the instant invention. This highly specific Marker could form a sensitive test target for spread of prostate cancer cells to prostatic margins and lymph nodes, and could be sampled in surgery during radical prostatectomy. The inventors have found that some Markers (Markers 1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 12, 17, 18, 20, 23 (see Fig. 2)), are expressed at higher levels in cancer than BPH. Accordingly, these Markers may be used to distinguish between cancer and BPH. Four Markers (Markers 6, 13, 14, and 25) were not detected in BPH samples but were detected in various cancer samples. The latter four Markers may therefore represent important

biomarkers for distinguishing cancer from BPH. Highly specific yet low abundance Markers could be invaluable if secreted in protein form, as they could permit non-invasive testing of body fluids.

According to a further aspect of the invention there is provided a method for distinguishing prostate cancer from BPH comprising testing a biological sample for aberrant levels of one or more of the Markers selected from the group consisting of Marker 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 17, 18, 20, 23, 25 and 26. Such a method might involve measuring the expression levels of one or more of these Markers in a biological sample taken from an individual, comparing the Marker expression level detected with control values or historical values and depending on the value detected determining whether or not the individual has prostate cancer or BPH.

Each of the Marker genes useful in the present invention are already in the public domain. However, the inventors are not aware of any prior art disclosing an association of any of the Marker genes with prostate cancer. Thus, although each of the Markers/genes identified herein are themselves already known, their association with prostate disorders leading to the present invention is unknown.

Marker 1 is the human smooth muscle protein, 22kDa. The gene coding for this protein has been cloned and sequenced (Thweatt et al., Biochem. Biophys. Res. Commun. 187:1-7 (1992)). These authors identify 22kD smooth muscle protein as a fibrobast or smooth muscle protein. There is no teaching in this paper of an association with prostate cancer. The sequence of 22kDa smooth muscle protein is disclosed in Thweatt et al. and is present in the EMBL database under accession number HS22SM. For the purpose of this application, the gene and amino acid sequences and sequence positions referred to herein, refers to that in the EMBL database under accession number HS22SM, unless stated otherwise or apparent from the context.

Marker 2 is the Rho GDP dissociation inhibitor protein. The gene coding for this protein has been cloned and sequenced (Leffers et al., Exp. Cell Res. 209:165-174 (1993)). These authors identify Rho GDP dissociation inhibitor protein as a keratinocyte protein. There is no teaching in this paper of an association with prostate cancer. The sequence of Rho GDP dissociation inhibitor protein is disclosed in Leffers et al., and is present in the EMBL database under accession number HSRHO1. For the purpose of this application, the gene and amino acid sequences and sequence positions referred to herein, refers to that in the EMBL

database under accession numbers: HSRHO1, unless stated otherwise or apparent from the context.

Marker 3 is the human ferritin H protein. The gene coding for this protein has been cloned and sequenced (Hentze et al., Proc. Natl. Acad. Sci. U.S.A. 83:7226-7230 (1986)). These authors identify ferritin H protein as an intracellular iron-binding protein. There is no teaching in this paper of an association with prostate cancer. The sequence of ferritin H protein is disclosed in Hentze et al., and is present in the EMBL database under accession number HSFERHC2. For the purpose of this application, the gene and amino acid sequences and sequence positions referred to herein, refers to that in the EMBL database under accession number HSFERHC2, unless stated otherwise or apparent from the context.

Marker 4 represents the human cDNA DKFZp564I1922. This cDNA has been cloned and sequenced (Wambutt, R. et al., direct EMBL entry HSM800288)). These authors identifiy DKFZp564I1922 as a brain transcript. There is no teaching in this paper of an association with prostate cancer. The sequence of DKFZp564I1922 is disclosed in the EMBL database under accession number HSM800288. For the purpose of this application, the gene and amino acid sequences and sequence positions referred to herein, refers to that in the EMBL database under accession number HSM800288, unless stated otherwise or apparent from the context.

Marker 5 is the human smooth muscle cell calponin protein. The gene coding for this protein has been cloned and sequenced (Miano et al., Gene 197:215-224 (1997)). These authors identify smooth muscle cell calponin as a muscle protein. There is no teaching in this paper of an association with prostate cancer. The sequence of smooth muscle cell calponin is disclosed in Miano et al., and is present in the EMBL database under accession number HS3701910. For the purpose of this application, the gene and amino acid sequences and sequence positions referred to herein, refers to that in the EMBL database under accession number HS3701910, unless stated otherwise or apparent from the context.

Marker 6 is the human mRNA KIAA0211. This cDNA has been cloned and sequenced (Nagase et al., DNA Res. 3:321-329 (1996)). These authors identify KIAA0211 as a brain transcript. There is no teaching in this paper of an association with prostate cancer. The sequence KIAA0211 is disclosed in Nagase et al., and is present in the EMBL database under accession number HSD966. For the purpose of this application, the gene and amino

acid sequences and sequence positions referred to herein, refers to that in the EMBL database under accession number HSD966, unless stated otherwise or apparent from the context.

Marker 7 is the human mRNA KIAA0120. This cDNA has been cloned and sequenced (Nagase et al., DNA Res. 2:37-43 (1995)). These authors identify KIAA0120 as a brain transcript. There is no teaching in this paper of an association with prostate cancer. The sequence KIAA0120 is disclosed in Nagase et al., and is present in the EMBL database under accession number HSORFF. For the purpose of this application, the gene and amino acid sequences and sequence positions referred to herein, refers to that in the EMBL database under accession number HSORFF, unless stated otherwise or apparent from the context.

Marker 8 is the human ribosomal protein S25. The gene coding for this protein has been cloned and sequenced (Li et al., Gene 107:329-333 (1991)). These authors identify ribosomal protein S25 as a leukaemia cell line protein. There is no teaching in this paper of an association with prostate cancer. The sequence of human ribosomal protein S25 is disclosed in Li et al., and is present in the EMBL database under accession number HSRPS25. For the purpose of this application, the gene and amino acid sequences and sequence positions referred to herein, refers to that in the EMBL database under accession number HSRPS25, unless stated otherwise or apparent from the context.

Marker 9 is the human 80K-H protein (kinase C substrate). The gene coding for this protein has been cloned and sequenced (Sakai et al., Genomics 5:309-315 (1989)). These authors identify 80K-H protein (kinase C substrate) as a fibroblast and epidermal carcinoma cell protein. There is no teaching in this paper of an association with prostate cancer. The sequence of 80K-H protein (kinase C substrate) is disclosed in Sakai et al., and is present in the EMBL database under accession number HSG19P1A. For the purpose of this application, the gene and amino acid sequences and sequence positions referred to herein, refers to that in the EMBL database under accession number HSG19P1A, unless stated otherwise or apparent from the context.

Marker 10 is the human alpha-2-macroglobulin protease inhibitor protein. The gene coding for this protein has been cloned and sequenced (Kan et al., Proc. Natl. Acad. Sci. U.S.A. 82:2282-2286 (1985)). These authors identify alpha-2-macroglobulin protease inhibitor protein as a serum protein. There is no teaching in this paper of an association with prostate cancer. The sequence of alpha-2-macroglobulin protease inhibitor protein is disclosed in Kan et al., and is present in the EMBL database under accession number

M11313. For the purpose of this application, the gene and amino acid sequences and sequence positions referred to herein, refers to that in the EMBL database under accession number M11313, unless stated otherwise or apparent from the context.

Marker 11 is the human SNC73 protein. The gene coding for SNC73 protein has been cloned and sequenced (Zheng S. et al. direct EMBL submission AF067420). These authors identify SNC73 as an mRNA downregulated in colorectal cancer. There is no teaching in this entry of an association with prostate cancer. The sequence of SNC73 is disclosed in the EMBL database under accession number AF067420. For the purpose of this application, the gene and amino acid sequences and sequence positions referred to herein, refers to that in the EMBL database under accession number AF067420, unless stated otherwise or apparent from the context.

Marker 12 is the human myosin light chain protein. The gene coding for human myosin light chain protein has been cloned and sequenced (Kumar et al., Biochemistry 28:4027-4035 (1989)). These authors identify human myosin light chain protein as a muscle protein. There is no teaching in this paper of an association with prostate cancer. The sequence of human myosin light chain protein is disclosed in Kumar et al., and is present in the EMBL database under accession number HSMLC2A. For the purpose of this application, the gene and amino acid sequences and sequence positions referred to herein, refers to that in the EMBL database under accession number HSMLC2A, unless stated otherwise or apparent from the context.

Marker 13 is a human adult uterus protein. The gene DKFZp586DO918 has been cloned and sequenced and submitted to EMBL database under Accession No. HSM800155. There is no disclosure of an association with prostate cancer. For the purpose of this application, the gene and predicted amino acid sequences and sequence positions referred to herein, refers to that in the EMBL database under accession number HSM800155 unless stated otherwise or apparent from the context.

Marker 14 is the laminin B2 protein. The gene for laminin B2 has been cloned and sequenced by Kallunki et al., J. Biol. Chem. 266:221-228 (1991). These authors identify laminin B2 protein as associated with myeloid disease. There is no teaching in this paper of an association with prostate cancer. The sequence of laminin B2 protein is disclosed in Kallunki et al. and is present in the EMBL database under accession number HSLB2A26. For the purpose of this application, the gene and amino acid sequences and sequence positions

referred to herein, refers to that in the EMBL database under accession number HSLB2A26, unless stated otherwise or apparent from the context.

Marker 15 is the human PRSM1 protein. The gene for PRSM1 has been cloned and sequenced (Scott et al., Gene 174:135-143 (1996)). These authors identify PRSM1 as a putative metallopeptidase. There is no teaching in this paper of any association with prostate cancer. The sequence of PRSM1 is disclosed in Scott et al. and is present in the EMBL database under accession number HSU58048. For the purpose of this application, the gene and amino acid sequences and sequence positions referred to herein, refers to that in the EMBL database under accession number HSU58048, unless stated otherwise or apparent from the context.

Marker 16 is the human ribosomal large subunit L12 protein. The gene for ribosomal large subunit L12 protein has been cloned and sequenced (Chu et al., Nucleic Acids Res. 21:749-749 (1993)). These authors identify human ribosomal large subunit L12 as a ribosomal protein. There is no teaching in this paper of any association with prostate cancer. The sequence of human ribosomal large subunit L12 is disclosed in Chu et al. and is present in the EMBL database under accession number HSL12A. For the purpose of this application, the gene and amino acid sequences and sequence positions referred to herein, refers to that in the EMBL database under accession number HSL12A, unless stated otherwise or apparent from the context.

Marker 17 is the human 7SK45 RNA. The gene for 7SK45 RNA has been cloned and sequenced (Murphy et al., Nucleic Acids Res. 14:9243-9260 (1986)). These authors identify human 7SK45 as a small cytoplasmic RNA. There is no teaching in this paper of any association with prostate cancer. The sequence of 7SK45 is disclosed in Murphy et al. and is present in the EMBL database under accession number HS7K45. For the purpose of this application, the gene sequences and sequence positions referred to herein, refers to that in the EMBL database under accession numbers: HS7K45, unless stated otherwise or apparent from the context.

Marker 18 is the human KIAA0588 protein (human protocadherin gamma). This protein is abundantly expressed in brain cells. The gene for this protein has been cloned and sequenced (Nagase et al., DNA Research. 5:31-39 (1998)). There is no teaching in this paper of an association with prostate cancer. The sequence of the KIAA0588 protein is present in the EMBL database under accession number AB011160. For the purpose of this application,

the gene sequence and sequence positions referred to herein, refers to that in the EMBL database under accession number AB011160, unless stated otherwise or apparent from the context.

Marker 19 is the human factor V protein. This gene has been cloned and sequenced (Jenny et al., Proc. Natl. Acad. Sci. U.S.A. 84:4846-4850 (1987)). The authors identify human factor V protein as a fetal liver protein. There is no teaching in this paper of an association with prostate cancer. The sequence of human factor V protein is disclosed in Jenny et al., and is present in the EMBL database under the accession number HSFAV. For the purpose of this application, the gene and amino acid sequences and sequence positions referred to herein, refers to that in the EMBL database under accession number HSFAV, unless stated otherwise or apparent from the context.

Marker 20 is the SAP-1 protein. The gene for this protein has been cloned and sequenced (Dalton & Treisman, Cell 68:597-612 (1992)). The authors identify SAP-1 as a protein recruited by serum response factor to the c-fos serum response element. There is no teaching in this paper of an association with prostate cancer. The sequence of SAP-1 is disclosed in Dalton and Treisman and is present in the EMBL database under the accession number HSSERREFB. For the purpose of this application, the gene and amino acid sequences and sequence positions referred to herein, refers to that in the EMBL database under accession number HSSERREFB, unless stated otherwise or apparent from the context.

Marker 21 is the Tip 60 protein, 60kDa. The gene for this protein has been cloned and sequenced (Kamine et al., Virology 216:357-366 (1996)). These authors identify Tip 60 as a human Tat interacting protein. There is no teaching in this paper of an association with prostate cancer. The sequence of Tip 60 is disclosed in Kamine et al. and is present in the EMBL database under the accession number HSU74667. For the purpose of this application, the gene and amino acid sequences and sequence positions referred to herein, refers to that in the EMBL database under accession number HSU74667, unless stated otherwise or apparent from the context.

Marker 22 is the SRP14 protein subunit. The gene for this protein has been cloned and sequenced (Leffers H. 1993, direct submission to EMBL). This author identifies SRP14 as a human signal recognition particle subunit. There is no teaching in this paper of and association with prostate cancer. The sequence of SRP14 is disclosed in Leffers and is present in the EMBL database under accession number HSSRP14A. For the purpose of this

application, the gene and amino acid sequences and sequence positions referred to herein, refers to that in the EMBL database under accession number HSSRP14A, unless stated otherwise or apparent from the context.

Marker 23 is bone small proteoglycan I (biglycan) protein. The gene for this protein has been cloned and sequenced (Fisher et al., J. Biol. Chem. 264:4571-4576 (1989)). These authors identify bone small proteoglycan I (biglycan) as a protein derived from human bone-derived cells. There is no teaching in this paper of an association with prostate cancer. The sequence of bone small proteoglycan I (biglycan) protein is disclosed in Fisher et al. and is present in the EMBL database under accession number HSHPGI. For the purpose of this application, the gene and amino acid sequences and sequence positions referred to herein, refers to that in the EMBL database under accession number HSHPGI, unless stated otherwise or apparent from the context.

Marker 24 is the human KIAA0045 gene. The gene for this protein has been cloned and sequenced (Nomura N et al., DNA Res 1:223-229 (1994)). These authors identify KIAA0045 as a protein from the human immature myeloid cell line KG-1. There is no teaching in this paper of an association with prostate cancer. The sequence of KIAA0045 is disclosed in the EMBL database under accession number HSKG1C (accession no. D28476). For the purpose of this application, the gene and amino acid sequences and sequence positions referred to herein, refers to that in the EMBL database under accession number HSKG1C, unless stated otherwise or apparent from the context.

Marker 25 is b-cell receptor associated protein, also known as REA (repressor of estrogen activity). The gene for this protein has been cloned and sequenced (Montano et al., Proc Natl Acad Sci U.S.A. 96:6947-6952 (1999)). These authors identify REA as a protein derived from human breast. There is no teaching in this paper of an association with prostate cancer. The sequence of b-cell receptor associated /REA protein, is disclosed in Montano et al. and is present in the EMBL database under accession number AF150962. For the purpose of this application, the gene and amino acid sequences and sequence positions referred to herein, refers to that in the EMBL database under accession number AF150962, unless stated otherwise or apparent from the context.

Marker 26 is the cystatin B protein. The gene for cystatin B has been cloned and sequenced (Pennacchio et al., Science 271:1731-1734 (1996)). These authors identify cystatin B protein as associated with progressive myoclonus epilepsy disease. There is no teaching in

this paper of an association with prostate cancer. The sequence of cystatin B protein is disclosed in Pennachio et al. and is present in the EMBL database under accession number HSCST4BA. For the purpose of this application, the gene and amino acid sequences and sequence positions referred to herein, refers to that in the EMBL database under accession number HSCST4BA unless stated otherwise or apparent from the context.

Although each of the nucleic acid and/or protein sequences of the above 26 Markers are in the public domain (via EMBL, other databases or scientific publications) the nucleic acid sequence of each Marker as present in the EMBL database is provided in the sequence listing herein, retaining the same numbering system, i.e. Marker 1 = SEQ ID NO: 1.

Each of the above disclosures provides a cDNA and, where a protein is encoded, an amino acid sequence. It will be apparent that the sequences disclosed for each of the 26 Markers is the representative sequence assigned to the gene in question. In normal individuals there are two copies of each gene, a maternal and paternal copy, which will likely have some sequence differences, moreover within a population there will exist numerous allelic variants of the gene sequence. It will be appreciated that the diagnostic methods and other aspects of this invention extend to the detection etc. of any of these sequence variants. Preferred sequence variants are those that possess at least 90% and preferably at least 95% sequence identity (nucleic acid or amino acid) to the Marker sequence depicted in the particular EMBL Accession number identified above. Nucleic acid sequence identity can also be gauged by hybridisation studies whereby, under stringent hybridisation and wash conditions, only closely related sequences (for example, those with >90% identity) are capable of forming a hybridisation complex.

The levels of each of the Markers can be assessed from relative amounts of mRNA, cDNA, genomic DNA or polypeptide sequence present in the test sample. Where RNA is used, it may be desired to convert the RNA to a complementary cDNA and during this process it may be desirable to incorporate a suitable detectable label into the cDNA.

In a preferred embodiment the method of the invention relies on detection of mRNA transcript levels. This involves assessment of the relative mRNA transcript levels of the Marker in a sample, and comparison of sample data to control data. Each Marker transcript can be detected individually, or, is preferably detected amongst a panel of markers from which a transcript profile can be generated. These additional markers may include some or all of the other Markers specifically identified herein, it may include other known prostate disorder

markers such as PSA, prostatic acid phosphatase, prostate-specific membrane antigen or others (e.g. those reviewed in Gao et al., Prostate 31:264-281 (1977)); markers associated with cell proliferation, invasiveness, tumour angiogenesis, metastasis or any other aspect of carcinogenesis; or markers for other diseases and conditions. Levels of Marker mRNA in the test sample can be detected by any technique known in the art. These include Northern blot analysis, reverse transcriptase-PCR amplification (RT-PCR), microarray analysis and RNAse protection.

In one embodiment, levels of Marker RNA in a sample can be measured in a Northern blot assay. Here, tissue RNA is fractionated by electrophoresis, fixed to a solid membrane support, such as nitrocellulose or nylon, and hybridised to a probe or probes capable of selectively hybridising with the Marker RNA to be detected. The actual levels may be quantitated by reference to one or more control housekeeping genes. Probes may be used singly or in combination. This may also provide information on the size of mRNA detected by the probe. Housekeeping genes are genes which are involved in the general metabolism or maintenance of the cell, and are considered to be expressed at a constant level irrespective of cell type, physiological state or stage in the cell cycle. Examples of suitable housekeeping genes are: beta actin, GAPDH, histone H3.3 or ribosomal protein L13 (Koehler et al., Quantitation of mRNA by Polymerase Chain Reaction. Springer-Verlag, Germany (1995)).

To gauge relative expression levels, a control sample can be run alongside the test sample or, the test result/value can be compared to Marker expression levels expected in a normal or control tissue. These control values can be generated from prior test experiments using normal or control tissues, to generate mean or normal range values for each Marker.

In another embodiment, the Marker nucleic acid in a tissue sample is amplified and quantitatively assayed. The polymerase chain reaction (PCR) procedure can be used to amplify specific nucleic acid sequences through a series of iterative steps including denaturation, annealing of oligonucleotide primers (designed according to the published Marker sequence to be detected), and extension of the primers with DNA polymerase (see, for example, Mullis, et al., U.S. patent No. 4,683,202; Loh et al., Science 243:217 (1988)). In reverse transcriptase-PCR (RT-PCR) this procedure is preceded by a reverse transcription step to allow a large amplification of the number of copies of mRNA (Koehler et al., *supra*). Other known nucleic acid amplification procedures include transcription-based amplification systems (TAS) such as nucleic acid based sequence application (NASBA) and 3SR (Kwoh et

al., Proc Natl. Acad Sci USA 86:1173 (1989), Gingeras et al., PCT application WO 88/10315), the ligase chain reaction (LCR, see European Application No. 320308), Strand Displacement Amplification (SDA), "race", "one sided PCR" and others (Frohman, PCR Protocols: a Guide to Methods and Applications. Academic Press, NY (1990); Ohara et al., Proc Natl Acad Sci. USA 86:5673-5677 (1989)). Quantitation of RT-PCR products can be done while the reaction products are building up exponentially, and can generate diagnostically useful clinical data. In one embodiment, analysis is carried out by reference to one or more housekeeping genes which are also amplified by RT-PCR. Quantitation of RT-PCR product may be undertaken, for example, by gel electrophoresis visual inspection or image analysis, HPLC (Koehler et al., *supra*) or by use of fluorescent detection methods such as intercalation labelling, Taqman probe (Higuchi et al., Biotechnology 10:413-417 (1992)), Molecular Beacon (Piatek et al., Nature Biotechnol. 4:359-363 (1998)), primer or Scorpion primer (Whitcombe et al., Nature Biotech 17:804-807 (1999)). or other fluorescence detection method, relative to a control housekeeping gene or genes as discussed above.

In a preferred embodiment of the invention, the method of the invention involves assessing Marker transcript levels using microarray analysis. Probes are made that selectively hybridise to the sequences of the target Marker genes in the test sample. These probes, perhaps together with other Marker probes and control probes, are bound at discrete locations on a suitable support medium such as a nylon filter or microscope slide to form a transcript profiling array. The diagnostic method involves assessing the relative mRNA transcript levels of one or more of the Markers in a clinical sample. This can be done by radioactively labelling, or non-radioactively labelling the tissue mRNA, which can be optionally purified from total RNA, in any of a number of ways well known to the art (Sambrook et al., *supra*). The probes can be directed to any part or all of the target Marker mRNA. Example 1 illustrates the use of microarray/RNA transcript detection.

Microarray (also termed hybridisation array, gene array or gene chip) technology wherein nucleic acid molecules attached to solid substrates at predefined locations in small areas and at high density are used, in conjunction with hybridisation reactions, for identifying and discriminating target nucleic acid sequences, has advanced rapidly in the past few years. These chips or microarrays allow massive parallel data acquisition and are used, for example, in polymorphism detection, clinical mutation detection, expression monitoring, fingerprinting and sequencing.

A variety of methods are currently available for making arrays of biological molecules. The 'dot or slot blot' approach, whereby an ordered array of DNA is vacuum blotted using a manifold, or hand blotted by capillary action, onto a porous membrane, such as nylon or nitrocellulose has been around for many years (Maniatis et al., Molecular Cloning-A Laboratory Manual, First Edition, Cold Spring Harbor, (1982)). Methods for preparing a plurality of oligonucleotide sequences and for attaching these to solid supports at high density are also known in the art. For example, US Patent No. 4,562,157 describes a method of using photo-activatable cross-linking groups to immobilise pre-synthesised ligands on surfaces. Fodor et al. (Nature 364:555-556 (1993)) and US Patent No. 5,143,854 describe the 'light-directed chemical synthesis' method for synthesising ligands, including oligonucleotides, directly onto a substrate surface at the desired location. US 5,700,637 also describes methods for *in situ* synthesis of oligonucleotides on solid support surfaces. In addition, such methods for preparing microarrays can easily be automated. International Publication No. WO 95/35505 discloses an automated capillary dispensing device and method for applying biological macromolecules to solid supports. International Publication No. WO 97/44134 also describes devices for delivery of small volumes of liquid (which may contain biological macromolecules) in a precise manner to produce microsized spots on a solid surface to generate a microarray. Similarly, International Publication No. WO 98/10858 also describes an apparatus for the automated synthesis of molecular arrays. Techniques exist for applying the oligonucleotides to the array at high density and for example, techniques exist for applying well in excess of 10^3 distinct polynucleotides per 1 cm^2.

Microarray technology makes it possible to simultaneously study the expression of many thousands of genes in a single experiment. Analysis of gene expression in human tissue (e.g. biopsy tissue) can assist in the diagnosis and prognosis of disease and the evaluation of risk for disease. A comparison of levels of expression of various genes from patients with defined pathological disease conditions with normal patients enables an expression profile, characteristic of disease, to be created. There are currently two main approaches to analyse gene expression using microarrays. In the first approach, cDNA fragments, often generated by PCR, for each of the genes under study are attached to an array. Typically, mRNA isolated from the test samples (i.e. induced or un-induced) is reverse transcribed into cDNA with incorporation of a fluorescent label. The cDNA is sheared and hybridised to the array. If a control test sample is to be run at the same time, mRNA from this sample can be reverse

transcribed with incorporation of a different fluorescent label to enable direct comparison of the expression level of each test gene on the same array (see WO 95/35505). The second approach is similar to the first except that an oligonucleotide microarray is used. Because of the differences in hybridisation properties between short oligonucleotide probes, each gene is usually represented by several oligonucleotides (typically 20 or more) on the chip. In addition, a partner control oligonucleotide identical to each oligonucleotide, except for one of the central nucleotides, is included on the array to serve as an internal control for hybridisation sensitivity.

With oligonucleotide arrays, the capture oligonucleotide molecules may be individually synthesised on a standard oligonucleotide synthesiser. These oligonucleotide (oligos) may then be attached to the substrate matrix by any of a variety of techniques known in the art such as by using photochemical reagents, such as disclosed in US Patent No. 4,542,102 and 4,713,326. US Patent No. 4,562,157 also describes a method of using photo-activatable cross-linking groups to immobilise pre-synthesised ligands on surfaces. Alternatively, the oligonucleotides can be synthesised directly onto the solid surface using photolithography techniques, such as disclosed in US Patent No. 5,143,854, or other methods such as disclosed in US Patent No. 5,700,637, or International Publication No's: WO 95/35505, WO 97/44134 or WO 98/10858. Schena et al. (TIBTECH 16:301-306 (1998)) reviews the recent advances in microarray technology including the various means of constructing these arrays. Problems facing current photolithographic techniques for oligonucleotide synthesis involve the low yield of synthesis at each synthesis step, and also the efficiency of nucleotide addition at each synthesis step which can range from about 80% to 97%, with purines generally having a lower efficiency than pyrimidines (Thomas & Burke. Exp. Opin. Ther. Patents 8(5):503-508, (1998)). When constructing an array with relatively few capture oligonucleotides, say 500 or less, or with long oligonucleotides say 30-mers or more it may be preferable to synthesise the oligos separately and affix them to the solid support later rather than in situ synthesis.

In another embodiment of the invention, total Marker RNA or DNA is quantified and compared to levels in control tissue or expected levels from pre tested standards. DNA and/or RNA may be quantified using techniques well known in the art. Messenger RNA is often quantitated by reference to internal control mRNA levels within the sample, often relative to housekeeping genes (Koehler et al., *supra*).

In transcript profiling, several or many mRNAs are detected in the same procedure. One or more of these mRNAs may be diagnostic of cancer cells (i.e prostate cancer) in a tissue sample. In one embodiment, combinations of probes can be used to classify the cancer cells into clinically relevant types, according to the complex expression pattern of Markers measured on the array. Such classification may help to define which tumours are growing aggressively, or harbour latent signs of aggression, or are less aggressive or benign. The array provides a quantitative measure of Marker RNAs. This is done by comparison of Marker RNA signal with control signal. In a preferred embodiment hybridisation signals generated are measured by computer software analysis of images on phosphorimage screens exposed to radioactively labelled tissue RNA hybridised to a microarray of probes on a solid support such as a nylon membrane. In another, quantities are measured by densitometry measurements of radiation-sensitive film (e.g. X-ray film), or estimated by visual means. In another embodiment quantities are measured by use of fluorescently labelled probe, which may be a mixture of tumour and normal RNA differentially labelled with different fluorophores, allowing quantities of Marker mRNAs to be expressed as a ratio versus the normal level. The solid support in this type of experiment is generally a glass microscope slide, and detection is by fluorescence microscopy and computer imaging.

The detection of specific interactions may be performed by detecting the positions where the labelled target sequences are attached to the array. Radiolabelled probes can be detected using conventional autoradiography techniques. Use of scanning autoradiography with a digitised scanner and suitable software for analysing the results is preferred. Where the label is a fluorescent label, the apparatus described, e.g. in International Publication No. WO 90/15070, US Patent No. 5, 143,854 or US Patent No. 5,744,305 may be advantageously applied. Indeed, most array formats use fluorescent readouts to detect labelled capture:target duplex formation. Laser confocal fluorescence microscopy is another technique routinely in use (M.J.Kozal et al., Nature Medicine 2:753-759 (1996)). Mass spectrometry may also be used to detect oligonucleotides bound to a DNA array (Little et al, Analytical Chemistry 69: 4540-4546, (1997)). Whatever the reporter system used, sophisticated gadgetry and software may be required in order to interpret large numbers of readouts into meaningful data (such as described, for example, in US Patent No. 5,800,992 or International Publication No. WO 90/04652).

Typically microarray data is expressed as a relative value. In one embodiment a housekeeping gene or a set of housekeeping genes which remain constant independent of tissue type are used as a normalisation standard. In another embodiment the aggregate signal from the whole array or a subset of probes on the array used as a normalisation standard. By taking the ratio of Marker against the control normalisation standard for each diseased tissue sample, values for each Marker are generated which can be compared to known standards for normal, diseased or tumour tissue, and can be analysed for diagnostic levels of expression.

In the preferred embodiment of the microarray test, each Marker RNA measurement is generated as a value relative to an internal standard (i.e. a housekeeping gene) known to be constant or relatively constant. The histone H3.3 and ribosomal protein L19 housekeeping genes have been shown to be cell-cycle independent and constitutively expressed in all tissues (Koehler et al., *supra*). For normalisation of data, several different housekeeping genes can be used to generate an average housekeeping measurement.

A microarray or RT-PCR test to detect prostate disorder Markers can be used where tissue samples containing mRNA are available. Prostate biopsies are often taken for pathological analysis where initial screening using PSA, DRE or ultrasound indicates possible abnormality. Needle biopsies taken in this way typically weigh 100mg, and would typically yield 50-100mg total RNA using standard methods (e.g. using Trizol, Gibco BRL, UK). This would generate enough material for microarray analysis: for example a typical filter microarray of 4000 probes measuring 70x110 mm will require 20mg total RNA. A small diagnostic microarray featuring 50 probes could therefore require less than 1μg total RNA. RT-PCR tests require less starting material, typically 10ng, and could be used for a set or subset of Markers where biopsy material is limited or precious. Another advantage of RT-PCR is that it is quicker to perform and analyse than microarray (<1 day as compared to >1 week). Often, much more tissue is available for analysis.

Many prostate cancers are detected incidentally when transurethral resection of the prostate (TURP) is carried out to alleviate bladder outflow blockage. TURP tissue chips generated from this procedure typically weigh >100mg, and several tens to hundreds of chips are generated in a typical operation.

Samples for RNA extraction must be treated promptly to avoid RNA degradation (Sambrook et al., *supra*). This entails either prompt extraction using e.g. phenol-based reagents or snap freezing in e.g. liquid Nitrogen. Samples can be stored at -70°C or less until

RNA can be extracted at a later date. Proprietary reagents are available which allow tissue or cells to be conveniently stored for several days at room temperature and up to several months at 4°C (e.g. RNAlater, Ambion Inc., TX). Prior to extraction, methods such as grinding, blending or homogenisation are used to dissipate the tissue in a suitable extraction buffer. Typical protocols then use solvent extraction and selective precipitation techniques. Example 1 describes such a method.

In another embodiment, tissue is directly analysed for the presence of Marker nucleic acid. This can be by *in situ* hybridisation, where sections of tissue may be interrogated with specific probes to determine which morphological cell type in the sample displays a marker nucleic acid, such as sequences corresponding to Markers 1-26. *In situ* hybridisation typically comprises 3 steps. Firstly tissue is fixed, and sections are prepared by standard treatments known to those in the art (Polack and McGee, *In situ* hybridisation: principles and practice. Oxford University Press, 1998). Secondly, Marker mRNA and amplified Marker DNA can be detected by hybridisation with e.g. a biotin-, digoxygenin- or radio-labelled Marker probe, typically for 2-16 hours at 42°C in a suitable hybridisation buffer. A typical buffer might contain 50% formamide, 5% dextran sulphate, 2xSSC and 10-20ng of probe per 7µl (Herrington and McGee, Diagnostic Molecular Pathology. IRL Press, Oxford, 1992). The probe can be made of DNA or RNA. Lastly, following stringency washes, the probes in hybridisation complexes are detected with chromogenic or fluorescent reagents, which can be visualised by microscopy, or by autoradiography in the case of radiolabelled probes. Signal amplification systems using e.g. tyramide can be used to increase sensitivity (Polack and McGee, *supra). In situ* PCR using oligonucleotide probes complementary to the nucleic acid of any of Markers 1 - 26 is therefore envisaged. *In situ* hybridisation can follow *in situ* PCR, giving greater specificity (Polack and McGee, *supra*). Techniques for quantitation of signal, and quantification of positive cells in a section are available to the pathologist using image analysis. *In situ* Marker visualisation permits localisation of signal in mixed-tissue specimens commonly found in tumours, and is compatible with many histological staining procedures. In one embodiment, several probes can be differentially labelled and hybridised simultaneously to the same section, and detected using appropriate reagents. In another, serial sections from the same sample can be analysed with a panel of probes. Quantitation may involve comparison with one or more control housekeeping genes as discussed above.

In another embodiment oligonucleotide probes capable of selectively hybridising to nucleic acid of one or other of Markers 1 - 26, can be used to detect levels of Marker gene expression.

Primers or probes for use in any of the methods of the invention may be manufactured using any convenient method of synthesis. Examples of such methods may be found in standard textbooks, for example "Protocols for Oligonucleotides and Analogues; Synthesis and Properties," Methods in Molecular Biology Series; Volume 20; Ed. Sudhir Agrawal, Humana ISBN: 0-89603-247-7 (1993); 1st Edition. If required the primer(s) may be labelled to facilitate detection.

There are many conventional detectable labels such as radioisotopes, fluorescent labels, chemiluminescent compounds, labelled binding proteins, magnetic labels, spectroscopic Markers and linked enzymes that might be used. One particular example well known in the art is end-labelling with ^{32}P. Fluorescent labels are preferred because they are less hazardous than radiolabels, they provide a strong signal with low background and various different fluorophors capable of absorbing light at different wavelengths and/or giving off different colour signals exist to enable comparative analysis in the same analysis. For example, fluorescein gives off a green colour, rhodamine gives off a red colour and both together give off a yellow colour.

Preferred primers for amplification are between 15 and 60 bp, more preferably between 17 and 35bp in length. Probe sequences can be anything from about 25 nucleotides in length upwards. If the target sequence is a gene of 2kb in size the probe sequence can be the complete gene sequence complement and thus may also be 2kb in size. Preferably, the probe sequence is a genomic, or more preferably a cDNA, fragment of the target sequence and may be between 50 and 2000 bp, preferably between 200 and 750 bp. It will be appreciated that multiple probes each capable of selectively hybridising to a different target sequence of the individual Marker nucleic acid, maybe across the complete length of the Marker sequence, may be prepared and used together in a diagnostic test. The primers or probes may be completely homologous to the target sequence or may contain one or more mismatches to assist specificity in binding to the correct template sequence. Any sequence which is capable of selectively hybridising to the target sequence of interest may be used as a suitable primer or probe sequence. It will also be appreciated that the probe or primer sequences must hybridise to the target template nucleic acid. If the target nucleic acid is double stranded (genomic or

cDNA) then the probe or primer sequence can hybridise to the sense or antisense strand. If however the target is mRNA (single stranded sense strand) the primer/probe sequence will have to be the antisense complement.

An example of a suitable hybridisation solution when a nucleic acid is immobilised on a nylon membrane and the probe nucleic acid is greater than 500 bases or base pairs is: 6 x SSC (saline sodium citrate), 0.5% SDS (sodium dodecyl sulphate), 100μg/ml denatured, sonicated salmon sperm DNA. The hybridisation being performed at 68°C for at least 1 hour and the filters then washed at 68°C in 1 x SSC, or for higher stringency, 0.1 x SSC/0.1% SDS.

An example of a suitable hybridisation solution when a nucleic acid is immobilised on a nylon membrane and the probe is an oligonucleotide of between 12 and 50 bases is: 3M trimethylammonium chloride (TMACl), 0.01M sodium phosphate (pH 6.8), 1mM EDTA (pH 7.6) , 0.5% SDS,100μg/ml denatured, sonicated salmon sperm DNA and 0.1 dried skimmed milk. The optimal hybridisation temperature (Tm) is usually chosen to be 5°C below the Ti of the hybrid chain. Ti is the irreversible melting temperature of the hybrid formed between the probe and its target. If there are any mismatches between the probe and the target, the Tm will be lower. As a general guide, the recommended hybridisation temperature for 17-mers in 3M TMACl is 48-50°C; for 19-mers,it is 55-57°C; and for 20-mers, it is 58-66°C.

A suitable hybridisation protocol is described in Example 5, however, operable variations to this method will be apparent to the person skilled in the art.

Where the cDNA molecules of Markers 1 - 26 encode proteins or parts of cellular proteins, these may themselves act as prostate disease Markers. To detect proteins in tissue, cells, body fluids or extracts of these sample types, specific antibody can be used. These antibodies can be prepared using the Marker protein/polypeptides.

Methods of making and detecting labelled antibodies are well known (Campbell; Monoclonal Antibody Technology, in: Laboratory Techniques in Biochemistry and Molecular Biology, Volume 13. Eds: Burdon R et al. Elsevier, Amsterdam (1984)). The term antibody includes both monoclonal antibodies, which are a substantially homogeneous population, and polyclonal antibodies which are heterogeneous populations. The term also includes inter alia, humanised and chimeric antibodies. Monoclonal antibodies to specific antigens may be obtained by methods known to those skilled in the art, such as from hybridoma cells, phage display libraries or other methods. Monoclonal antibodies may be inter alia, human, rat or

mouse derived. For the production of human monoclonal antibodies, hybridoma cells may be prepared by fusing spleen cells from an immunised animal, e.g. a mouse, with a tumour cell. Appropriately secreting hybridoma cells may thereafter be selected (Koehler & Milstein, Nature 256:495-497 (1975); Cole et al., "Monoclonal antibodies and Cancer Therapy", Alan R Liss Inc, New York N.Y. pp 77-96 (1985)). Such antibodies may be of any immunoglobulin class including IgG, IgM, IgE, IgA, IgD and any subclass thereof.

Polyclonal antibodies can be generated by immunisation of an animal (such as a mouse, rat, goat, horse, sheep etc) with an antigen, such as one of the Marker protein used in this invention.

The Marker polypeptides can be prepared by various techniques known to the person skilled in the art. RNA transcripts can be used to prepare a polypeptide of the invention by *in vitro* translation techniques according to known methods (Sambrook *et al. supra*). Alternatively, the Marker polypeptides can be synthesised chemically. For example, by the Merryfield technique (J. Amer. Chem. Soc. **85**:2149-2154, (1968)). Numerous automated polypeptide synthesisers, such as Applied Biosystems 431A Peptide Synthesizer also now exist. Alternatively, and preferably, the Marker polypeptides are produced from a nucleotide sequence encoding the polypeptide using recombinant expression technology. A variety of expression vector/host systems may be used to express the Marker coding sequences. These include, but are not limited to microorganisms such as bacteria expressed with plasmids, cosmids or bacteriophage; yeasts tranformed with expression vectors; insect cell systems transfected with baculovirus expression systems; plant cell systems transfected with plant virus expression systems, such as cauliflower mosaic virus; or mammalian cell systems (for example those transfected with adenoviral vectors); selection of the most appropriate system is a matter of choice. Preferably, the Marker protein is expressed in eukaryotic cells, especially mammalian, insect and yeast cells. Mammalian cells provide post-translational modifications to recombinant Marker proteins, which include folding and/or phosphorylation.

Expression vectors usually include an origin of replication, a promoter, a translation initiation site, optionally a signal peptide, a polyadenylation site, and a transcription termination site. These vectors also usually contain one or more antibiotic resistance Marker gene(s) for selection. As noted above, suitable expression vectors may be plasmids, cosmids or viruses such as phage or retroviruses. The coding sequence of the polypeptide is placed under the control of an appropriate promoter, control elements and transcription terminator so

that the nucleic acid sequence encoding the polypeptide is transcribed into RNA in the host cell transformed or transfected by the expression vector construct. The coding sequence may or may not contain a signal peptide or leader sequence for secretion of the polypeptide out of the host cell. Expression and purification of the Marker polypeptides can be easily performed using methods well known in the art (for example as described in Sambrook et al. *supra*).

The Marker polypeptides so produced can then be used to inoculate animals, from which serum samples, containing the specific antibody against the introduced Marker protein/polypeptide, can later be obtained.

Rodent antibodies may be humanised using recombinant DNA technology according to techniques known in the art. Alternatively, chimeric antibodies, single chain antibodies, Fab fragments may also be developed against the polypeptides of the invention (Huse et al., Science 256:1275-1281 (1989)), using skills known in the art. Antibodies so produced have a number of uses which will be evident to the molecular biologist or immunologist skilled in the art. Such uses include, but are not limited to, monitoring enzyme expression, development of assays to measure enzyme activity and use as a therapeutic agent. Enzyme linked immunosorbant assays (ELISAs) are well known in the art and would be particularly suitable for detecting the Marker proteins or polypeptide fragments thereof in a test sample.

The expression system described in Example 3 can produce protein for use as an antigen for the generation of antibodies for use in an ELISA assay to detect Marker protein in body fluids or by immunohistochemistry (as described in Example 4) or other means. In addition, an antibody could be used individually or as part of a panel of antibodies, together with a control antibody which reacts to a common protein, on a dipstick or similar diagnostic device.

Levels of Marker gene expression can also be detected by screening for levels of polypeptide (Marker protein). For example, monoclonal antibodies immunoreactive with a Marker protein can be used to screen a test sample. Such immunological assays can be done in any convenient format known in the art. These include Western blots, immunohistochemical assays and ELISA assays. Functional assays can also be used, such as protein binding determinations.

In another preferred embodiment antibodies directed against a Marker protein or proteins can be used, to detect, prognose, diagnose and stage prostate cancer or its precursor lesions, or related prostate disorders. Various histological staining methods known in the art,

including immunochemical staining methods, may also be used. Silver stain is but one method of detecting Marker proteins. For other staining methods useful in the present invention see, for example, A Textbook of Histology, Eds. Bloom and Fawcett, W.B. Saunders Co., Philadelphia (1964).

According to a further aspect of the invention there is provided use of an antibody selective for a Marker protein selected from the group consisting of: Marker 1 - Marker 26, in an assay to diagnose or prognose or monitor a prostate disorder.

Once an individual is diagnosed as suffering from a prostate disorder they may be subjected to any of a number of therapeutic treatments including antibody therapy, antisense therapy, chemotherapy etc. Antibodies targeted to the Marker proteins identified herein can be used to study the effects of the therapeutic treatment on Marker expression levels and thus may provide a guide to efficacy of the therapy.

According to a further aspect of the invention there is provided use of an antibody selective for a Marker protein selected from the group consisting of: Marker 1 - Marker 26, in an assay to monitor therapeutic efficacy.

In another aspect of the invention there is provided a method for treating a patient suffering from a prostate disorder comprising administering to said patient an effective amount of an antibody specific for a protein selected from the group consisting of: Marker 1 - Marker 26. In a preferred embodiment the prostate disorder is prostate cancer.

One therapeutic means of inhibiting or dampening the expression levels of a particular gene (for example one of the Markers identified herein) is to use antisense therapy. Antisense therapy utilises antisense nucleic acid molecules that are synthetic segments of DNA or RNA ("oligonuclotides"), designed to mirror specific mRNA sequences and block protein production. Once formed, the mRNA binds to a ribosome, the cell's protein production "factory" which effectively reads the RNA sequence and manufactures the specific protein molecule dictated by the gene. If an antisense molecule is delivered to the cell (for example as native oligonucleotide or via a suitable antisense expression vector), it binds to the messenger RNA because its sequence is designed to be a complement of the target sequence of bases. Once the two strands bind, the mRNA can no longer dictate the manufacture of the encoded protein by the ribosome and is rapidly broken down by the cell's enzymes, thereby freeing the antisense oligonucleotide to seek and disable another identical messenger strand of mRNA.

Thus, according to another aspect of the invention there is provided a method for treating a patient suffering from a prostate disorder comprising administering to said patient an effective amount of an anti-sense molecule capable of binding to the mRNA of a Marker gene selected from the group consisting of: Marker 1 - Marker 26, and inhibiting expression of the protein product of the Marker gene.

Complete inhibition of protein production is not essential, indeed may be detrimental. It is likely that inhibition to a state similar to that in normal tissues would be desired.

This aspect of antisense therapy is particularly applicable if the prostate disorder is a direct cause of over-expression of the Marker gene(s) in question, although it is equally applicable if said Marker gene(s) indirectly cause the prostate disorder. Having identified the particular Marker genes (1 - 26) over-expressed in prostate disorders, and with knowledge of the gene and mRNA sequence the person skilled in the art is able to design suitable antisense nucleic acid therapeutic molecules and administer them as required.

Antisense oligonucleotide molecules with therapeutic potential can be determined experimentally using well established techniques. To enable methods of down-regulating expression of a Marker gene of the present invention in mammalian cells, an example antisense expression construct can be readily constructed for instance using the pREP10 vector (Invitrogen Corporation). Transcripts are expected to inhibit translation of the gene in cells transfected with this type of construct. Antisense transcripts are effective for inhibiting translation of the native gene transcript, and capable of inducing the effects (e.g., regulation of tissue physiology) herein described. Oligonucleotides which are complementary to and hybridisable with any portion of Marker gene mRNA are contemplated for therapeutic use. U.S. Patent No. 5,639,595, "Identification of Novel Drugs and Reagents", issued Jun. 17, 1997, wherein methods of identifying oligonucleotide sequences that display *in vivo* activity are thoroughly described, is herein incorporated by reference. Expression vectors containing random oligonucleotide sequences derived from Marker polynucleotides are transformed into cells. The cells are then assayed for a phenotype resulting from the desired activity of the oligonucleotide. Once cells with the desired phenotype have been identified, the sequence of the oligonucleotide having the desired activity can be identified. Identification may be accomplished by recovering the vector or by polymerase chain reaction (PCR) amplification and sequencing the region containing the inserted nucleic acid material. Antisense molecules can be synthesised for antisense therapy. These antisense molecules may be DNA, stable

derivatives of DNA such as phosphorothioates or methylphosphonates, RNA, stable derivatives of RNA such as 2'-O-alkylRNA, or other oligonucleotide mimetics. U.S. Patent No. 5,652,355, "Hybrid Oligonucleotide Phosphorothioates", issued July 29, 1997, and U.S. Patent No. 5,652,356, "Inverted Chimeric and Hybrid Oligonucleotides", issued July 29, 1997, which describe the synthesis and effect of physiologically-stable antisense molecules, are incorporated by reference. Antisense molecules may be introduced into cells by microinjection, liposome encapsulation or by expression from vectors harboring the antisense sequence.

As noted above, antisense nucleic acid molecules may also be provided as RNAs, as some stable forms of RNA are now known in the art with a long half-life that may be administered directly, without the use of a vector. In addition, DNA constructs may be delivered to cells by liposomes, receptor mediated transfection and other methods known to the art.

Marker RNA measurements can also be carried out on blood or serum samples. Preferably, the RNA is obtained from a peripheral blood sample. In the case of soluble RNA in the blood serum, the low abundance of mRNA expected would necessitate a sensitive test such as RT-PCR (Kopreski et al., Clin Cancer Res 5:1961-5 (1999)). In some cases, there may be cancer cells present in the blood (Renneberg et al., Urol Res 27:23-7 (1999)). A whole blood gradient may be performed to isolate nucleated cells and total RNA is extracted such as by the Rnazole B method (Tel-Test Inc., Friendsworth, Tex.) or by modification of methods known in the art such as described in Sambrook et al., (*supra*).

Amplification of some genes occurs at the DNA level. Increases in the number of copies per cell of specific genes is well documented (Lewin, Genes, Fifth edition. Oxford University Press (1997)), and can occur in a stably maintained chromosomal location or as autonomously replicating extrachromosomal DNA. Increased transcript levels often occur in direct correlation with the amount of DNA for the specific gene amplified. This is a known mechanism for increased expression of several genes implicated in cancer, e.g. HER2 (de Cremoux et al., Int J Cancer 83:157-61 (1999)) and androgen receptor (Koivisto et al., Am J Pathol. 152:1-9 (1998)). Several methods are available for testing a sample for gene amplification. One embodiment of the invention is to use PCR to detect amplification of the Marker genes. This may be done quantitatively, for example via a comparison of the amplified gene in tumour tissue versus normal or cell line. The method employed is

analogous to quantitative RT-PCR already described. Marker gene amplification could also be detected *in situ* in cells by probe hybridisation. Fluorescently-labelled probes (used in Fluorescent in situ Hybridisation or FISH) are ideal for this purpose as they allow subnuclear localisation via microscopy, and permit estimates of gene copy number by subnuclear spot intensity or number (Mark et al. Exp Mol Pathol 66:170-8 (1999)). Quantitative analysis of DNA for amplifications of Marker genes can also be carried out by Southern analysis, a method which is widely known to those skilled in the art (Sambrook et al., *supra*). DNA can be extracted from clinical material using established methods (Sambrook et al., *supra*). The methods of the invention can therefore also be directed to measuring genomic DNA levels of one or more of the identified Markers.

If a Marker mRNA encodes a secreted protein, that protein is likely to be present in body fluids. Proteins, especially secreted proteins, can be quantified in blood, serum, urine, semen and other fluids. Specific antibodies can often detect abundant proteins in ELISA tests on body fluid samples without enrichment. Prostate-specific antigen (PSA) falls into this category. PSA is an important Marker produced by prostatic epithelial cells and almost always expressed in prostate cancer, though not exclusively so. Clinically significant levels of proteins such as PSA are defined by appropriate studies, and protein levels are typically given as ng per mL of sample. Serum protein tests with greater specificity for cancer are needed. Detection of rare proteins may require that the protein is concentrated by e.g. precipitation. Thus, in a further embodiment of the invention diagnosis or prognosis or prostate disorder, or stage monitoring of the prostate disorder or therapeutic efficacy assessment is performed by testing for aberrant levels of one or more Marker proteins of the invention, which Marker protein is a secreted protein, in a bodily fluid.

The inventors predict that Markers 9, 19 and 24 (at least) are secreted proteins.

All the essential materials and reagents required for detecting or monitoring prostate disorder Markers in a test sample may be assembled together in a kit. Such a kit may comprise one or more diagnostic cDNA probes or oligonucleotide primers together with control probes/primers. The kit may contain probes immobilised on a microarray substrate such as a filter membrane or silicon-based substrate. The kit may also comprise samples of total RNA derived from tissues of various physiological states, such as normal, BPH, confined tumour and metastatic tumour, for example, to be used as controls. The kit may also comprise appropriate packaging and instructions for use in the methods of the invention.

According to another aspect of the present invention there is provided a diagnostic kit for diagnosing or prognosing or monitoring a prostate disorder comprising, one or more diagnostic probe(s) and/or diagnostic primer(s) and/or antibodies capable of selectively hybridising or binding to one or more of the Markers 1 -26.

It will be appreciated that the term "diagnostic kit" is not intended to limit the kit to diagnostic use only, it also encompasses other uses such as in prognostic, stage monitoring and therapeutic efficacy studies.

In a preferred embodiment, the diagnostic (detection) probes are provided on a microarray.

Such kits may further comprise appropriate buffer(s) and/or polymerase(s) such as thermostable polymerases, for example taq polymerase. They may also comprise companion/constant primers and/or control primers or probes. A companion/constant primer is one that is part of the pair of primers used to perform PCR. Such primer usually complements the template strand precisely. The kit may also contain control normal prostate RNA labelled with one fluorophore (E.g.. Cy5). In use, patient RNA derived from biopsy or body fluids or cells can be labelled with another fluorophore (e.g. Cy3), the RNAs could then be mixed and hybridised to the array. Instrumentation to detect fluorescence ratio e.g. of. Cy3:Cy5, would be required to detect Marker over-expression. Preferred cDNA probes for detection of Marker mRNAs are selected from Markers 1-26.

In another embodiment the kit comprises one or more specific probes suitable for hybridisation to mRNA in tissue sections *in situ*. The kit may also contain hybridisation buffer and detection reagents for colourimetric or fluorescence microscopy detection. Preferred probes for detection of Marker mRNAs *in situ* are selected from the sequences of Markers 1-26.

In another embodiment the kit comprises a set of specific oligonucleotide primers, optionally labelled, for quantitation by RT-PCR of one or more Marker mRNAs. These primers may be Scorpion primers (Whitcombe et al., Nature Biotechnol. 17:804-807 (1999)) allowing accurate quantitation of specific PCR product. Alternatively, Taqman or Molecular Beacon probes may be provided in the kit for this purpose. One form of the kit would be a microtitre plate containing specific reagents in several wells, to which aliquots of extracted RNA could be pipetted. The microtitre plate could be thermocycled on a suitable machine, which could also be capable of reading fluorescence emissions from plate wells (e.g. Perkin

Elmer 7700). Preferred oligonucleotide primers for detection of Marker mRNAs are selected from Markers 1-26.

In another embodiment the kit comprises one or more antibodies specific for one or other of the Markers identified herein for use in immunohistochemical analysis.

In another embodiment the kit is an ELISA kit comprising one or more antibodies specific for one or other of the Markers identified herein.

In another aspect of the invention, one or more of the 26 Markers can also be used in biochemical assays to identify agents which modulate the activity of the Marker proteins. The design and implementation of such assays will be evident to the biochemist of ordinary skill. The protein, particularly if it is a biochemical enzyme, may be used to turn over a convenient substrate whilst incorporating or losing a labelled component to define a test system. Test compounds are introduced into the test system and measurements made to determine their effect on enzyme activity. Such assays are useful to identify inhibitors of the enzyme which may prove valuable as therapeutic agents.

The inventors believe that Marker 19 (Factor V) is an example of a suitable biochemical enzyme that can be used in a suitable biochemical assay to identify modulators.

In a further aspect of the invention, each of the Markers can be used to characterise cell cultures in a screen for therapeutic agents, such as a high throughput screen. Effects of test compounds may be assayed by changes in mRNA or protein of any of Markers 1-26. As described above, cells (i.e. mammalian, bacterial etc) can be engineered to express one of the Markers identified herein.

Thus, according to a further aspect of the invention there is provided a method of testing potential therapeutic agents for the ability to suppress a prostate disorder phenotype comprising contacting a test compound with a cell engineered to express one of the Markers identified herein; and determining whether said test compound suppressed expression of the Marker.

Thus, according to a further aspect of the invention there is provided a screening assay or method for identifying potential anti-prostate disorder therapeutic compounds comprising contacting an assay system capable of detecting the effect of a test compound against expression levels of one or more of the Markers selected from the group consisting of: Marker 1 - Marker 26, with a test compound and assessing the change in expression level of the particular Marker under study.

If the Markers are subject to regulatory pathways influenced by a primary therapeutic target, the Markers may act as surrogate Markers of drug action on this target. Such a Marker can be of use in a cell-culture system in a high throughput screen, or in a toxicity study, or in a clinical trial.

Compounds that modulate the expression of DNA or RNA of any of the Marker polypeptides may be detected by a variety of assay systems. A suitable assay system may be a simple "yes/no" assay to determine whether there is a change in expression of a reporter gene, such as beta-galactosidase, luciferase, green fluorescent protein or others known to the person skilled in the art (reviewed by Naylor, Biochem. Pharmacol. 58:749-57 (1999)). The assay system may be made quantitative by comparing the expression or function of a test sample with the levels of expression or function in a standard sample. Systems in which transcription factors are used to stimulate a positive output, such as transcription of a reporter gene, are generally referred to as "one-hybrid systems" (Wang, M.M. and Reed, R.R. (1993) Nature **364**:121-126). Using a transcription factor to stimulate a negative output (growth inhibition) may thus be referred to as a "reverse one-hybrid system" (Vidal et al, 1996, *supra*). Therefore, in an embodiment of the present invention, a reporter gene is placed under the control of a promoter of one of the 26 Markers identified herein.

In a further aspect of the invention we provide a cell or cell line comprising a reporter gene under the control of a Marker promoter, which Marker is one selected from the group consisting of: Marker 1 - Marker 26.

We also provide a method for identifying inhibitors of transcription of one or more of the Markers selected from the group consisting of: Marker 1 - Marker 26 which method comprises contacting a potential therapeutic agent with a cell or cell line as described above and determining inhibition of Marker transcription by the potential therapeutic agent by reference to a lack or reduced expression of the reporter gene.

A further aspect of the invention is directed to a modulator (activator or inhibitor) compound identified using any of the screening methods described above.

Any convenient test compound or library of test compounds may be used in conjunction with the test assay. Particular test compounds include low molecular weight chemical compounds (preferably with a molecular weight less than 1500 daltons) suitable as pharmaceutical or veterinary agents for human or animal use, or compounds for non-administered use such as cleaning/sterilising agents or for agricultural use.

Further features of the invention include:

A method of treatment of a patient suffering from a prostate disorder, comprising administration to the patient of a compound capable of reducing the transcription or expression of any one of Markers 1 - 26.

A method of treatment of a patient suffering from a prostate disorder, comprising administration to the patient an antisense nucleic acid molecule targeted against the mRNA of any one of Markers 1 - 26.

Use of an antisense nucleic acid molecule or an antibody directed against any one of Markers 1 - 26, in the manufacture of a medicament for treating a prostate disorder.

Each aspect of the invention involves detection or use of one or more of Markers 1 - 26. A preferred sub-group of Markers are those selected from the group consisting of Markers: 8, 13, 14, 15, 17, 18, 19, 20, 22, 24, 25 and 26.

The invention will be further described by way of the following non-limiting examples and figures in which data illustrating over-expression of markers is included:

Figure 1. Filter microarray data showing over-expression of Markers 1- 18 and 20-25 in prostate cancer relative to normal prostate. Overexpression value for Marker 26 (not shown) was found to be at the same level as Marker 13. Values given are mean expression level for 9 prostate cancer samples in those samples where over-expression of 1.5-fold or more was detected. Expression is given relative to normal prostate level (i.e. compared to the mean of 3 normal prostate datasets. These 3 datasets comprised 2 different RNA mixtures, each with at least 10 normal sample components).

Figure 2. Filter microarray data showing over-expression of Markers 1-5, 7-12, 17, 18, 20 and 23 in prostate cancer relative to BPH. Values given are mean expression level for 9 prostate cancer samples in those samples where over-expression of 1.5-fold or more was detected. Expression is given relative to BPH level (i.e. compared to the mean of 13 BPH datasets).

Example 1: Identification and Evaluation of Markers of Prostate Disease by cDNA Microarray Analysis

The microarray analysis protocol described in the following example was developed as a means to determine the relative abundances of mRNA species that are expressed in various tissues. Microarray analysis was used to identify differentially expressed RNA species

isolated from diseased primary human prostate tissue (benign and malignant), and normal prostate tissue. This involved comparison of quantitative data from dozens of RNA hybridisation experiments. 1877 probes each capable of targeting a distinct human gene, were used in duplicate in the microarray analysis, gridded out in a series of 7x7 arrays on a nylon filter support measuring approximately 70mm x 110mg. Where available, 2 different probes were used per gene. Housekeeping genes were included to serve as controls for hybridisation and data normalisation. The total number of spots in the array, and hence datapoints in the subsequent analysis was 4704. This figure includes hundreds of blanks for measuring background.

Each probe sequence (cDNA) was initially present in a plasmid clone in one of the following vectors; pINCY (Incyte Pharmaceuticals, CA, USA), pSPORT or pBS (Gibco BRL, UK).

For cDNA microarray generation, bacterial clones housing the plasmid containing segments of genes were independently cultured in 100µl of bacterial medium (containing 50mg/ml ampicillin) in a well of a 96 well microtitre plate and incubated at 37°C for 16 hours. Untreated culture was used to provide template for the polymerase chain reaction (PCR). 1µL of culture was added to a PCR mixture containing 2mM of the following oligonucleotide primers which bind to plasmid sequence and enable PCR amplification of the cloned insert: TTGGGTAACGCCAGGGTTTCCCAGTCAC and CCCCAGGCTTTACACTTTATGCTTCCGGC), dNTPs, buffer and thermostable DNA polymerase (ABL Ltd, UK), in 96-well format in duplicate. PCR amplification was carried out with 35 cycles of (15s at 94°C / 30s at 60°C / 4 min at 72°C) followed by 10min at 72°C. Duplicate reactions were mixed, and amplified PCR products (approximately 10nl) were spotted onto a set of nylon membranes (Hybond N, Amersham, UK) using a Q-BOT robotic system (Genetix Ltd. UK). Membranes were allowed to dry, then denatured and neutralised by laying on filter paper soaked as follows: 0.5M NaOH (5 min), 1.5M NaCl/1M Tris-HCl pH7.0 (5 min), 0.5M Tris pH8.0 (5 min), 2x SSC (standard sodium citrate, 5 min). The filter was then used in hybridisation experiments using RNA from normal or diseased tissue samples.

Normal prostate RNA was purchased from a commercial supplier (Clontech, Palo Alto CA). RNA from diseased prostatic tissue was isolated using a commercial reagent (Trizol, Gibco BRL, UK) according to the manufacturer's recommended procedure. Samples were

treated with DNAse I (also Gibco BRL), again as recommended by the manufacturer. For 1st strand label synthesis, 20μg total RNA was reverse transcribed into cDNA in the presence of radiolabelled nucleotide (^{33}P), using Superscript II enzyme (Gibco BRL) according to manufacturer's instructions. Labelled RNA was purified using GFX columns (Amersham Pharmacia Biotech, St. Albans, UK) according to manufacturers instructions, and added to 10 ml/filter of Church hybridisation buffer described below.

Prior to hybridisation, filters were wetted briefly in 2xSSC, then incubated in 5ml prewarmed Church hybridisation solution (0.5M sodium phosphate, 7% SDS, 1mM EDTA, pH 7.2) at 65°C for 2-6 hours. Probe was denatured at 100°C for 5 min, placed on ice for at least 5 min, then added to 5ml fresh Church hybridisation buffer in the presence of the filter, and mixed by swirling the bottle with the lid on for even probe distribution. Hybridisation was carried out at 65°C for 12-16 hours. Filters were rinsed briefly in pre-warmed Church wash solution (40mM sodium phosphate, 1% SDS, pH 7.2), then incubated twice in the same solution at 65°C for 20 min. Filters were drained briefly, wrapped in Saran wrap (Dow Chemical Company, USA) and exposed to phosphor screens for 3-6 days (screens and cassettes supplied by Molecular Dynamics, CA). The screens were then scanned using a Storm 830 phosphorimager (Molecular Dynamics). Array Vision software (Incyte Pharmaceuticals, CA, USA) was used to visualise the hybridisation images and generate quantitative numbers for each spot.

Typically, for comparison of data from different arrays, data is generated as a value relative to an internal standard for each array. Following export of ArrayVision data into Microsoft Excel format, spot measurements for data were normalised to housekeeping gene hybridisation signals known to be constant or relatively constant, using a simple Excel macro. A total of 38 spot values (from 16 different clones) representing 12 different housekeeping genes were used to generate an average housekeeping measurement. The gene and EMBL accession number of the housekeeping genes used are listed in Table 1. Background values were not subtracted, but to eliminate spurious low-level signals indistinguishable from noise, a local background value from 3 blank spots out of every 7x7 array was taken, and every value <2-fold this value was ignored.

To generate representative values for normal prostate, data from 3 array hybridisation experiments to normal prostate RNA was normalised to the mean housekeeping gene value, then the mean of these 3 values was taken as the control dataset.

Individual diseased prostate samples were compared to the mean normal value, and over-expression was defined as either a) a 1.5-fold increase above the mean normal value or b) presence of a value in the diseased sample versus no value for normal sample due to the 2-fold background cut-off imposed (a ratio could not be generated for these data).

On the basis of microarray analysis, hundreds of genes appeared to be over-expressed in diseased tissue compared to normal tissue. Over-expression levels ranged from several - fold to 50-fold. As can be seen from **Figure 1**, Markers 1-18 and Markers 20-25 were over-expressed in the range 1.5-fold to 7.5-fold in prostate cancer compared to normal prostate. Marker 19 was expressed in a high proportion of tumours tested, but never detected in normal prostate. Marker 19 could therefore represent a highly specific Marker.

In addition, Markers over-expressing in cancer relative to BPH were identified. This analysis was done by taking the mean normalised value for 13 BPH samples, and comparing each cancer sample to it. Over-expression in cancer was defined as a minimum of 1.5-fold increase over the expression in BPH, or, presence of a signal in cancer versus absence in the control (much rarer in this case due to the 13 samples of BPH rather than 3 normal samples above, and the observation that most genes over-expressed in BPH were also over-expressed in cancer). As can be seen from **Figure 2**, the expression of Markers 1-5, 7-12, 17, 18, 20 and 23 was increased by 1.5 to 4- fold in prostate cancer compared to BPH. Markers 6, 13, 14 and 25 were not detected in BPH, but were expressed in a high proportion of tumours tested, making them highly specific markers for prostate cancer.

One example of a diagnostic test could be the use of the gene probes characterised in Example 1 in a small transcript profiling array. This could comprise for example 2-50 probes or more together with blanks for background quantitation and control probes for data normalisation. Measurement of certain mRNA levels for groups of genes over-expressed in cancer versus BPH could allow molecular classification of the disease, which may improve on current classification and staging methods. One application of this could be differential diagnosis, e.g. testing of TURP chips for cancerous cells. At present TURP operations are carried out on patients with urinary blockage due to prostate enlargement. Most such patients have benign prostatic hyperplasia (BPH), but a small proportion (~10%) have cancer. Despite this, up to 50% of prostate cancer primary diagnosis may be detected via this route. Histopathological analysis is routinely carried out to detect the cancerous individuals. A diagnostic microarray test could eliminate or reduce time-consuming and costly

histopathological analysis. **Table 2** lists the 26 genes we have identified as being Markers for prostatic disease.

Control number	Gene	NCBI genInfo identifier	Spots on Array
1	Human ribosomal protein S14 gene	g337498	2
2	Human ribosomal protein S14 gene	g337498	2
3	Human acidic ribosomal phosphoprotein P0	g190231	2
4	Human alpha-tubulin mRNA	g340020	2
5	Human mRNA for TEF-5 protein	g1848081	2
6	Human transcriptional enhancer factor	g339440	2
7	Human transcription factor RTEF-1	g1561727	2
8	Human hypoxanthine phosphoribosyltransferase	g184349	2
9	Human hypoxanthine phosphoribosyltransferase	g184349	2
10	Human mRNA for transcription factor, TEF3	g1403337	2
11	Human acidic ribosomal phosphoprotein P1	g190233	2
12	Human mRNA for ribosomal protein L19	g36127	4
13	Human hH3.3B gene for histone H3.3	g761715	4
14	Human pancreatic phospholipase A-2 (PLA-2)	g190008	2
15	Human mRNA for ribosomal protein L19	g36127	4
16	Human acidic ribosomal phosphoprotein P1	g190233	2

Table 1. Names and identities of control housekeeping genes used to normalise microarray data. This set was identified from a wider set, by comparison of variability of candidate housekeeping gene signals over 13 filter hybridisation experiments as a group showing minimal variability.

Marker	Gene	EMBL ID
1	smooth muscle protein, 22kDa	HS22SM
2	Rho GDP dissociation inhibitor protein	HSRHO1
3	ferritin H	HSFERHC2
4	cDNA DKFZp564I1922	HSM800288
5	smooth muscle cell calponin	HS3701910
6	mRNA for KIAA0211 gene	HSD966
7	mRNA for KIAA0120 gene	HSORFF
8	ribosomal protein S25 mRNA	HSRPS25
9	80K-H protein (kinase C substrate) mRNA	HSG19P1A
10	alpha-2-macroglobulin; protease inhibitor	HSA2M
11	SNC73	AF067420
12	myosin light chain	HSMLC2A
13	DKFZp586DO918	HSM800155
14	laminin B2	HSLB2A26
15	PRSM1	HSU58048
16	ribosomal large subunit L12	HSL12A
17	7SK45	HS7K45
18	human protocadherin gamma (KIAA0588)	AB011160
19	factor V	HSFAV
20	SAP-1	HSSERREFB
21	Tip 60	HSU74667
22	SRP14	HSSRP14A
23	bone small proteoglycan I	HSHPGI
24	KIAA0045	HSKG1C
25	b-cell receptor associated protein	AF150962
26	cystatin B	HSCST4BA

Table 2. Sequences identified as markers of prostate disease; Markers 1-26. Expression of these Markers is increased >1.5-fold in prostate cancer compared to normal, or >1.5-fold in

prostate cancer compared to BPH, or in the case of Marker 19, not detected at all in normal prostate. In addition, Markers 6, 13, 14 and 25 were not detected in BPH.

Example 2: Quantitative RT-PCR test

RT-PCR is a convenient method for assaying the relative abundance of mRNA species expressed in various cells, tissues and organs.

Reproducibly accurate amplification of an mRNA can be achieved by firstly synthesising a DNA template from the mRNA using a reverse transcriptase, then by PCR using a heat-stable DNA polymerase (e.g. Taq polymerase). An optimised amplification protocol is essential for quantitation of nucleic acids by PCR because small differences in efficiency of the reaction can greatly influence the reaction rates, with a subsequent effect on PCR yield. Optimisation of PCR is not usually problematic (refer to: Koehler *supra,* and references therein). Quantitation of RT-PCR products can be done while the reaction products are building up exponentially, preferably following each round of amplification (known as real-time PCR). For quantitation, analysis is carried out by reference to one or more housekeeping genes which are also amplified by RT-PCR, often in a different reaction tube. Quantitation of RT-PCR product may be undertaken, for example, by gel electrophoresis visual inspection or image analysis, HPLC (Koehler et al., *supra*) or by use of suitable detection methods such as described above.

In practise, specific pairs of oligonucleotide primers corresponding to Markers 1-26 can be used to trigger amplification of specific marker mRNAs. The Marker mRNA is quantitated relative to a standard housekeeping mRNA using a real-time RT-PCR assay. For the purpose of this example, the housekeeping gene is ribosomal protein L19, but any such gene could be used.

Firstly, RNA is isolated from a clinical or tissue sample as in example 1. Aliquots of RNA are used in a reverse transcription reaction using random hexamers or oligo-dT as primers, and for example Moloney Murine Leukaemia Virus (MuLV) or Avian Myeloblastosis Virus (AMV) reverse transcriptase. A typical 20ml reverse transcription reaction comprises the following: 10 mM each dNTP, 50 mM Tris-HCl, pH8.3, 50 mM KCl, 50 mM dithiothrietol, 0.5 mM spermidine, 10ng oligo-dT primer, up to 1μg RNA, dissolved in DEPC-treated water. This mixture is incubated at 65°C for 15 min, then chilled on ice for 5 min. 5 units of AMV reverse transcriptase and 10 units of RNAse inhibitor are added. The

mixture is then incubated at 42°C for at least 1 hr. Secondly, a PCR is carried out using the first-strand cDNA as template. Primers specific to the sequence of interest, e.g. one of Markers 1-26, must be designed, typically 20-25 nucleotides long and 200-800 bp apart. This can be done with the aid of software such as Oligo (NBI, Cambio, UK). A typical 50 ml PCR contains: 10 mM Tris/HCl pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$, 0.01% gelatin, 0.2 mM dNTPs, 20-50 pmol primers, 50-100 ng cDNA and 1-2 units AmpliTaq Gold polymerase (PE Applied Biosystems, CA). In the same tube or a different tube, a reaction control is also initiated using paired primers specific for ribosomal protein L19, a ubiquitously expressed gene. Taqman probes internal to the amplicons can be used to detect specific amplification, and would be included in the reaction mix at suitable concentrations (refer to PE Applied Biosystem product information). A Taqman probe resides between the oligonucleotide primers and fluoresces only when hybridised to the denatured amplicon and partially digested by the 5'-3' exonucleotide activity of the Taq polymerase. Different reporter dyes can be used on the 2 Taqman probes such that the Marker fluorescence differs from the control fluorescence. The mixture is heated to 94°C for 15-20min, then cycled 35-40 times. Optimum cycling conditions would depend on the Tm of the primers, but typically might be 60°C for 1 min, 72°C for 3 min and 94°C for 1 min. Cycling is carried out on a thermal cycler capable of fluorescence detection (e.g. the ABI Prism 7700 Sequence Detection System). This detects emissions at the specific wavelengths of the probes following each cycle, and generates a cycle-by cycle amplification plot reflecting the increase in amplicon. The range of the exponential phase of the curve, or the point on the curve where fluorescence begins to significantly increase over background, is directly proportional to the number of starting copies of the template cDNA. By calibrating the cycle at which fluorescence increases against standard plots for known amounts of starting DNA, a quantitative measure of the level of Marker is obtained. By reference (e.g. ratio) to the signal for ribosomal protein L19, which acts as a control for successful amplification and is itself quantifiable, a measure of marker expression directly comparable to other samples, tumour or normal, is obtained. This may be further refined by multiple measurements of other housekeeping genes for accurate normalisation, other markers from the set Marker 1-26, or other markers as desired.

Example 3: Generation of antibody specific to Protein encoded by Markers 1-26

Antibody specific to Markers 1-26 could be used to detect protein products of these Markers in tissue samples and body fluids. Ideally for diagnosis, easily accessible samples are required, involving minimally invasive techniques. Typically, these samples include blood, plasma, serum, saliva, urine and semen. These fluids are far removed from prostate cells and, in a normal individual, would not be expected to contain mRNA of prostatic origin. However, protein may be transported across cell membrane barriers into body fluids. Prostate-specific antigen is one such protein, and is easily assayed by a simple blood test using specific antibody against the protein. For new markers, e.g. Markers 1-26, it may be necessary to generate novel antibody. This example describes a method of generating protein in vitro, which can then be used to raise antibody.

One way to generate protein is to express the cloned gene in *E. coli*. This requires that the cDNA, e.g. Markers 1-26 is cloned into a vector capable of expressing the encoded protein in the bacterial host. Optionally, it may be desirable to incorporate a molecular tag into the protein so that it can be easily purified. One such tag in wide use is the 6xHis tag. A protein with this tag is easily isolated from a cell culture extract by affinity chromatography, and can be used in relatively pure form to inoculate an animal for antibody generation.
Having subcloned the cDNA or cDNA fragment into an appropriate vector, e.g. pHAT10, 11, 12 or 20 (Clontech, Palo Alto, CA) and confirmed the integrity of the insert such as sequence and orientation. The subclone is cultured on a suitable scale, cells are harvested by centrifugation and then lysed, and the protein extract incubated with TALON resin according to manufacturer's instructions (Clontech, Palo Alto, CA). The resin is washed and recombinant protein is eluted by adjusting the pH or imidazole concentration. If desired, the 6xHIS tag can be enzymatically removed using a specific protease (Clontech). To generate a specific antibody, the purified protein is injected into a host animal, usually rabbit, sheep or goat. After boost injections, the serum is periodically collected and tested for antibody. Polyclonal antibody can be purified form the serum using standard techniques, or the animal's spleen can be harvested for the production of hybridomas. Techniques for the production of antibodies, both polyclonal and monoclonal, are well known to those skilled in the art (Catty D (ed.) Antibodies: A Practical Approach. Vol 1 (1988), Vol2 (1989)).

Example 4: Use of an antibody-based diagnostic test using the antibody of Example 3.

Antibody to one of the Marker 1-26 proteins can be used to specifically detect protein. Antibody technology is well advanced, and assay techniques that can be used to determine protein levels, such as Markers 1-26 of the present invention, in a sample derived from a host are well known to those skilled in the art. Such assays include radio-immune assays, immunohistochemistry assays, Western blot analysis and ELISA assays. Among these, ELISAs are favoured for detecting an expressed protein in biological fluids, and immunohistochemistry assays are favoured to detect a gene's expression in tissue samples. It will be apparent to the person skilled in the art that the protocol described below is only one example of an antibody-based test suitable to identify Marker protein in a test sample.

An antibody preparation, described in Example 3, is used to immunohistochemically stain a variety of normal and diseased tissues using standard procedures (Beesley, Immunocytochemistry: A Practical Approach. IRL Press, Oxford UK (1993)). Briefly, frozen blocks of tissue are cut into 6 micron sections and placed on microscope slides. After fixation in cold acetone, the sections are dried at room temperature, then washed with phosphate-buffered saline and blocked. The slides are incubated with the antibody at a dilution of 1:500, washed, incubated with biotinylated goat anti-rabbit antibody (if the primary antibody was raised in rabbit), washed again, and incubated with avidin labelled with horseradish peroxidase. After a final wash, the slides are incubated with 3-amino-9-ethylcarbazole which gives a red stain. The slides are counterstained with hematoxylin, mounted and examined under a microscope.

Example 5: Hybridisation Test

Hybridisation technology is well advanced. It will be apparent to the person skilled in the art that the protocol described below is only one example of a hybridisation protocol suitable to identify Marker nucleic acid in a test sample.

Hybridisation probe.

Hybridisation probes may be generated from any fragment of DNA or RNA encoding the specific Marker nucleic sequence of interest. Such fragments can be for example, restriction fragments isolated following restriction enzyme digestion of nucleic acid containing the Marker nucleotide sequence or synthetic oligonucleotides specific for a region of the Marker gene or a complementary sequence thereto.

A hybridisation probe can be generated from a synthetic oligonucleotide or a dephosphorylated restriction fragment sequence by addition of a radioactive 5' phosphate group from [γ-^{32}P]ATP by the action of T4 polynucleotide kinase. 20 pmoles of the oligonucleotide are added to a 20µl reaction containing 100mM Tris, pH7.5 , 10mM $MgCl_2$, 0.1mM spermidine, 20mM dithiothreitol (DTT), 7.55µM ATP, 55µCi [γ-^{32}P]ATP and 2.5u T4 polynucleotide kinase (Pharmacia Biotechnology Ltd, Uppsala, Sweden). The reaction is incubated for 30 minutes at 37°C and then for 10 minutes at 70°C prior to use in hybridisation. Methods for the generation of hybridisation probes from oligonucleotides or from DNA and RNA fragments are described in Chapters 11 and 10 respectively in Sambrook et al. (*ibid*). A number of proprietary kits are also available for these procedures.

Filter preparation

The sample DNA could be isolated and run on an agarose gel and Southern blotted onto a nitrocellulose or nylon filter using standard techniques.

Hybridisation conditions

Filters containing the nucleic acid are pre-hybridised in 100ml of a solution containing 6x SSC, 0.1%SDS and 0.25% dried skimmed milk (Marvel™) at 65°C for a minimum of 1 hour in a suitable enclosed vessel. A proprietary hybridisation apparatus such as model HB-1 (Techne Ltd) provides reproducible conditions for the experiment.

The pre-hybridisation solution is then replaced by 10ml of a probe solution containing 6xSSC, 0.1% SDS, 0.25% dried skimmed milk (e.g. Marvel™) and the oligonucleotide probe generated above. The filters are incubated in this solution for 5 minutes at 65°C before allowing the temperature to fall gradually to below 30°C. The probe solution is then discarded and the filters washed in 100ml 6xSSC, 0.1% SDS at room temperature for 5 minutes. Further washes (1-3) are then made in fresh batches of the same solution at 30°C and then, optionally, in 10°C increments up to 60°C for 5 minutes per wash.

After washing, the filters are dried and used to expose an X-ray film such as Hyperfilm™ MP (Amersham International) at -70°C in a light-tight film cassette using a fast tungstate intensifying screen to enhance the photographic image. The film is exposed for a suitable period (normally overnight) before developing to reveal the photographic image of the radio-active areas on the filters. Related nucleic acid sequences are identified by the presence of a photographic image compared to totally unrelated sequences which should not

produce an image. Generally, related sequences will appear positive at the highest wash temperature (60°C). However, related sequences may only show positive at the lower wash temperatures (50, 40 or 30°C).

These results will also depend upon the nature of the probe used. Longer nucleic acid fragment probes will need to be hybridised for longer periods at high temperature but may remain bound to related sequences at higher wash temperatures and/or at lower salt concentrations. Shorter, mixed or degenerate oligonucleotide probes may require less stringent washing conditions such as lower temperatures and/or higher Na^+ concentrations. A discussion of the considerations for hybridisation protocols is provided in Sambrook et al. (Chapter 11).

To prepare 20 x SSC,175.3 g of NaCl and 88.2 g of sodium citrate is dissolved in approximately 800ml of water, the pH is adjusted to 7.0 using 10 N solution of NaOH and the volume is adjusted to 1 litre with water, before autoclaving.

Example 6: *In situ* Hybridisation Test

In *in situ* hybridisation, sections of tissue may be interrogated with specific Marker probe to determine which morphological cell type in the sample displays a Marker nucleic acid, such as mRNA sequences corresponding to Markers 1-26. Standard treatments and protocols are well known to those in the art (Terenghi, Methods in Molecular Biology, 86:137-142 (1998); Polack and McGee, *In situ* hybridisation: principles and practice. Oxford University Press (1998)). Briefly, tissue is fixed for 16 hours in 4% paraformaldehyde solution, and 6 micron sections are prepared and placed on microscope slides, and allowed to dry at room temperature. Slides are rehydrated in 0.2% Triton X-100/phosphate buffered saline for 15 min and washed twice in the same solution. Slides are incubated in a prewarmed proteinase K solution (1mg/mL in 0.1M Tris-HCL pH 8.0, 50mM EDTA) for 15-20 min, and immersed in 0.1M glycine for 5 min to stop digestion. Slides are immersed in 4% paraformaldehde for 3 min, and rinsed briefly in phosphate buffered saline (PBS), twice. Slides are placed in 0.1M triethanolamine, acetic anhydride is added to 0.25% with stirring, and incubation is continued for 10 min. Slides are rinsed in double-distilled water and allowed to dry at 37°C for 10 min. 10 mL of probe (10-20 ng) in hybridisation buffer (50% deionised formamide, 5x standard sodium citrate (SSC), 10% dextran sulphate, 5x Denhardt's solution, 2% SDS, 100 mg/mL herring sperm DNA), denatured by heating to 95°C for 5 min

then chilled quickly on ice-water, is placed on the tissue section and a coverslip is sealed on to the slide using rubber solution. A suitable probe is a digoxygenin-labelled 300-600 base pair cDNA clone or fragment thereof, corresponding to one of Markers 1-26, prepared by one of several methods known to the art. The slides are then incubated in a sealed chamber for 16h at 37°C. The rubber solution is peeled away and coverslips are removed by immersion of the slides in 2x SSC, 0.1% SDS. Slides are washed 4 times for 5 min each at room temperature, and twice for 10 min at 65°C, in 2x SSC, 0.1% SDS, and rinsed briefly in 2x SSC. Hybridisation is detected as follows: Nonspecific binding is blocked by incubating the slides in blocking buffer (0.1M Tris-HCL pH 7.5, 0.1M NaCl, 2mM Mg Cl_2, 3% bovine serum albumin) for 10 min. Slides are then flooded with anti-digoxygenin-conjugated alkaline phosphatase (1/500 dilution of stock; Boehringer, Mannheim, Germany) in blocking buffer, and incubated 2 h at room temperature, then washed in blocking buffer 3 times for 3 min each. Slides are placed in buffer 2 (0.1 M Tris-HCl pH 9.5, 0.1 M NaCl, 50 mM $MgCl_2$) for 10 min, then immersed in substrate buffer (100mL buffer 2 containing 25 mg levamisole, plus 35 mg nitroblue tetrazolium chloride dissolved in 277 mL 70% dimethylformamide, plus 17 mg 5-bromo-4-indoyl-phosphate dissolved in 222 mL 100% dimethylformamide) for 10-30 min in the dark. The slides are then immersed in 20 mM Tris-HCl pH7.5, 5 mM EDTA for 5 min, rinsed with tap water for 5-10 min, mounted in aqueous mountant and examined under a microscope.

Example 7. Screening for therapeutic agents capable of inhibiting expression of any of Markers 1-26.

Markers 1-26 may be employed in a process for screening compounds which either inhibit, promote or modulate the expression of Markers 1-26. Examples of potential Marker 1-26 agonists are small molecules such as organic molecules or peptides, antibodies or oligonucleotides which bind to Markers 1-26 and inhibit expression or activity. One assay for therapeutic agents uses cultured cells, and measures transcript abundance using a microarray as described in Example 1. Typical prostate cell lines such as LNCaP, PC-3 and DU145 are available from the European Collection of Cell Cultures, Salisbury UK. Cell culture is a standard technique well known to those practised in the art. Briefly, in one suitable example, a seed vial of LNCaP cells is inoculated into 500mL culture flasks containing 20-50ml RPMI medium containing 10% fetal calf serum and 1% glutamate (all components Gibco, UK).

This is placed in a CO_2 incubator at 37°C for several days, and is used to seed fresh cultures to obtain several parallel cultures. Test compound is added, at a suitable concentration, to a set of flasks, which are serially harvested over the span of a time-course, for example 0, 1, 2, 4, 8 and 12 hours. Duplicates for each time point can also be used. Control, untreated samples must also be taken. Culture medium is poured from the culture flasks, and cells are washed with phosphate-buffered saline (PBS) containing 0.25% EDTA. Cells are released from the culture flask by addition of 0.25% crystalline trypsin (chymotripsin-free, Sigma Chemical Co., Poole, UK) in PBS, in sufficient quantity to cover the bottom of the flask, and incubation continued at 37°C for 2-8 min. Trypsin digestion is monitored by knocking the flask sharply to loosen the cells and observing them under an inverted microscope. As soon as cells are detached and single, digestion is halted by addition of 50ml RPMI medium containing 10% fetal calf serum and 1% soybean trypsin inhibitor (Sigma Chemical Co., Poole, UK). Suspended cells are transferred to centrifuge bottles and harvested in a standard centrifuge at 5k rpm for 10 min, and resuspended in Trizol for RNA extraction as in Example 1.

Microarray analysis can be used to assay levels of Marker 1-26 expression as in Example 1, or RT-PCR as in Example 2.

Claims:

1. A method for diagnosing or prognosing or monitoring a prostate disorder comprising testing a biological sample from an individual for aberrant levels of one or more of the Markers selected from the group consisting of: Marker 1 to Marker 26.

2. A method as claimed in claim 1, wherein the Marker(s) are selected form the group consisting of Markers: 8, 13, 14, 15, 17, 18, 19, 20, 22, 24, 25 and 26.

3. A method for distinguishing prostate cancer from BPH comprising testing a biological sample for aberrant levels of one or more of the Markers selected from the group consisting of Marker 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 17, 18, 20, 23, 25 and 26.

4. A method as claimed in any of claims 1-3 wherein mRNA transcript levels and/or protein levels of the Marker(s) is/are measured.

5. A method as claimed in claim 4, wherein mRNA transcript levels are measured using reverse-transcriptase polymerase chain reaction (RT-PCR).

6. A method for measuring Marker mRNA transcript levels as claimed in claim 4, which method involves the use of one or more oligonucleotide probes each capable of selectively hybridising to nucleic acid of a Marker of interest to determine the expression level of said Marker of interest.

7. A method for measuring Marker protein levels as claimed in claim 4, which method involves the use of one or more antibodies each capable of selectively binding to a Marker protein or protein fragment of interest to determine the expression level of said Marker of interest.

8. Use of an antibody selective for a Marker protein selected from the group consisting of: Marker 1 - Marker 26, in an assay to monitor therapeutic efficacy.

9. A diagnostic kit for diagnosing or prognosing or monitoring a prostate disorder comprising, one or more diagnostic probe(s) and/or one or more diagnostic primer(s) and/or one or more antibodies capable of selectively hybridising or binding to one or more of the Markers 1 -26.

10. A screening assay for identifying potential anti-prostate disorder therapeutic compounds comprising contacting an assay system capable of detecting the effect of a test compound on expression levels of one or more of the Markers selected from the group consisting of: Marker 1 - Marker 26, with a test compound and assessing the change in expression level of the particular Marker under study.

11. A method of testing potential therapeutic agents for the ability to suppress a prostate disorder phenotype comprising contacting a test compound with a cell engineered to express one of the Markers identified herein; and determining whether said test compound suppressed expression of the Marker.

12. A compound or agent identified by the screening assay according to claim 10 or the method according to claim 11.

13. A method of treatment of a patient suffering from a prostate disorder, comprising administration to the patient of a compound capable of reducing the transcription or expression of any one of Markers 1 - 26.

Figure 1

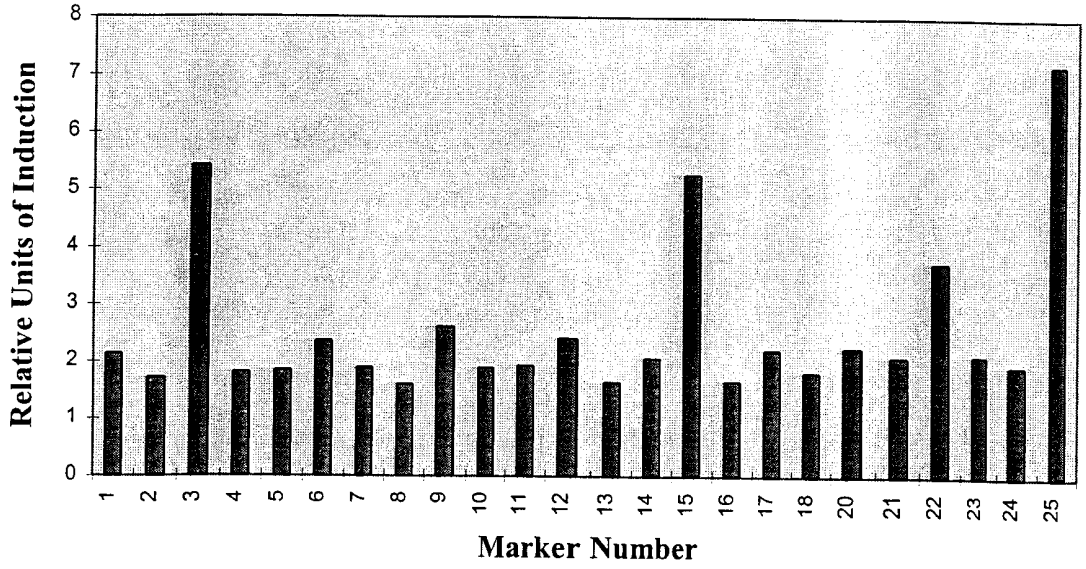
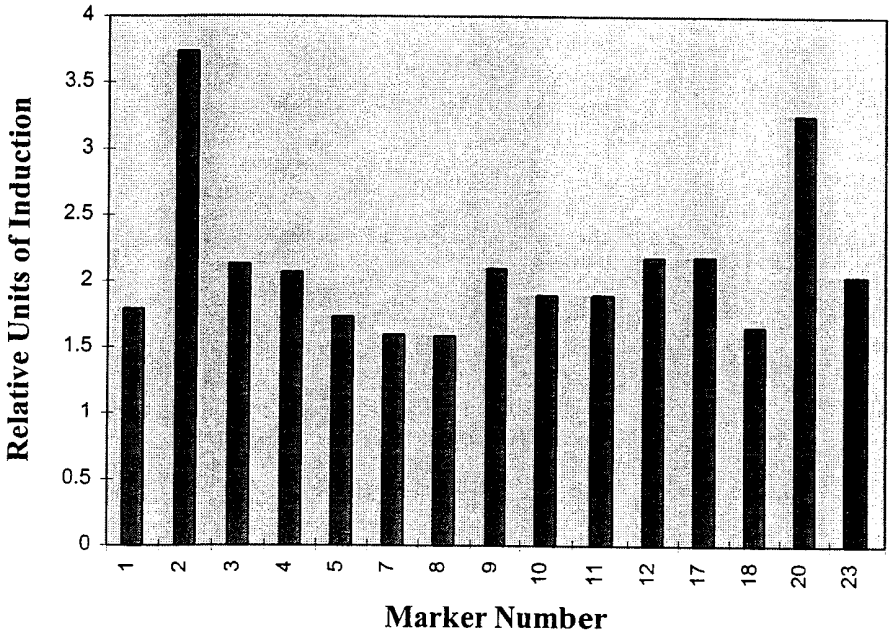


Figure 2



SEQUENCE LISTING

<110> AstraZeneca AB

<120> Diagnostic Methods.

<130> LDSG/Z70615/WO

<140>
<141>

<150> 9926805.4
<151> 1999-11-13

<160> 26

<170> PatentIn Ver. 2.1

<210> 1
<211> 1083
<212> DNA
<213> Homo sapiens

<400> 1
atcctgtctg tccgaaccca gacacaagtc ttcactcctt cctgcgagcc ctgaggaagc 60
cttctttccc cagacatggc caacaagggt ccttcctatg gcatgagccg cgaagtgcag 120
tccaaaatcg agaagaagta tgacgaggag ctggaggagc ggctggtgga gtggatcata 180
gtgcagtgtg gccctgatgt gggccgccca gaccgtgggc ccttgggctt ccaggtgtgg 240
ctgaagaatg gcgtgattct gagcaagctg gtgaacagcc tgtaccctga tggctccaag 300
ccggtgaagg tgcccgagaa cccaccctcc atggtcttca agcagatgga gcaggtggct 360
cagttcctga aggcggctga ggactctggg gtcatcaaga ctgacatgtt ccagactgtt 420
gacctctttg aaggcaaaga catggcagca gtgcagagga ccctgatggc tttgggcagc 480
ttggcagtga ccaagaatga tgggcactac cgtggagatc ccaactggtt tatgaagaaa 540
gcgcaggagc ataagaggga attcacagag agccagctgc aggagggaaa gcatgtcatt 600
ggccttcaga tgggcagcaa cagaggggcc tcccaggccg gcatgacagg ctacggacga 660
cctcggcaga tcatcagtta gagcggagag ggctagccct gagcccggcg ctcccccagc 720
tccttggctg cagccatccc gcttagcctg cctcacccac acccgtgtgg taccttcagc 780
cctggccaag ctttgaggct ctgtcactga gcaatggtaa ctgcacctgg gcagctcctc 840
cctgtgcccc cagcctcagc ccaacttctt acccgaaagc atcactgcct tggcccctcc 900
ctcccggcgg cccccatcac ctctactgtc tcctccctgg gctaagcagg ggagaagcgg 960
gctgggggta gcctggatgt gggcgaagtc cactgtcctc cttggcggca aaagcccatt 1020
gaagaagaac cagcccagcc tgccccctat cttgtacctg gaatattttt ggggttggaa 1080
ctc 1083

<210> 2
<211> 1819
<212> DNA
<213> Homo sapiens

<400> 2
cctgaaccgc gcggccgaac cctccggtgt cccgacccag gctaagcttg agcatggctg 60
agcaggagcc cacagccgag cagctggccc agattgcagc ggagaacgag gaggatgagc 120
actcggtcaa ctacaagccc ccggcccaga agagcatcca ggagatccag gagctggaca 180
aggacgacga gagcctgcga aagtacaagg aggccctgct gggccgcgtg gccgtttccg 240
cagaccccaa cgtccccaac gtcgtggtga ctggcctgac cctggtgtgc agctcggccc 300
cgggcccccc ggagctggac ctgacgggcg acctggagag cttcaagaag cagtcgtttg 360
tgctgaagga gggtgtggag taccggataa aaatctcttt ccgggttaac cgagagatag 420
tgtccggcat gaagtacatc cagcatacgt acaggaaagg cgtcaagatt gacaagactg 480
actacatggt aggcagctat gggccccggg ccgaggagta cgagttcctg accccgtgg 540
aggaggcacc caagggtatg ctggcccggg gcagctacag catcaagtcc cgcttcacag 600
acgacgacaa gaccgaccac ctgtcctggg agtggaatct caccatcaag aaggactgga 660
aggactgagc ccagccagag gcgggcaggg cagagtgatg gacggaagac ggacaggcgg 720
atgtgtcccc cccagccccc cccctcccca taccaaggtg ctgagcaggc cctccgtgcc 780

cctccaccct ggtccgcctc cctggcctgg ctcaaccgag tgcctccgac ccccctcctc 840
agccctcccc cacccacagg cccagcctcc tcggtctcct gtctcgttgc tgcttctgcc 900
tgtgctgtgg gggagagagg ccgcagccag gcctctgctg ccctttctgt gccccccagg 960
ttctatctcc ccgtcacacc cgaggcctgg cttcaggagg gagcggagca gccattctcc 1020
aggccccgtg gttgcccctg gacgtgtgcg tctgctgctc cggggtggag ctggggtgtg 1080
ggatgcacgg cctcgtgggg gccgggccgt cctccagccc cgctgctccc tggccagccc 1140
ccttgtcgct gtcggtcccg tctaaccatg atgccttaac atgtggagtg taccgtgggg 1200
cctcactagc ctctactccc tgtgtctgca tgagcatgtg gcctccccgt cccttccccg 1260
gtggcgaacc cagtgaccca gggacacgtg gggtgtgctg gtgctgctcc ccagcccacc 1320
aatgcctggc cagcctgccc ccttccctgg acagggctgt ggagatggct ccggcggctt 1380
ggggaaagcg aaattgccaa cactcaagtc acctcagtac catccaggag gctgggtatt 1440
gtcctgcctc tgccttttct gtctcagcgg cagtgcccag agcccacacc cccccaagag 1500
ccctcgatgg acaggcctga cccaccccac ctggggccag ccaggagccc cgcctgggcc 1560
atcagtattt attgcctccg tccgtgccgt ccctgggcca ctggctggcg cctcttcccc 1620
cagcctctca gtgccaccac ccccggcagc cttccctgac ccagccagga caaacaaggg 1680
accaagtgca cacattgctg agagccgtct cctataggtc ccccgcccca tccccggtgt 1740
tggtgttgtg tctgccaggc tcaggcagag gcgcctgtcc ctgcttcttt tctgaccggg 1800
aaataaatgc ccctgaagg 1819

<210> 3
<211> 996
<212> DNA
<213> Homo sapiens

<400> 3
agcttctctg atccctagta taaacacttc agtgttcccc tttcagtctt actactttga 60
ccgcgatgat gtggctttga agaactttgc caaatacttt cttcaccaat ctcatgagga 120
gagggaacat gctgagaaac tgatgaagct gcagaaccaa cgaggtggcc gaatcttcct 180
tcaggatatc aaggtgaaca aaagatccta ggggtgtcat acttcatcat ctggcagtgt 240
tcgggtatca gaaatcactt aaactagcaa ttgcccttat aaagtgatga tacactgggc 300
ttttgccttt tgtgcttttt taggcttacc atctaaacta aattaggcaa atagtaatgt 360
ccctttttgcc aaaacgtggt ggttagagat gatgggcttg ctgacttcta ggttagttgg 420
tagagatgca ttaacctatt ctcattcaga aaccagactg tgatgactgg gagagcgggc 480
tgaatgcaat ggagtgtgca ttacatttgg aaaaaaatgt gaatcagtca ctactggaac 540
tgcacaaact ggccactgac aaaaatgacc cccatgtgag tattggaacc ccaggaaata 600
aatggaggaa atcatttgcc ttagggattg ggaaagctgc ccactaactg tcttgcccca 660
ttgttttgca gttgtgtgac ttcattgaga cacattacct gaatgagcag gtgaaagcca 720
tcaaagaatt gggtgaccac gtgaccaact tgcgcaagat gggagcgccc gaatctggct 780
tggcggaata tctctttgac aagcacaccc tgggagacag tgataatgaa agctaagcct 840
cgggctaatt tccccatagc cgtggggtga cttccctggt caccaaggca gtgcatgcat 900
gttggggttt cctttacctt ttctataagt tgtaccaaaa catccactta agttctttga 960
tttgtaccat tccttcaaat aaagaaattt ggtacc 996

<210> 4
<211> 2353
<212> DNA
<213> Homo sapiens

<400> 4
gggttgagca caaggaggag aacgaccaca aagtcttcta cgggggtgac ctgaaagtgg 60
actgtgtggc caccgggctt cccaatcccg agatctcctg gagcctccca gacgggagtc 120
tggtgaactc cttcatgcag tcggatgaca gcggtggacg caccaagcgc tatgtcgtct 180
tcaacaatgg gacactctac tttaacgaag tggggatgag ggaggaagga gactacacct 240
gctttgctga aaatcaggtc gggaaggacg agatgagagt cagagtcaag gtggtgacag 300
cgcccgccac catccggaac aagacttgct tggcggttca ggtgccctat ggagacgtgg 360
tcactgtagc ctgtgaggcc aaaggagaac ccatgcccaa ggtgacttgg ttgtccccaa 420
ccaacaaggt gatccccacc tcctctgaga agtatcagat ataccaagat ggcactctcc 480
ttattcagaa agcccagcgt tctgacagcg gcaactacac ctgcttggtc aggaacagcg 540
cgggagagga taggaagacg gtgtggattc acgtcaacgt ccagccaccc aagatcaacg 600
gtaaccccaa ccccatcacc actgtgcggg agatagcagc cgggggcagt cggaaactga 660
ttgactgcaa agctgaaggc atccccaccc cgagggtgtt atgggctttt cccgagggtg 720
tggttctgcc agctccatac tatggaaacc ggatcactgt ccatggcaac ggttccctgg 780

acatcaggag tttgaggaag agcgactccg tccagctggt atgcatggca cgcaacgagg 840
gaggggaggc caggttgatc ctgcagctca ctgtcctgga gcccatggag aaacccatct 900
tccacgaccc gatcagcgag aagatcacgg ccatggcggg ccacaccatc agcctcaact 960
gctctgccgc ggggacccca acacccagcc tggtgtgggt ccttcccaat ggcaccgatc 1020
tgcagagtgg acagcagctg cagcgcttct accacaaggc tgacggcatg ctacacatta 1080
gcggtctctc ctcggtggac gccggggcct accgctgcgt ggcccgcaat gccgctggcc 1140
acacggagag gctggtctcc ctgaaggtgg gactgaagcc agaagcaagc aagcagtatc 1200
ataacctggt cagcatcatc aatggtgaga ccctgaagct cccctgcacc cctcccgggg 1260
ctgggcaggg acgtttctcc tggacgctcc ccaatggcat gcatctggag ggcccccaaa 1320
ccctgggacg cgtttctctt ctggacaatg gcaccctcac ggttcgtgag gcctcggtgt 1380
ttgacagggg tacctatgta tgcaggatgg agacggagta cggcccttcg gtcaccagca 1440
tccccgtgat tgtgatcgcc tatcctcccc ggatcaccag cgagcccacc ccggtcatct 1500
acacccggcc cgggaacacc gtgaaactga actgcatggc tatggggatt cccaaagctg 1560
acatcacgtg ggagttaccg gataagtcgc atctgaaggc aggggttcag gctcgtctgt 1620
atggaaacag atttcttcac ccccagggat cactgaccat ccagcatgcc acacagagag 1680
atgccggctt ctacaagtgc atggcaaaaa acattctcgg cagtgactcc aaaacaactt 1740
acatccacgt cttctgaaat gtggattcca gaatgattgc ttaggaactg acaacaaagc 1800
ggggtttgta agggaagcca ggttggggaa taggagctct taaataatgt gtcacagtgc 1860
atggtggcct ctggtgggtt tcaagttgag gttgatcttg atctacaatt gttgggaaaa 1920
ggaagcaatg cagacacgag aaggagggct cagccttgct gagacacttt cttttgtgtt 1980
tacatcatgc caggggcttc attcagggtg tctgtgctct gactgcaatt tttcttttt 2040
tgcaaatgcc actcgactgc cttcataagc gtccatagga tatctgagga acattcatca 2100
aaaataagcc atagacatga acaacacctc actaccccat tgaagacgca tcacctagtt 2160
aacctgctgc agtttttaca tgatagactt tgttccagat tgacaagtca tctttcagtt 2220
atttcctctg tcacttcaaa actccagctt gcccaataag gatttagaac cagagtgact 2280
gatatatata tatattttaa ttcagagtta catacataca gctaccattt tatatgaaaa 2340
aaaaaaaaaa aaa 2353

<210> 5
<211> 1504
<212> DNA
<213> Homo sapiens

<400> 5
tgcagacgga acttcagccg ctgcctctgt tctcagcgtc agtgccgcca ctgcccccgc 60
cagagcccac cggccagcat gtcctctgct cacttcaacc gaggccctgc ctacgggctg 120
tcagccgagg ttaagaacaa gctggcccag aagtatgacc accagcggga gcaggagctg 180
agagagtgga tcgagggggt gacaggccgt cgcatcggca acaacttcat ggacggcctc 240
aaagatggca tcattctttg cgaattcatc aataagctgc agccaggctc cgtgaagaag 300
atcaatgagt caacccaaaa ttggcaccag ctggagaaca tcggcaactt catcaaggcc 360
atcaccaagt atggggtgaa gccccacgac atttttgagg ccaacgacct gtttgagaac 420
accaaccata cacaggtgca gtccaccctc ctggctttgg ccagcatggc gaagacgaaa 480
ggaaacaagg tgaacgtggg agtgaagtac gcagagaagc aggagcggaa attcgagccg 540
gggaagctaa gagaagggcg gaacatcatt gggcttcaga tgggcagcaa caagtttgcc 600
agccagcagg gcatgacggc ctatggcacc cggcgccacc tctacgaccc caagctgggc 660
acagaccagc ctctggacca ggcgaccatc agcctgcaga tgggcaccaa caaaggagcc 720
agccaggctg gcatgactgc gccagggacc aagcggcaga tcttcgagcc ggggctgggc 780
atggagcact gcgacacgct caatgtcagc ctgcagatgg gcagcaacaa gggcgcctcg 840
cagcggggca tgacggtgta tgggctgcca cgccaggtct acgaccccaa gtactgtctg 900
actcccgagt acccagagct gggtgagccc gcccacaacc accacgcaca caactactac 960
aattccgcct aggtccacaa ggccttcact gttttccccc caagggaggc tgctgctgct 1020
cttggctgga gccagccagg gccagccgac cccctctccc tgcatggcat cctccagccc 1080
ctgtagaact caacctctac agggttagag tttggagaga gcagactggc ggggggccca 1140
ttgggggaa ggggaccctc cgctctgtag tgctacaggg tccaacatag aacagggtgt 1200
ccccaacagc gcccaaagga cgcactgagc aacgctattc cagctgtccc cccactccct 1260
cacaagtggg taccccccagg accagaagct cccccagcaa agcccccaga gcccaggctc 1320
ggcctgcccc caccccattc ccgcagtggg agcaaactgc atgcccagag acccagcgga 1380
cacacgcggt ttggtttgca gcgactggca tactatgtgg atgtgacagt ggcgtttgta 1440
atgagagcac tttctttttt ttctatttca ctggagcaca ataaatggct gtaaaatcta 1500
cacg 1504

<210> 6

<211> 5086
<212> DNA
<213> Homo sapiens

<400> 6
gcctgcgccg cgcagcccgc ctcgggcggg aggcgggagg cgggaggccg ggcccaggcc 60
ggggcagccc ctccaccgca ccgtcctggg ccggtgccca ggtccgagtc gccttccgcc 120
tgcccccccc gccaatcccc cgccgcggca gccccagcca ggtcccgccg ccaggccggc 180
tcccgccccc gctccgcccc cggagccgca gccccgcccg ccatcgccgt cgccatgttg 240
tggctcccgc agccggcgct ggggacgcgc gcggccgaga ctctggcctg cagtcgccgc 300
cgccgccgcc aggtggtgtt tggactctag accatgtgcc taggtagaag tttttccttt 360
ctccgcagct ctgctccgct agcaacgctc gccacaccct tgttttgaga tcctctctaa 420
ggagcggaga gtttaatagg caagaaggaa gggagaagac agaaggaaga cgctcccccg 480
tacggagaca gagggagggg gggctccaaa gccgaaagag gaggtcccta cctgccacgg 540
ataccagtca gcccttgcca gcatccagcc atgggggata tgaaaacccc agattttgat 600
gaccttctgg ctgcctttga catcccagac cccaccagcc ttgatgccaa ggaggccatc 660
cagacaccca gtgaggagaa tgagagtccc ctcaaacctc caggcatatg tatggatgaa 720
agtgtgtcct tgtctcactc aggatcagcc cccgatgtgc cggccgtgag tgtcattgtc 780
aagaacacca gccgccagga gtcatttgaa gcggagaaag accacattac tcccagtctc 840
ctacacaatg gattccgggg ctcagatctg cctccagatc cccacaactg tgggaaattt 900
gattctactt ttatgaatgg agacagtgcc aggagtttcc ctggcaaact ggagcctccc 960
aagtcagagc cattacccac cttcaaccag ttcagtccaa tctccagccc agaacctgag 1020
gatcccatca aagataacgg atttgggata aagcccaaac actctgacag ttatttccca 1080
ccccctcttg ggtgcggggc tgtgggaggc ccagtcctgg aggctctggc taagtttccg 1140
gttccagagc tgcatatgtt tgatcatttt tgtaagaaag aacccaagcc agaacccctg 1200
cccttgggga gccagcagga acacgagcaa agtgggcaga acacagtgga acctcacaag 1260
gatccggatg ccactcgatt cttcggggaa gctttggagt tcaacagcca tcctagcaac 1320
agtattggag agtccaaggg gcttgcccgg gagcttggta cctgctcatc agtccccct 1380
aggcagcgtc taaagccagc tcattccaag ctgtcctctt gtgtggcagc cttggtggcc 1440
ttgcaggcca aaagagtggc tagtgtcact aaggaggatc agcctggcca cacaaaggat 1500
ctctcagggc ccactaaaga gagttctaaa ggtagcccca aaatgcccaa gtcaccaaag 1560
agtccccgga gccctctgga ggccactaga aaaagtatca agccatcgga cagccctcgt 1620
agcatctgca gtgacagcag cagcaaaggc tcaccgtctg tggctgccag ctccccacca 1680
gcaattccca aagtgagaat caaaaccatt aagacatcat caggggaaat caaacggact 1740
gtcacaagga tcctgccaga tcctgatgat ccaagtaagt cccctgttgg gtcacctcta 1800
gggagcgcca ttgcagaggc ccccagcgag atgccagggg atgaggtgcc tgtggaagag 1860
cactttcctg aggcaggcac aaattcaggg agcccccagg gggccaggaa aggggacgag 1920
agcatgacaa aggccagtga ctcgtcatct cccagctgca gttctgggcc ccgggtccca 1980
aaggggggctg ccccaggctc acagacaggc aagaagcaac agagcacagc actgcaggca 2040
tccaccctgg cccctgccaa cctcctgccc aaagccgtgc acttggccaa cctgaacctc 2100
gtcccccaca gtgttgctgc atcagtgaca gccaagtctt cagtgcaaag acggagccag 2160
ccacagctta cacaaatgtc ggtgcccctg gtccaccagg tgaaaaaggc tgccccactg 2220
attgtagagg tcttcaacaa ggtccttcac agctccaacc ccgtgccccct ctatgcgcca 2280
aatctcagcc cgcctgcgga cagcaggatc cacgtgccgg ccagtgggta ctgctgcctg 2340
gagtgtggag acgcatttgc cttagagaag agcctgagcc agcactatgg ccggcggagc 2400
gtccacattg aggtactgtg cacactgtgc tccaagacgc tgctcttctt caacaagtgc 2460
agcctgctcc ggcacgcccg tgaccacaag agcaaggggc tcgtcatgca gtgttcccag 2520
ctgctggtga agcctatctc tgcggaccaa atgttcgtgt cggcccctgt gaactccacg 2580
gcaccagcag ccccagcccc ttcatcctct cccaaacatg gcctcacttc gggcagtgcc 2640
agtccccctc ctccagcctt gccactctac ccagaccctg tgaggctcat ccggtactca 2700
atcaagtgtc ttgaatgtca caagcagatg cgggactaca tggtcctggc tgcacatttc 2760
cagaggacaa cagaggagac agaggggctg acctgccagg tatgccagat gctgctgccc 2820
aaccagtgca gtttctgtgc ccaccagcgg attcatgcac acaagtcccc ctactgctgc 2880
ccggagtgtg gggtcctctg ccgctctgcc tacttccaga cccatgtaaa ggagaattgc 2940
ctgcactatg cccgcaaggt gggctacagg tgcatccact gtggtgtcgt ccacctgacc 3000
ttggccttgc tgaaaagcca catccaggag cgacactgcc aggttttcca caaatgtgca 3060
ttctgcccca tggccttcaa gactgccagc agcactgcag accacagtgc cacccagcac 3120
cccacccagc cccacagacc ctcccagctc atttataagt gctcctgtga aatggtcttc 3180
aacaagaaga ggcacattca gcagcatttt taccagaatg tcagcaagac gcaggtgggc 3240
gtcttcaagt gccctgagtg cccactcttg ttcgtgcaga agccggagtt gatgcaacac 3300
gtcaagagca cccacggtgt tccccgaaat gtggacgagc tgtcaaacct ccagtcttca 3360
gcggacacat cctcaagccg ccctggctct cgagttccca ctgagccacc agccactagt 3420
gtggctgctc ggagcagctc cctgccttct ggccgctggg gtaggcctga agcccaccgc 3480
agggtggaag ccaggccgcg gctgaggaac actggctgga cctgccagga gtgccaggag 3540

tgggttccag atcgggagag ctacgtgtcc cacatgaaaa agagccacgg tcggacattg 3600
aagcggtacc catgccggca gtgtgaacag tccttccaca cccccaacag cctgcgcaaa 3660
cacatccgca acaaccatga cacagtaaag aagttctaca cctgcgggta ctgcacagag 3720
gacagcccca gctttcctcg gccctccctt ctggagagcc acatcagcct tatgcatggc 3780
atcagaaacc ctgatttgag ccagacgtcc aaagtgaaac ctccgggtgg acattccct 3840
caggtgaacc atctgaaaag accagtcagt ggagtggggg acgctccagg caccagcaat 3900
ggcgcaactg tctcttccac caaaaggcac aagtcccttt ttcagtgcgc gaaatgtagt 3960
tttgccacag actcggggct cgagtttcag agccacatac ctcagcacca ggtggacagc 4020
tccacagccc aatgtctcct ctgtggtttg tgctacacct ctgccagctc cctcagccgc 4080
cacctcttca ttgtccacaa ggtgagagac caggaggagg aggaggaaga ggaggcggcg 4140
gcagcggaga tggcagtgga ggtggcagag ccagaggagg gctccgggga ggaggtgccc 4200
atggagacta gagagaatgg actggaagaa tgtgccggtg agcctttgtc agctgaccca 4260
gaggcgagga gattgctggg cccggccct gaggacgatg gtggccacaa tgatcacagt 4320
caaccacagg cctctcagga ccaggacagc cacacactgt cccctcaggt gtgaccggag 4380
actttgcagt gtgcatggtc aggggtggtg ccgaagtgtc ttccacctgc cctgcggacc 4440
gtggaaaata aaaggctctg cccccagtgt gagtgtgacc ggttgtaccc tggagtagtg 4500
tctgccctga gctgccagtg ctgggtatcc cccagcccca ggaaatgtgg ggtcggccag 4560
gaccctcaca gctctgaatt tgcttctgtt atttatggct tttcgctgct tcttggtgcc 4620
ccatctcttg tctgtgtcct tccaacccca agctgcttat gtggcccaac cccactgctg 4680
tcaactaggc ttgaacccca cagcggctgt gctcttctgg gaggttcccg cttgctgcct 4740
tcagccaggg cgctcctcag agctctattt tcctgcagac accagctctc cttcctgcct 4800
ttagatcctg agaaggaggg aaatgagggg tgctgacaca gtccctctgg gagagctctg 4860
cctagtctgg tttggcgagg gcccttgatc accttgcccc tcctccctgt cttctctgat 4920
tcttttccct caaaatagtc ctgagaacta attgtcacag acattggaat atttgtactg 4980
ctctcgtgcc atttgagagg ctgctgcccc aggcaggcca gcccctactc ctcttggcta 5040
cactcatgtt gctcagacta tatttcaaat aaaaaatctt ctcacc 5086

<210> 7
<211> 1360
<212> DNA
<213> Homo sapiens

<400> 7
gcccttgcct tgagtcagtg cgctgctctc cagcccgctt gaacgctccc cgcagccacc 60
gccacccatt ggaatggcca acaggggacc tgcatatggc ctgagccggg aggtgcagca 120
gaagattgag aaacaatatg atgcagatct ggagcagatc ctgatccagt ggatcaccac 180
ccagtgccga aaggatgtgg gccggcccca gcctggacgc gagaacttcc agaactggct 240
caaggatggc acggtgctat gtgagctcat taatgcactg taccccgagg ggcaggcccc 300
agtaaagaag atccaggcct ccaccatggc cttcaagcag atggagcaga tctctcagtt 360
cctgcaagca gctgagcgct atggcattaa caccactgac atcttccaaa ctgtggacct 420
ctgggaagga aagaacatgg cctgtgtgca gcggacgctg atgaatctgg gtgggctggc 480
agtagcccga gatgatgggc tcttctctgg ggatcccaac tggttcccta agaaatccaa 540
ggagaatcct cggaacttct cagataacca gctgcaagag ggcaagaacg tgatcgggtt 600
acagatgggc accaaccgcg gggcgtctca ggcaggcatg actggctacg ggatgccacg 660
ccagatcctc tgatcccacc ccaggccttg cccctgccct cccacgaatg gttaatatat 720
atgtagatat atattttagc agtgacattc ccagagagcc ccagagctct caagctcctt 780
tctgtcaggg tgggggttc agcctgtcct gtcacctctg aggtgcctgc tggcatcctc 840
tccccatgc ttactaatac attcccttcc ccatagccat caaaactgga ccaactggcc 900
tcttcctttc ccctgggacc aaaatttagg ggcctcagtc cctcaccgcc atgccctggc 960
ctattctgtc tctccttctt ccccctggcc tgttctgtct ctgagctctg tgtcctccgt 1020
tcattccatg gctgggagtc actgatgctg cctctgcctt ctgatgctgg actggccttg 1080
cttctacaag tatgcttctc ccacagctgt ggctgcagga acttaattta tagggaggag 1140
cctgtggcag ctgctgcccc agccacagct gcactgactg tgctcaccac acatctgggg 1200
cagccttccc tggcaggggc cctcgtggct tctcattttc cattcccttc actgtggcta 1260
aggggtgggg tgaggggatg gagagggagg gctgcctacc atggtctggg gcttgaggaa 1320
gatgagtttg ttgatttaaa taaagaattt gtcatttttg 1360

<210> 8
<211> 497
<212> DNA
<213> Homo sapiens

<400> 8
tttttttttt ttttttgtcc gacatcttga gacgaggctg cggtgtctgc tgctattctc 60
cgagcttcgc aatgccgcct aaggacgaca agaagaagaa ggacgctgga aagtcggcca 120
agaaagacaa agacccagtg aacaaatccg ggggcaaggc caaaaagaag aagtggtcca 180
aaggcaaagt tcgggacaag ctcaataact tagtcttgtt tgacaaagct acctatgata 240
aactctgtaa ggaagttccc aactataaac ttataacccc agctgtggtc tctgagagac 300
tgaagattcg aggctccctg gccagggcag cccttcagga gctccttagt aaaggactta 360
tcaaactggt ttcaaagcac agagctcaag taatttacac cagaaatacc aagggtggag 420
atgctccagc tgctggtgaa gatgcatgaa taggtccaac cagctgtaca tttggaaaaa 480
taaaacttta ttaaatc 497

<210> 9
<211> 2056
<212> DNA
<213> Homo sapiens

<400> 9
ggaaccgcgg ctgctggaca agaggggtgc ggtggatact gacctttgct ccggcctcgt 60
cgtgaagaca cagcgcatct ccccgctgta ggcttctccc acagaacccg tttcgggcct 120
cagagcgtct ggtgagatgc tgttgccgct gctgctgctg ctacccatgt gctgggccgt 180
ggaggtcaag aggccccggg gcgtctccct caccaatcat cacttctacg atgagtccaa 240
gcctttcacc tgcctggacg gttcggccac catcccattt gatcaggtca acgatgacta 300
ttgcgactgc aaagatggct ctgacgagcc aggcacggct gcctgtccta atggcagctt 360
ccactgcacc aacactggct ataagccccc gtatatcccc tccaaccggg tcaacgatgg 420
tgtttgtgac tgctgcgatg gaacagacga gtacaacagc ggcgtcatct gtgagaacac 480
ctgcaaagag aagggccgta aggagagaga gtccctgcag cagatggccg aggtcacccg 540
cgaagggttc cgtctgaaga agatccttat tgaggactgg aagaaggcac gggaggagaa 600
gcagaaaaag ctcattgagc tacaggctgg gaagaagtct ctggaagacc aggtggagat 660
gctgcggaca gtgaaggagg aagctgagaa gccagagaga gaggccaaag agcagcacca 720
gaagctgtgg gaagagcagc tggctgctgc caaggcccaa caggagcagg agctggcggc 780
tgatgccttc aaggagctgg atgatgacat ggacgggacg gtctcggtga ctgagctgca 840
gactcacccg gagctggaca cagatgggga tggggcgttg tcagaagcgg aagctcaggc 900
cctcctcagt ggggacacac agacagacgc cacctctttc tacgaccgcg tctgggccgc 960
catcagggac aagtaccggt ccgaggcact gcccaccgac cttccagcac cttctgcccc 1020
tgacttgacg gagcccaagg aggagcagcc gccagtgccc tcgtcgccca cagaggagga 1080
ggaggaggag gaggaggagg aagaagaggc tgaagaagag gaggaggagg aggattccga 1140
ggaggcccca ccgccactgt caccccccgca gccggccagc cctgctgagg aagacaaaat 1200
gccgccctac gacgagcaga cgcaggcctt catcgatgct gcccaggagg cccgcaacaa 1260
gttcgaggag gccgagcggt cgctgaagga catggaggag tccatcagga acctggagca 1320
agagatttct tttgactttg gccccaacgg ggagtttgct tacctgtaca gccagtgcta 1380
cgagctcacc accaacgaat acgtctaccg cctctgcccc ttcaagcttg tctcgcagaa 1440
acccaaactc gggggctctc ccaccagcct tggcacctgg ggctcatgga ttggccccga 1500
ccacgacaag ttcagtgcca tgaagtatga gcaaggcacg ggctgctggc agggccccaa 1560
ccgctccacc accgtgcgcc tcctgtgcgg gaaagagacc atggtgacca gcaccacaga 1620
gcccagtcgc tgcgagtacc tcatggagct gatgacgcca gccgcctgcc cggagccacc 1680
gcctgaagca cccaccgaag acgaccatga cgagctctag ctggatgggc gcagagaacc 1740
tcaagaaggc atgaagccag cccctgcagt gccgtccacc cgcccctctg ggcctgcctg 1800
tggctctgtt gccctcctct gtggcggcag gacctttgtg gggcttcgtg ccctgctctg 1860
gggcccaggc ggggctggtc cacattccca ggccccaaca gcctccaaag atgggtaaag 1920
gagcttgccc tccctgggcc ccccaccttg gtgactcgcc ccaccacccc cagccctgtc 1980
cctgccaccc ctcctagtgg ggactagtga atgacttgac ctgtgacctc aatacaataa 2040
atgtgatccc ccaccc 2056

<210> 10
<211> 4577
<212> DNA
<213> Homo sapiens

<400> 10
gctacaatcc atctggtctc ctccagctcc ttctttctgc aacatgggga agaacaaact 60
ccttcatcca agtctggttc ttctcctctt ggtcctcctg cccacagacg cctcagtctc 120
tggaaaaccg cagtatatgg ttctggtccc ctccctgctc cacactgaga ccactgagaa 180

gggctgtgtc cttctgagct acctgaatga gacagtgact gtaagtgctt ccttggagtc 240
tgtcagggga aacaggagcc tcttcactga cctggaggcg gagaatgacg tactccactg 300
tgtcgccttc gctgtcccaa agtcttcatc caatgaggag gtaatgttcc tcactgtcca 360
agtgaaagga ccaacccaag aatttaagaa gcggaccaca gtgatggtta agaacgagga 420
cagtctggtc tttgtccaga cagacaaatc aatctacaaa ccagggcaga cagtgaaatt 480
tcgtgttgtc tccatggatg aaaactttca ccccctgaat gagttgattc cactagtata 540
cattcaggat cccaaaggaa atcgcatcgc acaatggcag agtttccagt tagagggtgg 600
cctcaagcaa ttttcttttc ccctctcatc agagcccttc cagggctcct acaaggtggt 660
ggtacagaag aaatcaggtg gaaggacaga gcacctttc accgtggagg aatttgttct 720
tcccaagttt gaagtacaag taacagtgcc aaagataatc accatcttgg aagaagagat 780
gaatgtatca gtgtgtggcc tatacacata tgggaagcct gtccctggac atgtgactgt 840
gagcatttgc agaaagtata gtgacgcttc cgactgccac ggtgaagatt cacaggcttt 900
ctgtgagaaa ttcagtggac agctaaacag ccatggctgc ttctatcagc aagtaaaaac 960
caaggtcttc cagctgaaga ggaaggagta tgaaatgaaa cttcacactg aggcccagat 1020
ccaagaagaa ggaacagtgg tggaattgac tggaaggcag tccagtgaaa tcacaagaac 1080
cataaccaaa ctctcatttg tgaaagtgga ctcacacttt cgacagggaa ttcccttctt 1140
tgggcaggtg cgcctagtag atgggaaagg cgtccctata ccaaataaag tcatattcat 1200
cagaggaaat gaagcaaact attactccaa tgctaccacg gatgagcatg gccttgtaca 1260
gttctctatc aacaccacca acgttatggg tacctctctt actgttaggg tcaattacaa 1320
ggatcgtagt ccctgttacg gctaccagtg ggtgtcagaa gaacacgaag aggcacatca 1380
cactgcttat cttgtgttct ccccaagcaa gagctttgtc caccttgagc ccatgtctca 1440
tgaactaccc tgtggccata ctcagacagt ccaggcacat tatattctga atggaggcac 1500
cctgctgggg ctgaagaagc tctcctttta ttatctgata atggcaaagg gaggcattgt 1560
ccgaactggg actcatggac tgcttgtgaa gcaggaagac atgaagggcc atttttccat 1620
ctcaatccct gtgaagtcag acattgctcc tgtcgctcgg ttgctcatct atgctgtttt 1680
acctaccggg gacgtgattg gggattctgc aaaatatgat gttgaaaatt gtctggccaa 1740
caaggtggat ttgagcttca gcccatcaca aagtctccca gcctcacacg cccacctgcg 1800
agtcacagcg gctcctcagt ccgtctgcgc cctccgtgct gtggaccaaa gcgtgctgct 1860
catgaagcct gatgctgagc tctcggcgtc ctcggtttac aacctgctac cagaaaagga 1920
cctcactggc ttccctgggc ctttgaatga ccaggacgat gaagactgca tcaatcgtca 1980
taatgtctat attaatggaa tcacatatac tccagtatca agtacaaatg aaaaggatat 2040
gtacagcttc ctagaggaca tgggcttaaa ggcattcacc aactcaaaga ttcgtaaacc 2100
caaaatgtgt ccacagcttc aacagtatga aatgcatgga cctgaaggtc tacgtgtagg 2160
tttttatgag tcagatgtaa tgggaagagg ccatgcacgc ctggtgcatg ttgaagagcc 2220
tcacacggag accgtacgaa agtacttccc tgagacatgg atctgggatt tggtggtggt 2280
aaactcagca ggggtggctg aggtaggagt aacagtccct gacaccatca ccgagtggaa 2340
ggcaggggcc ttctgcctgt ctgaagatgc tggacttggt atctcttcca ctgcctctct 2400
ccgagccttc cagcccttct ttgtggagct tacaatgcct tactctgtga ttcgtggaga 2460
ggccttcaca ctcaaggcca cggtcctaaa ctaccttccc aaatgcatcc gggtcagtgt 2520
gcagctggaa gcctctcccg ccttccttgc tgtcccagtg gagaaggaac aagcgcctca 2580
ctgcatctgt gcaaacgggc ggcaaactgt gtcctgggca gtaaccccaa agtcattagg 2640
aaatgtgaat ttcactgtga gcgcagaggc actagagtct caagagctgt gtgggactga 2700
ggtgccttca gttcctgaac acggaaggaa agacacagtc atcaagcctc tgttggttga 2760
acctgaagga ctagagaagg aaacaacatt caactcccta ctttgtccat caggtggtga 2820
ggtttctgaa gaattatccc tgaaactgcc accaaatgtg gtagaagaat ctgcccgagc 2880
ttctgtctca gttttgggag acatattagg ctctgccatg caaaacacac aaaatcttct 2940
ccagatgccc tatggctgtg gagagcagaa tatggtcctc tttgctccta acatctatgt 3000
actggattat ctaaatgaaa cacagcagct tactccagag gtcaagtcca aggccattgg 3060
ctatctcaac actggttacc agagacagtt gaactacaaa cactatgatg gctcctacag 3120
caccttggg gagcgatatg gcaggaacca gggcaacacc tggctcacag cctttgttct 3180
gaagactttt gcccaagctc gagcctacat cttcatcgat gaagcacaca ttacccaagc 3240
cctcatatgg ctctcccaga ggcagaagga caatggctgt ttcaggagct ctgggtcact 3300
gctcaacaat gccataaagg gaggagtaga agatgaagtg accctctccg cctatatcac 3360
catcgccctt ctggagattc ctctcacagt cactcaccct gttgtccgca atgccctgtt 3420
ttgcctggag tcagcctgga agacagcaca agaaggggac catggcagcc atgtatatac 3480
caaagcactg ctggcctatg cttttgccct ggcaggtaac caggacaaga ggaaggaagt 3540
actcaagtca cttaatgagg aagctgtgaa gaaagacaac tctgtccatt gggagcgccc 3600
tcagaaaccc aaggcaccag tggggcattt ttacgaaccc caggctccct ctgctgaggt 3660
ggagatgaca tcctatgtgc tcctcgctta tctcacggcc cagccagccc caacctcgga 3720
ggacctgacc tctgcaacca acatcgtgaa gtggatcacg aagcagcaga atgcccaggg 3780
cggtttctcc tccacccagg acacagtggt ggctctccat gctctgtcca aatatggagc 3840
cgccacattt accaggactg ggaaggctgc acaggtgact atccagtctt cagggacatt 3900
ttccagcaaa ttccaagtgg acaacaacaa tcgcctgtta ctgcagcagg tctcattgcc 3960
agagctgcct ggggaataca gcatgaaagt gacaggagaa ggatgtgtct acctccagac 4020

ctccttgaaa tacaatattc tcccagaaaa ggaagagttc ccctttgctt taggagtgca 4080
gactctgcct caaacttgtg atgaacccaa agcccacacc agcttccaaa tctccctaag 4140
tgtcagttac acagggagcc gctctgcctc caacatggcg atcgttgatg tgaagatggt 4200
ctctggcttc attcccctga agccaacagt gaaaatgctt gaaagatcta accatgtgag 4260
ccggacagaa gtcagcagca accatgtctt gatttacctt gataaggtgt caaatcagac 4320
actgagcttg ttcttcacgg ttctgcaaga tgtcccagta agagatctca aaccagccat 4380
agtgaaagtc tatgattact acgagacgga tgagtttgca atcgctgagt acaatgctcc 4440
ttgcagcaaa gatcttggaa atgcttgaag accacaaggc tgaaaagtgc tttgctggag 4500
tcctgttctc tgagctccac agaagacacg tgtttttgta tctttaaaga cttgatgaat 4560
aaacactttt tctggtc 4577

<210> 11
<211> 1651
<212> DNA
<213> Homo sapiens

<400> 11
ggagcccccg ccctgggatt cccaggtgtt ttcatttggt gatcagcact gaacacagaa 60
gagtcatgac ggagtttggg ctgagctggg ttttccttgt tgctattttt aaaggtgtcc 120
agtgtgaggt gcagctggtg gagtctgggg gagacttggt ccagcctggg gggtccctga 180
gactctcctg tgcagcctct ggattcacct tcagtagtta tgctatgcac tgggtccgcc 240
aggctccagg gaagggactg aaatatgttt caggtattag tagtaatggg cgtagaacat 300
attatgcaaa ttctgtgaag ggcagattca ccatctccag agacaattcc aagaacacgt 360
tgtatcttca aatgggcagc ctgagagctg aggacatggc tgtgtattac tgtgtgtccg 420
ggggaatcta tgatagtagt ggtccctttg actactgggg ccagggaacc ctggtcaccg 480
tctcctcagc atccccgacc agccccaagg tcttcccgct gagcctctgc agcacccagc 540
cagatgggaa cgtggtcatc gcctgcctgg tccagggctt cttcccccag gagccactca 600
gtgtgacctg gagcgaaagc ggacagggcg tgaccgccag aaacttccca cccagccagg 660
atgcctccgg ggacctgtac accacgagca gccagctgac cctgccggcc acacagtgcc 720
tagccggcaa gtccgtgaca tgccacgtga agcactacac gaatcccagc caggatgtga 780
ctgtgccctg cccagttccc tcaactccac ctaccccatc tccctcaact ccacctaccc 840
catctccctc atgctgccac ccccgactgt cactgcaccg accggccctc gaggacctgc 900
tcttaggttc agaagcgaac ctcacgtgca cactgaccgg cctgagagat gcctcaggtg 960
tcaccttcac ctggacgccc tcaagtggga agagcgctgt tcaaggacca cctgagcgtg 1020
acctctgtgg ctgctacagc gtgtccagtg tcctgccggg ctgtgccgag ccatggaacc 1080
atgggaagac cttcacttgc actgctgcct accccgagtc caagaccccg ctaaccgcca 1140
ccctctcaaa atccggaaac acattccggc ccgaggtcca cctgctgccg ccgccgtcgg 1200
aggagctggc cctgaacgag ctggtgacgc tgacgtgcct ggcacgcggc ttcagcccca 1260
aggacgtgct ggttcgctgg ctgcaggggt cacaggagct gccccgcgag aagtacctga 1320
cttgggcatc ccggcaggag cccagccagg gcaccaccac cttcgctgtg accagcatac 1380
tgcgcgtggc agccgaggac tggaagaagg gggacacctt ctcctgcatg gtgggccacg 1440
aggccctgcc gctggccttc acacagaaga ccatcgaccg cttggcgggt aaacccaccc 1500
atgtcaatgt gtctgttgtc atggcggagg tggacggcac ctgctactga gccgcccgcc 1560
tgtccccacc cctgaataaa ctccatgctc ccccaaaaaa aaaaaaaaaa aaaaaaaaaa 1620
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa a 1651

<210> 12
<211> 1120
<212> DNA
<213> Homo sapiens

<400> 12
acttcttcgc accagggaag ccccacccac cagaacgcca agatgtccag caagcgggcc 60
aaagccaagg ccaccaagaa gcggccacag cgggccacat ccaatgtctt cgcaatgttt 120
gaccagtccc agatccagga gtttaaggag gctttcaaca tgattgacca gaaccgtgat 180
ggcttcattg acaaggagga cctgcacgac atgctggcct cgctggggaa gaaccccaca 240
gacgaatacc tggagggcat gatgagcgag gccccggggc catacaactt caccatgttc 300
ctcaccatgt ttggggagaa gctgaacggc acggaccccg aggatgtgat tcgcaacgcc 360
tttgcctgct tcgacgagga atcctcaggt ttcatccatg aggaccacct ccggaagctg 420
ctcaccacca tgggtgaccg cttcacagat gaggaagtgg acgagatgta ccgggaggca 480
cccgttgata agaaaggcaa cttcaactac gtggagttca cccgcatcct caaacatggc 540
gccaaggata aacacgacta ggccatcccc agccccctga cacccagccc ccgccagtca 600

ccccctcccg cacacacccg tccataccag ctccctgccc atgaccctcg ctcagggatc 660
cccctttgag ggttagggtc ccagttccca gtggaagaaa caggccagga gagtgcgtgc 720
cgagctgagg cagatgttcc cacagtgacc ccagagccct gggctatagt ctctgacccc 780
tccaaggaaa gaccaccttc tggggacatg ggctggaggg caggacctag aggcaccaag 840
ggaaccgcat tccggggctg ttccccgagg aggaagggaa gcctctgtgt gccccccagg 900
aggaagaggc cctgagtcct gggatcagac acccccttcac gtgtatccca cacaaatgca 960
agctcaccaa ggtcccctct cagtcccctt ccctacaccc tgacgccaga tgccgcacac 1020
ccaacgccac cagccatggg agtgtgctca ggagtcgcgg ggcagacgtg acatctgtcc 1080
agagggggca gaatctccaa tagaggactg agacaacatg 1120

<210> 13
<211> 2722
<212> DNA
<213> Homo sapiens

<400> 13
gggagtaggt cagacagggt tttcactgtc tagatttagc tggttagatt ttcctctaaa 60
atcttaagca gatgctgaat ccataaagcc aaggagtgct gagggctgga gtgggtgggc 120
cctccctccc acagccccag ggagatacat ctggactcct gggaatctgg attaagaaag 180
ttgggatgaa agaggaggac tgctcaggcc agttgtgagc cagggagaat gttccagtct 240
ggttttggtc tggtggctcc tagaaaggga tgtggcccag aagtgggact caaacagaga 300
gatgaagtag gggaggggag tcctccacgg gcatctctct cctggcggct ggatagcttt 360
accctagggt ggggccagcc ctgacaccct catcccctcc tcagcccagc agcctctaga 420
gcagcgtggt gctctcttga tggccaggaa tccagaatct gcccttcccc aaccttggtt 480
tctccttctg taaagcagag tgatggctga gaagtcctct ctagctgcag aagcctgagc 540
ttcccggtgt cagtcagctc tgcaacgtgg gagccttgga ggcgggtggc ctggctgtgc 600
ttcatatgcc caggagtgga gaaggaggct gagaagctgg tttcccagct ctgatctcca 660
caggcacctc ctataaccct cctctcaccc accccgctac ggtctgagag atctgaaata 720
acctttccca gtgggcaggg ttgccagggt tgaggggaca gcacatacca cccccaccca 780
acctgttcga ggggccctgc atggcacggg atgagtccct gccctgtgca gctgcctggc 840
agtggctggg acaaggatct tgcagccagc acagaggcct cttcaaaggc ctctccctct 900
tggcactcca ggcaaggcag gtgcccgctt ccccaacacc tccaggcagt gaccctaggg 960
catgccccag caggtctccg agcagccact gggacccgtc tcagcacatc ctggcctttg 1020
aaagtctgat atcctgagag gagggcaggt tttagggccg cagttccagc cagcgtcccc 1080
agcctggctt ccctgccatg gactcagtag ctcgtggggc ttcttaccac ccaccagccc 1140
cgctggggtg cggcctggct gtgggcaaag gaggacttgc ctggagattt gagagaagat 1200
tccttctacc agggctgctg aggggccagg cctgcatcag gggctaggct ctggctgggc 1260
ccggaggctg agactaaggc tttcgaccct ggtgcctcca tgtggatgct gcctcagaca 1320
aaggcagtga gccttccctg ccaaagtgcc catcccatgg gctcggcctc actggtcact 1380
gttagcccat gaacacgtgt gggcctcggt cacgtggctt tgagggcagt ctgaccaggc 1440
tagaccacac gtgccgtgac agggggtgcc attcccctcg caggctctaa tgtgcccaca 1500
tgtagcctgg cagtccaaag accaagaatc aacttgcaaa tctgccatta aactgctgtg 1560
cgacttcagg catatcactg ccttctctgg gcttcagtgt ccttttcata cctagaagtc 1620
tgcggtctga ggctctttgg gttcagacac actgttctag gcttctgtag gggaccttgt 1680
gatctgccgt gcccctcctc cctgttcttt tctgtcctcc ccaccccacc ctcagaagct 1740
gcttgctctg cccccaggac aggagcttga cggatgaagt gcagccagcc acccaggtgc 1800
catttccagt ctgacttcca gaaatgtgca ccatgtccta gagcacagac ccattggctg 1860
gagcctcctg ggagggttca aaccatcagc tctatgagaa atgcccagaa aggctttgcc 1920
gactccatcc gtctgtggag gctgcctgcc tccggggtgg gatgggtggt ttctcctcca 1980
attcagaccc aagaggtagc ccccgagggc atgtacctgg tgggaagcag ctcaggtacc 2040
cttgggggtt gcagggccct tacgcaggta tttctctctc tctcctctct ggggtgcgtg 2100
tgtgcgtgcg cgtgtgcgtg cctatgcttt tctctgtggg cacatcagga tgcccctcgg 2160
agagcatgtg cacgtgtccc cacctgagcg agcgtgtgtg tgtgctcctc tgcgtcccag 2220
gtttggacgt ctagggtttg gtgtgcctgt cttctgccct ccctgagccc acagggtcag 2280
tcaatgtatc ttctacgtgc ctctccctct gccttctctc acagtgcccc cggctccaga 2340
gctcaggggt aggggttctc ctgagggtgc aggggatcct tctcatctcc tggaccctcc 2400
agggcactct ggtccctatt ccccagctcc taggcagctg agccgggtcc cttaggggag 2460
gtgaccagga gctttggtgc agggagctct tggtggggca aagggctgga cccctgccag 2520
gtctgtggac atggttatat gcccgggaga ggggggtgca gggccccagg gatggccccc 2580
aatcccacct ctgtttattc tgtaaactgc aacctataaa taacctttag cattcctatt 2640
gtaacaaaat taatttttat gaaataaatt atatttccta gtctaataaa atccgggttg 2700
catttaaaaa aaaaaaaaaa aa 2722

<210> 14
<211> 3612
<212> DNA
<213> Homo sapiens

<400> 14
aagcttaaag tgaagacttg ctgaaattat ttcagcattt tactcctggc cattttccat 60
aggaaggctc ctaatagtgt ctaactcctg aaagcagaaa gtgaccttca aaactgaccc 120
atcttccctg gaatctctct aagagctgag ggattagggt tcttagagat ttgtgtatct 180
ttgatggcac taaaatgaca tagagtttac atgtatacac ctggaaaagg tattaaatgt 240
tatcagatgg ccctcaaagg tctagaacat ggtagctgtg cattgaggaa acattctgca 300
aaaaccattg tctcagaaag aaacatccag ctctatatta attgtatgct cagacccatc 360
cctgctggtt aggtcagcat ttgtacttgg tttgcatatt ataccattgt gtgacatcaa 420
caatctgcaa gtcacatttg ccgggctgcc tgtgcagaat ggcagctgct gggcctagag 480
tctagtgcag gtacttgtga agggatcatt cttactgaat atgtctgttc tcttctatgt 540
actttctgac cctccagggc agctggatac agtggacctg aataagctaa acgagattga 600
aggcacccta aacaaagcca aagatgaaat gaaggtcagc gatcttgata ggaaagtgtc 660
tgacctggag aatgaagcca agaagcagga ggctgccatc atggactata accgagatat 720
cgaggagatc atgaaggaca ttcgcaatct ggaggacatc aggaagacct taccatctgg 780
ctgcttcaac accccgtcca ttgaaaagcc ctagtgtctt tagggctgga aggcagcatc 840
cctctgacag gggggcagtt gtgaggccac agagtgcctt gacacaaaga ttacattttt 900
cagaccccca ctcctctgct gctgtccatc actgtccttt tgaaccagga aaagtcacag 960
agtttaaaga gaagcaaatt aaacatcctg aatcgggaac aaagggtttt atctaataaa 1020
gtgtctcttc catcacgttg ctaccttacc cacacttccc tctgatttgc gtgaggacgt 1080
ggcatcctac ttacgtacgt ggcataacac atcgtgtgag cccatgtatg ctggggtaga 1140
gcaagtagcc ctcccctgtc tcatcgatcc agcagaacct cctcagtctc agtactcttg 1200
tttctataag gaaaagtttt gctactaaca gtagcattgt gatggccagt atatccagtc 1260
catggataaa gaaaatgcat ctgcatctcc tgcccctctt ccttctaagc aaaaggaaat 1320
aaacatcctg tgccaaaggt attggtcatt tagaatgtcg gtagccatcc atcagtgctt 1380
ttagctatta tgagtgtagg acactgagcc atccgtgggt caggatgcaa ttatttataa 1440
aagtccccag gtgaacatgg ctgaagattt ttctagtata ttaataattg actaggaaga 1500
tgaacttttt ttcagatctt tgggcagctg ataatttaaa tctggatggg cagcttgcac 1560
tcaccaatag accaaaagac atcttttgat attcttataa atggaactta cacagaagaa 1620
atagggatat gataaccact aaagttttgt tttcaaaatc aaactaattc ttacagcttt 1680
tttattagtt agtcttggaa ctagtgttaa gtatctggca gagaacagtt aatccctaag 1740
gtcttgacaa aacagaagaa aaacaagcct cctcgtccta gtcttttcta gcaaagggat 1800
aaaacttaga tggcagcttg tactgtcaga atcccgtgta tccatttgtt cttctgttgg 1860
agagatgaga catttgaccc ttagctccag ttttcttctg atgtttccat cttccagaat 1920
ccctcaaaaa acattgtttg ccaaatcctg gtggcaaata cttgcactca gtatttcaca 1980
cagctgccaa cgctatcgag ttcctgcact ttgtgattta aatccactct aaaccttccc 2040
tctaagtgta gagggaagac ccttacgtgg agtttcctag tgggcttctc aacttttgat 2100
cctcagctct gtggttttaa gaccacagtg tgacagttcc ctgccacaca ccccttcct 2160
cctaccaacc caccttttgag attcatatat agcctttaac actatgcaac tttgtacttt 2220
gcgtagcagg ggctggggtg gggggaaaga aacctattat catggacaca ctggtgctat 2280
taattatttc aaatttatat ttttgtgtga atgttttgtg ttttgtttat ccatgctata 2340
gaacaaggaa tttatgtaga tatacttagt cctatttcta gaatgacact ctgttcactt 2400
tgctcaattt ttcctcttca ctggcacaag tatctgaata cctccttccc tcccttctag 2460
agttctttgg attgtactcc aaagaattgt gccttgtgtt tgcagcatct ccattctcta 2520
aattaatata attgctttcc tccacaccca gccacgtaaa gaggtaactt gggtcctctt 2580
ccattgcagt cctgatgatc ctaacctgca gcacggtggt tttacaatgt tccagagcag 2640
gaacgccagg ttgacaagct atggtaggat taggaaagtt tgctgaagag gatctttgac 2700
gccacagtgg gactagccag gaatgaggga gaaatgccct ttttggcaat tgttggagct 2760
ggataggtaa gttttataag ggagtacatt ttgactgagc acttagggca tcaggaacag 2820
tgctacttac tggtgggtag actgggagag gtggtgtaac ttagttcttg atgatcccac 2880
ttcctgtttc catctgcttg ggatatacca gagtttacca caagtgtttt gacgatatac 2940
tcctgagctt tcactctgct ggcttctccc aggcctcttc tactatggca ggagatgtgg 3000
tgtgctgttg caaagttttc acgtcatcgt ttcctggcta gttcatttca ttaagtggct 3060
acatcctaac atatgcattg gtcaaggttg cagcaagagg actgaagatt gactgccaag 3120
ctagtttggg tgaagttcac tccagcaagt ctcaggccac aatggggtgg tttggtttgg 3180
tttcctttta actttctttt tgttatttgc ttttctcctc cacctgtgtg gtatattttt 3240
taagcagaat tttatttttt aaaataaaag gttctttaca agatgatacc ttaattacac 3300
tcccgcaaca cagccattat tttattgtct agctccagtt atctgtattt tatgtaatgt 3360
aattgacagg atggctgctg cagaatgctg gttgacacag ggattattat actgctattt 3420

ttccctgaat tcttttcctt ggaattccaa ctgtggacct tttatatgtg ccttcacttt 3480
agctgtttgc cttactctac agccttgctc tccggggtgg ttaataaaat gcaacacttg 3540
gcatttttat gttataagaa aaacagtatt ttatttataa taaaatctga atattttgta 3600
acccttttata tt 3612

<210> 15
<211> 2474
<212> DNA
<213> Homo sapiens

<400> 15
agcgactcac aaaaggtcct cggccacggc gtgcgtcacc atggcgaccg ccgccggcag 60
ccgcgccccg cccctctggc gggaccggcc accatcctct cgcgaggagc atcccgtgcg 120
accggaagtg gggcggcgac cccggaagtc cccgccgggt gcagcttggt cggttcgatc 180
gccgccggga cctgacaccg cccggagttg gcgtcccttc tccctctccg agtgctgctc 240
ctgtcattgt ggccatggac gataccctgt tccagttgaa gttcacggcg aagcagctgg 300
agaagctggc caagaaggcg gagaaggact ccaaggcgga gcaggccaaa gtgaagaagg 360
cccttctgca gaaaaatgta gagtgtgccc gtgtgtatgc cgagaacgcc atccgcaaga 420
agaacgaagg tgtgaactgg cttcggatgg cgtcccgcgt agacgcagtg gcctccaagg 480
tggacacagc tgtgactatg aaggggtga ccaagaatat ggcccaggtg accaaagccc 540
tggacaaggc cctgagcacc atggacctgc agaaggtctc ctcagtgatg gacaggttcg 600
agcagcaggt gcagaacctg gacgtccata catcggtgat ggaggactcc atgagctcgg 660
ccaccaccct gaccacgccg caggagcagg tggacagcct catcatgcag atcgccgagg 720
agaatggcct ggaggtgctg gaccagctca gccagctgcc cgagggcgcc tctgccgtgg 780
gcgagagctc tgtgcgcagc caggaggacc agctgtcacg gaggttggcc gccttgagga 840
actagccgtg ccccgccggt gtgcaccgcc tctgccccgt gatgtgctgg aaggctcctg 900
tcctctcccc accgcgtctt gcctttgtgc tgaccccgcg gggctgcggc cggcagccac 960
tctgcgtctc tcacctgcca ggcctgcgtg gccttagggt tgttcctgtt cttttaggtt 1020
gggcggtggg tctgtgtcct ggtgttgagt ttctgcaaat ttctgggggt gatttctgtg 1080
actctgggcc cacagcgggg aggccaagag gggccctgtg gactttcacc cagcactgtg 1140
ggggccttca gactctgggg cagcagacat gctgcttccc atcagccaga gggggtcagg 1200
gctgccctgt tgccaaacaa ctccctgagg cctctccgca ccacctcagc gggcaggagg 1260
tcccaccatg tggacagaca tagcccaagg aggcaccaca ggtctatgtg tgctggggga 1320
tgtcaggtgc cacccaacgc tgtcctggtg gtatttacaa tgacatcctc ctcctccatc 1380
actccagggg tggtgtctcg gccgccccta ccagctggct gagccccctg gcctcctgcg 1440
ctccctcact tccctcagtt cccaaagctg cccagtccat ggggacagaa ccgtcactca 1500
gatccacatt caagtgtgcc caccctgcag tcttcatcct cactcagctg ctgcctctgg 1560
aggtgccttt ggccacatgt gctgtgctgt ttgtctcctc gacagggagc ctgtccacca 1620
gcaggctgcg gtcccagcgg gtgcgtctgc agctcctccc cttgggcagg ctggttctcc 1680
cggaggacct ttccttgggg ccctgcttca tgacgatgct gcctgtgtca ccctctacca 1740
tctgtaaaca actgggtgcc ttccccgacc acaccccaat gccttcccag cttggaagcc 1800
aaggcagctg atgaagggag ctcaggagag ccgtcttcag ctgggctggg gttggggctg 1860
ctgtgaggaa aacctgccat tgtggccctg gagagtcacc agcagctctt gggaaggact 1920
tgctgggagg ctgagagagg ctttgggcac agcctgctgt cttttccatt tcctaaagtt 1980
tacttcattg tcttgaggct tccaggtttt gtttttgttt ttgccaaagt agaaaaggca 2040
ggtggtgggc ggctggcagg gagtgcgggt ccccgcccct cttcagtcct gccctcccct 2100
cctcagtcct gcccaccccg tgcagcccat gctgaggctg cagtggtgtc gtgggtgtta 2160
cgtgcaggaa cgtggagacc ctgacgtggg ctcactgcat ttggttttct tttcagaact 2220
tgggagcccc cagggagggg ctagtgttgg taggtcctag acgtggttcc ctccagcctc 2280
cccaaaatca accctggtgt tgagagaacg tccttctgtc catcgtgggt aacagccttg 2340
gggagggtgc agagctctgc agagccatgg gccaggtggg gctgcctcag tcctgtcccc 2400
ttgggcactg aggagagggg cccattcacc tttctcctag aatgctgttg taaataaaca 2460
aatggatccc tgga 2474

<210> 16
<211> 612
<212> DNA
<213> Homo sapiens

<400> 16
ggaggccaag gtgcaacttc cttcggtcgt cccgaatccg ggttcatccg acaccagccg 60
cctccaccat gccgccgaag ttcgacccca acgagatcaa agtcgtatac ctgaggtgca 120

ccggaggtga agtcggtgcc acttctgccc tggcccccaa gatcggcccc ctgggtctgt 180
ctccaaaaaa agttggtgat gacattgcca aggcaacggg tgactggaag ggcctgagga 240
ttacagtgaa actgaccatt cagaacagac aggcccagat tgaggtggtg ccttctgcct 300
ctgccctgat catcaaagcc ctcaaggaac caccaagaga cagaaagaaa cagaaaaaca 360
ttaaacacag tgggaatatc acttttgatg agattgtcaa cattgctcga cagatgcggc 420
accgatcctt agccagagaa ctctctggaa ccattaaaga gatcctgggg actgcccagt 480
cagtgggctg taatgttgat ggccgccatc ctcatgacat catcgatgac atcaacagtg 540
gtgctgtgga atgcccagcc agttaagcac aaaggaaaac atttcaataa aggatcattt 600
gacaactggt ga 612

<210> 17
<211> 1100
<212> DNA
<213> Homo sapiens

<400> 17
ctgcagtatt tagcatgccc cacccatctg caaggcattc tggatagtgt caaaacagcc 60
ggaaatcaag tccgtttatc tcaaacttta gcattttggg aataaatgat atttgctatg 120
ctggttaaat tagattttag ttaaatttcc tgctgaagct ctagtacgat aagcaacttg 180
acctaagtgt aaagttgaga tttccttcag gtttatatag cttgtgcgcc gcctgggtac 240
ctcggatgtg agggcgatct ggctgcgaca tctgtcaccc cattgatcgc cagggttgat 300
tcggctgatc tggctggcta ggcgggtgtc cccttcctcc ctcaccgctc catgtgcgtc 360
cctcccgaag ctgcgcgctc ggtcgaagag gacgaccatc cccgatagag gaggaccggt 420
cttcggtcaa gggtatacga gtagctgcgc tcccctgcta gaacctccaa acaagctctc 480
aaggtccatt tgtaggagaa cgtagggtag tcaagcttcc aagactccag acacatccaa 540
atgaggcgct gcatgtggca gtctgccttt cttttgaccc attacccatc taaggaccgg 600
tcttcggtca agggtatacg agtagctgcg ctcccctgct agaacctcca aacaagctct 660
caaggtccat ttgtaggaga acgtagggta gtcaagcttc caagactcca gacacatcca 720
aatgaggcgc tgcatgtggc agtctgcctt tcttttgacc cattacccat ctaagttaga 780
tgctttttta aatgtttttt aatttttaaa ttttaatttt ttttcattat ttatttttta 840
tttttgagac ggatctcggc tcactgtaac ctccacctcc cgggttcaag cgattctcct 900
gactcagcct cccgagtagc tgggattaca ggcgcgcgcc accatgcccg gctaattttt 960
gtatttttag taaagacggg ttttcgccat gttgaccagg gtggtcttga actcctgacg 1020
tcaggtgatc ctcctgcctc gtgctgggat tgcaggcgtg agccaccgct cccggccctt 1080
agatgctttt taatcagcaa 1100

<210> 18
<211> 4747
<212> DNA
<213> Homo sapiens

<400> 18
taggaatgaa ccatccctgg gagtatgtgg aaacaacgga ggagctctga cttcccaact 60
gtcccattct atgggcgaag gaactgctcc tgacttcagt ggttaagggc agaattgaaa 120
ataattctgg aggaagataa gaatgattcc tgcgcgactg caccgggact acaaagggct 180
tgtcctgctg ggaatcctcc tggggactct gtgggagacc ggatgcaccc agatacgcta 240
ttcagttccg gaagagctgg agaaaggctc tagggtgggc gacatctcca gggacctggg 300
gctggagccc cgggagctcg cggagcgcgg agtccgcatc atccccagag gtaggacgca 360
gcttttcgcc ctgaatccgc gcagcggcag cttggtcacg gcgggcagga tagaccggga 420
ggagctctgt atgggggcca tcaagtgtca attaaatcta gacattctga tggaggataa 480
agtgaaaata tatggagtag aagtagaagt aagggacatt aacgacaatg cgccttactt 540
tcgtgaaagt gaattagaaa taaaaattag tgaaaatgca gccactgaga tgcggttccc 600
tctaccccac gcctgggatc cggatatcgg gaagaactct ctgcagagct acgagctcag 660
cccgaacact cacttctccc tcatcgtgca aaatggagcc gacggtagta agtaccccga 720
attggtgctg aaacgcgccc tggaccgcga agaaaaggct gctcaccacc tggtccttac 780
ggcctccgac gggggcgacc cggtgcgcac aggcaccgcg cgcatccgcg tgatggttct 840
ggatgcgaac gacaacgcac cagcgtttgc tcagcccgag taccgcgcga gcgttccgga 900
gaatctggcc ttgggcacgc agctgcttgt agtcaacgct accgaccctg acgaaggagt 960
caatgcggaa gtgaggtatt ccttccggta tgtggacgac aaggcggccc aagttttcaa 1020
actagattgt aattcaggga caatatcaac aataggggag ttggaccacg aggagtcagg 1080
attctaccag atggaagtgc aagcaatgga taatgcagga tattctgcgc gagccaaagt 1140
cctgatcact gttctggacg tgaacgacaa tgccccagaa gtggtcctca cctctctcgc 1200

cagctcggtt cccgaaaact ctcccagagg gacattaatt gccctttaa atgtaaatga 1260
ccaagattct gaggaaaacg gacaggtgat ctgtttcatc caaggaaatc tgccctttaa 1320
attagaaaaa tcttacggaa attactatag tttagtcaca gacatagtct tggataggga 1380
acaggttcct agctacaaca tcacagtgac cgccactgac cggggaaccc cgcccctatc 1440
cacggaaact catatctcgc tgaacgtggc agacaccaac gacaacccgc cggtcttccc 1500
tcaggcctcc tattccgctt atatcccaga gaacaatccc agaggagttt ccctcgtctc 1560
tgtgaccgcc cacgaccccg actgtgaaga gaacgcccag atcacttatt ccctggctga 1620
gaacaccatc caaggggcaa gcctatcgtc ctacgtgtcc atcaactccg acactggggt 1680
actgtatgcg ctgagctcct tcgactacga gcagttccga gacttgcaag tgaaagtgat 1740
ggcgcgggac aacgggcacc cgcccctcag cagcaacgtg tcgttgagcc tgttcgtgct 1800
ggaccagaac gacaatgcgc ccgagatcct gtaccccgcc ctccccacgg acggttccac 1860
tggcgtggag ctggctcccc gctccgcaga gcccggctac ctggtgacca aggtggtggc 1920
ggtggacaga gactccggcc agaacgcctg gctgtcctac cgtctgctca aggccagcga 1980
gccgggactc ttctcggtgg gtctgcacac gggcgaggtg cgcacggcgc gagccctgct 2040
ggacagagac gcgctcaagc agagcctcgt agtggccgtc caggaccacg gccagccccc 2100
tctctccgcc actgtcacgc tcaccgtggc cgtggccgac agcatccccc aagtcctggc 2160
ggacctcggc agcctcgagt ctccagctaa ctctgaaacc tcagacctca ctctgtacct 2220
ggtggtagcg gtggccgcgg tctcctgcgt cttcctggcc ttcgtcatct tgctgctggc 2280
gctcaggctg cggcgctggc acaagtcacg cctgctgcag gcttcaggag gcggcttgac 2340
aggagcgccg gcgtcgcact ttgtgggcgt ggacggggtg caggctttcc tgcagaccta 2400
ttcccacgag gtttccctca ccacggactc gcggaagagt cacctgatct tcccccagcc 2460
caactatgca gacatgctcg tcagccagga gagctttgaa aaaagcgagc ccctttgct 2520
gtcaggtgat tcggtatttt ctaaagacag tcatgggtta attgagcaag ccccgcccaa 2580
cacggactgg cgtttctctc aggcccagag acccggcacc agcggctccc aaaatggcga 2640
tgacaccggc acctggccca acaaccagtt tgacacagaa atgctgcaag ccatgatctt 2700
ggcgtccgcc agtgaagctg ctgatgggag ctccaccctg ggaggggtg ccggcaccat 2760
gggattgagc gcccgctacg gaccccagtt caccctgcag cacgtgcccg actaccgcca 2820
gaatgtctac atcccaggca gcaatgccac actgaccaac gcagctggca agcgggatgg 2880
caaggcccca gcaggtggca atggcaacaa gaagaagtcg ggcaagaagg agaagaagta 2940
acatggaggc caggccaaga gccacagggc ggcctctccc caaccagccc agcttctcct 3000
tacctgcacc caggcctcag agtttcaggg ctaaccccca gaatactggt aggggccaag 3060
gccatgctcc ccttgggaaa cagaaacaag tgcccagtca gcacctaccc cttccccccc 3120
agggggttga atatgcaaaa gcagttccgc tgggaacccc catccaatca actgctgtac 3180
ccatgggggt agtggggtta ctgtagacac caagaaccat ttgccacacc ccgtttagtt 3240
acagctgaac tcctccatct tccaaatcaa tcaggcccat ccatcccatg cctccctcct 3300
ccccacccca ctccaacagt tcctctttcc cgagtaaggt ggttggggtg ttgaagtacc 3360
aagtaaccta caagcctcct agttctgaaa agttggaagg gcatcatgac ctcttggcct 3420
ctcctttgat tctcaatctt cccccaaagc atggtttggt gccagcccct tcacctcctt 3480
ccagagccca agatcaatgc tcaagttttg gaggacatga tcaccatccc catggtactg 3540
atgcttgctg gatttaggga gggcattttg ctaccaagcc tcttcccaac gccctgggga 3600
ccagtcttct gttttgtttt tcattgtttg acgtttccac tgcatgcctt gacttccccc 3660
acctcctcct caaacaagag actccactgc atgttccaag acagtatggg gtggtaagat 3720
aaggaaggga agtgtgtgga tgtggatggt gggggcatgg acaaagcttg acacatcaag 3780
ttatcaaggc cttggaggag gctctgtatg tcctcagggg actgacaaca tcctccagat 3840
tccagccata aaccaataac taggctggac ccttcccact acataatagg gctcagccca 3900
ggcagccagc tttgggctga gctaacagga ccaatggatt aaactggcat ttcagtccaa 3960
ggaagctcga agcaggttta ggaccaggtc cccttgagag gtcagagggg cctctgtggg 4020
tgctgggtac tccagaggtg ccactggtgg aagggtcagc ggagccccag caggaagggt 4080
gggccagcca ggccattctt agtccctggg ttggggaggc agggagctag ggcagggacc 4140
aaatgaacag aaagtctcag cccaggatgg ggcttcttca acagggcccc tgccctcctg 4200
aagcctcagt ccttcacctt gccaggtgcc gtttctcttc cgtgaaggcc actgcccagg 4260
tccccagtgc gccccctagt ggccatagcc tggttaaagt tccccagtgc ctccttgtgc 4320
atagaccttc ttctcccacc cccttctgcc cctgggtccc cggccatcca gcggggctgc 4380
cagagaaccc cagacctgcc cttacagtag tgtagcgccc cctccctctt tcggctggtg 4440
tagaatagcc agtagtgtag tgcggtgtgc ttttacgtga tggcgggtgg gcagcgggcg 4500
gcgggctccg cgcagccgtc tgtccttgat ctgcccgcgg cggcccgtgt tgtgttttgt 4560
gctgtgtcca cgcgctaagg cgacccccctc ccccgtactg acttctccta taagcgcttc 4620
tcttcgcata gtcacgtagc tcccacccca ccctcttcct gtgtctcacg caagttttat 4680
actctaatat ttatatggct ttttttcttc gacaaaaaaa taataaaacg tttcttctga 4740
aaagctg 4747

<210> 19
<211> 6909

<212> DNA
<213> Homo sapiens

<400> 19
gaattccgca gcccggagtg tggttagcag ctcggcaagc gctgcccagg tcctggggtg 60
gtggcagcca gcgggagcag gaaaggaagc atgttcccag gctgcccacg cctctgggtc 120
ctggtggtct tgggcaccag ctgggtaggc tgggggagcc aagggacaga agcggcacag 180
ctaaggcagt tctacgtggc tgctcagggc atcagttgga gctaccgacc tgagcccaca 240
aactcaagtt tgaatctttc tgtaacttcc tttaagaaaa ttgtctacag agagtatgaa 300
ccatatttta agaaagaaaa accacaatct accatttcag gacttcttgg gcctacttta 360
tatgctgaag tcggagacat cataaaagtt cactttaaaa ataaggcaga taagcccttg 420
agcatccatc ctcaaggaat taggtacagt aaattatcag aaggtgcttc ttaccttgac 480
cacacattcc ctgcggagaa gatggacgac gctgtggctc caggccgaga atacacctat 540
gaatggagta tcagtgagga cagtggaccc acccatgatg accctccatg cctcacacac 600
atctattact cccatgaaaa tctgatcgag gatttcaact cggggctgat tgggcccctg 660
cttatctgta aaaaagggac cctaactgag ggtgggacac agaagacgtt tgacaagcaa 720
atcgtgctac tatttgctgt gtttgatgaa agcaagagct ggagccagtc atcatcccta 780
atgtacacag tcaatggata tgtgaatggg acaatgccag atataacagt ttgtgcccat 840
gaccacatca gctggcatct gctgggaatg agctcggggc cagaattatt ctccattcat 900
ttcaacggcc aggtcctgga gcagaaccat cataaggtct cagccatcac ccttgtcagt 960
gctacatcca ctaccgcaaa tatgactgtg ggcccagagg gaaagtggat catatcttct 1020
ctcaccccaa aacatttgca agctgggatg caggcttaca ttgacattaa aaactgccca 1080
aagaaaacca ggaatcttaa gaaaataact cgtgagcaga ggcggcacat gaagaggtgg 1140
gaatacttca ttgctgcaga ggaagtcatt tgggactatg cacctgtaat accagcgaat 1200
atggacaaaa aatacaggtc tcagcatttg gataatttct caaaccaaat tggaaaacat 1260
tataagaaag ttatgtacac acagtacgaa gatgagtcct tcaccaaaca tacagtgaat 1320
cccaatatga aagaagatgg gattttgggt cctattatca gagcccaggt cagagacaca 1380
ctcaaaatcg tgttcaaaaa tatggccagc cgcccctata gcatttaccc tcatggagtg 1440
accttctcgc cttatgaaga tgaagtcaac tcttctttca cctcaggcag gaacaacacc 1500
atgatcagag cagttcaacc aggggaaacc tatacttata agtggaacat cttagagttt 1560
gatgaaccca cagaaaatga tgcccagtgc ttaacaagac catactacag tgacgtggac 1620
atcatgagag acatcgcctc tgggctaata ggactacttc taatctgtaa gagcagatcc 1680
ctggacaggc gaggaataca gagggcagca gacatcgaac agcaggctgt gtttgctgtg 1740
tttgatgaga acaaaagctg gtaccttgag gacaacatca acaagttttg tgaaaatcct 1800
gatgaggtga aacgtgatga ccccaagttt tatgaatcaa acatcatgag cactatcaat 1860
ggctatgtgc ctgagagcat aactactctt ggattctgct ttgatgacac tgtccagtgg 1920
cacttctgta gtgtggggac ccagaatgaa attttgacca tccacttcac tgggcactca 1980
ttcatctatg gaaagaggca tgaggacacc ttgaccctct tccccatgcg tggagaatct 2040
gtgacggtca caatggataa tgttggaact tggatgttaa cttccatgaa ttctagtcca 2100
agaagcaaaa agctgaggct gaaattcagg gatgttaaat gtatcccaga tgatgatgaa 2160
gactcatatg agatttttga acctccagaa tctacagtca tggctacacg gaaaatgcat 2220
gatcgtttag aacctgaaga tgaagagagt gatgctgact atgattacca gaacagactg 2280
gctgcagcat taggaattag gtcattccga aactcatcat tgaaccagga agaagaagag 2340
ttcaatctta ctgccctagc tctggagaat ggcactgaat tcgtttcttc gaacacagat 2400
ataattgttg gttcaaatta ttcttcccca agtaatatta gtaagttcac tgtcaataac 2460
cttgcagaac ctcagaaagc cccttctcac caacaagcca ccacagctgg ttccccactg 2520
agacacctca ttggcaagaa ctcagttctc aattcttcca cagcagagca ttccagccca 2580
tattctgaag accctataga ggatcctcta cagccagatg tcacagggat acgtctactt 2640
tcacttggtg ctggagaatt cagaagtcaa gaacatgcta agcgtaaggg acccaaggta 2700
gaaagagatc aagcagcaaa gcacaggttc tcctggatga aattactagc acataaagtt 2760
gggagacacc taagccaaga cactggttct ccttccggaa tgaggccctg ggaggacctt 2820
cctagccaag acactggttc tccttccaga atgaggccct gggaggaccc tcctagtgat 2880
ctgttactct taaaacaaag taactcatct aagattttgg ttgggagatg gcatttggct 2940
tctgagaaag gtagctatga aataatccaa gatactgatg aagacacagc tgttaacaat 3000
tggctgatca gcccccagaa tgcctcacgt gcttggggag aaagcacccc tcttgccaac 3060
aagcctggaa agcagagtgg ccacccaaag tttcctagag ttagacataa atctctacaa 3120
gtaagacagg atggaggaaa gagtagactg aagaaaagcc agtttctcat taagacacga 3180
aaaaagaaaa aagagaagca cacacaccat gctcctttat ctccgaggac ctttcaccct 3240
ctaagaagtg aagcctacaa cacattttca gaaagaagac ttaagcattc gttggtgctt 3300
cataaatcca atgaaacatc tcttcccaca gacctcaatc agacattgcc ctctatggat 3360
tttggctgga tagcctcact tcctgaccat aatcagaatt cctcaaatga cactggtcag 3420
gcaagctgtc ctccaggtct ttatcagaca gtgcccccag aggaacacta tcaaacattc 3480
cccattcaag accctgatca aatgcactct acttcagacc ccagtcacag atcctcttct 3540
ccagagctca gtgaaatgct tgagtatgac cgaagtcaca agtccttccc cacagatata 3600

agtcaaatgt cccctttctc agaacatgaa gtctggcaga cagtcatctc tccagacctc 3660
agccaggtga ccctctctcc agaactcagc cagacaaacc tctctccaga cctcagccac 3720
acgactctct ctccagaact cattcagaga aacctttccc cagccctcgg tcagatgccc 3780
atttctccag acctcagcca tacaaccctt tctccagacc tcagccatac aaccctttct 3840
ttagacctca gccagacaaa cctctctcca gaactcagtc agacaaacct ttccccagcc 3900
ctcggtcaga tgccctttc tccagacctc agccatacaa ccctttctct agacttcagc 3960
cagacaaacc tctctccaga actcagccat atgactctct ctccagaact cagtcagaca 4020
aacctttccc cagcccttgg tcagatgccc atttctccag acctcagcca tacaaccctt 4080
tctctagact tcagccagac aaacctctct ccagaactca gtcaaacaaa cctttcccca 4140
gccctcggtc agatgccctt ttctccagac cccagccata caacccttc tctagacctc 4200
agccagacaa acctctctcc agaactcagt cagacaaacc tttccccaga cctcagtgag 4260
atgcccctct ttgcagatct cagtcaaatt ccccttaccc cagacctcga ccagatgaca 4320
ctttctccag accttggtga gacagatctt tccccaaact ttggtcagat gtccctttcc 4380
ccagacctca gccaggtgac tctctctcca gacatcagtg acaccaccct tctcccggat 4440
ctcagccaga tatcacctcc tccagacctt gatcagatat tctacccttc tgaatctagt 4500
cagtcattgc ttcttcaaga atttaatgag tcttttcctt atccagacct tggtcagatg 4560
ccatctcctt catctcctac tctcaatgat acttttctat caaaggaatt taatccactg 4620
gttatagtgg gcctcagtaa agatggtaca gattacattg agatcattcc aaaggaagag 4680
gtccagagca gtgaagatga ctatgctgaa attgattatg tgccctatga tgacccctac 4740
aaaactgatg ttaggacaaa catcaactcc tccagagatc ctgacaacat tgcagcatgg 4800
tacctccgca gcaacaatgg aaacagaaga aattattaca ttgctgctga agaaatatcc 4860
tgggattatt cagaatttgt acaaagggaa acagatattg aagactctga tgatattcca 4920
gaagatacca catataagaa agtagttttt cgaaagtacc tcgacagcac ttttaccaaa 4980
cgtgatcctc gaggggagta tgaagagcat ctcggaattc ttggtcctat tatcagagct 5040
gaagtggatg atgttatcca agttcgtttt aaaaatttag catccagacc gtattctcta 5100
catgcccatg gactttccta tgaaaaatca tcagagggaa agacttatga agatgactct 5160
cctgaatggt ttaaggaaga taatgctgtt cagccaaata gcagttatac ctacgtatgg 5220
catgccactg agcgatcagg gccagaaagt cctggctctg cctgtcgggc ttgggcctac 5280
tactcagctg tgaacccaga aaaagatatt cactcaggct tgataggtcc cctcctaatc 5340
tgccaaaaag gaatactaca taaggacagc aacatgcctg tggacatgag agaatttgtc 5400
ttactattta tgacctttga tgaaaagaag agctggtact atgaaaagaa gtcccgaagt 5460
tcttggagac tcacatcctc agaaatgaaa aaatcccatg agtttcacgc cattaatggg 5520
atgatctaca gcttgcctgg cctgaaaatg tatgagcaag agtgggtgag gttacacctg 5580
ctgaacatag gcggctccca agacattcac gtggttcact ttcacggcca gaccttgctg 5640
gaaaatggca ataaacagca ccagttaggg gtctggcccc ttctgcctgg ttcatttaaa 5700
actcttgaaa tgaaggcatc aaaacctggc tggtggctcc taaacacaga ggttggagaa 5760
aaccagagag cagggatgca aacgccattt cttatcatgg acagagactg taggatgcca 5820
atgggactaa gcactggtat catatctgat tcacagatca aggcttcaga gtttctgggt 5880
tactgggagc ccagattagc aagattaaac aatggtggat cttataatgc ttggagtgta 5940
gaaaaacttg cagcagaatt tgcctctaaa ccttggatcc aggtggacat gcaaaaggaa 6000
gtcataatca cagggatcca gacccaaggt gccaaacact acctgaagtc ctgctatacc 6060
acagagttct atgtagctta cagttccaac cagatcaact ggcagatctt caaagggaac 6120
agcacaagga atgtgatgta ttttaatggc aattcagatg cctctacaat aaaagagaat 6180
cagtttgacc cacctattgt ggctagatat attaggatct ctccaactcg agcctataac 6240
agacctaccc ttcgattgga actgcaaggt tgtgaggtaa atggatgttc cacacccctg 6300
ggtatggaaa atggaaagat agaaaacaag caaatcacag cttcttcgtt taagaaatct 6360
tggtggggag attactggga acccttccgt gcccgtctga atgcccaggg acgtgtgaat 6420
gcctggcaag ccaaggcaaa caacaataag cagtggctag aaattgatct actcaagatc 6480
aagaagataa cggcaattat aacacagggc tgcaagtctc tgtcctctga aatgtatgta 6540
aagagctata ccatccacta cagtgagcag ggagtggaat ggaaaccata caggctgaaa 6600
tcctccatgg tggacaagat ttttgaagga aatactaata ccaaaggaca tgtgaagaac 6660
tttttcaacc ccccaatcat ttccaggttt atccgtgtca ttcctaaaac atggaatcaa 6720
agtattacac ttcgcctgga actctttggc tgtgatattt actagaattg aacattcaaa 6780
aacccctgga agagactctt taagacctca aaccatttag aatgggcaat gtattttacg 6840
ctgtgttaaa tgttaacagt tttccactat ttctctttct tttctattag tgaataaaat 6900
tttatacaa 6909

<210> 20
<211> 1933
<212> DNA
<213> Homo sapiens

<400> 20

ccgccgcctt ctactccgcc gcgggggtcg cagcggctgc cgcgccgtcc tcgagtttcc 60
agcgtgagga ggaggctgag ggcggagagg cgcatcgtgt tcgaggcgga gaccgagggg 120
gagccccgcg cgcggcgtcg ctcattgcta tggacagtgc tatcaccctg tggcagttcc 180
ttcttcagct cctgcagaag cctcagaaca agcacatgat ctgttggacc tctaatgatg 240
ggcagtttaa gcttttgcag gcagaagagg tggctcgtct ctgggggatt cgcaagaaca 300
agcctaacat gaattatgac aaactcagcc gagccctcag atactattat gtaaagaata 360
tcatcaaaaa agtgaatggt cagaagtttg tgtacaagtt tgtctcttat ccagagattt 420
tgaacatgga tccaatgaca gtgggcagga ttgagggtga ctgtgaaagt ttaaacttca 480
gtgaagtcag cagcagttcc aaagatgtgg agaatggagg gaaagataaa ccacctcagc 540
ctggtgccaa gacctctagc cgcaatgact acatacactc tggcttatat tcttcattta 600
ctctcaactc tttgaactcc tccaatgtaa agcttttcaa attgataaag actgagaatc 660
cagccgagaa actggcagag aaaaaatctc ctcaggagcc cacaccatct gtcatcaaat 720
ttgtcacgac accttccaaa aagccaccag ttgaacctgt tgctgccacc atttcaattg 780
gcccaagtat ttctccatct tcagaagaaa ctatccaagc tttggagaca ttggtttccc 840
caaaactgcc ttccctggaa gccccaacct ctgcctctaa cgtaatgact gcttttgcca 900
ccacaccacc catttcgtcc ataccccctt tgcaggaacc tcccagaaca ccttcaccac 960
cactgagttc tcacccagac atcgacacag acattgattc agtggcttct cagccaatgg 1020
aacttccaga gaatttgtct ctggagccta aagaccagga ttcagtcttg ctagaaaagg 1080
acaaagtaaa taattcatca agatccaaga aacccaaagg gttaggactg gcacccaccc 1140
ttgtgatcac gagcagtgat ccaagcccac tgggaatact gagcccatct ctccctacag 1200
cttctcttac accagcattt ttttcacaga cacccatcat actgactcca agccccttgc 1260
tctccagtat ccacttctgg agtactctca gtcctgttgc tcccctaagt ccagccagac 1320
tgcaaggtgc taacacactt ttccagtttc cttctgtact gaacagtcat gggccattca 1380
ctctgtctgg gctggatgga ccttccaccc ctggcccatt ttccccagac ctacagaaga 1440
cataacctat gcacttgtgg aatgagagaa ccgaggaacg aagaaacaga cattcaacat 1500
gattgcattt gaagtgagca attgatagtt ctacaatgct gataatagac tattgtgatt 1560
tttgccattc cccattgaaa acatcttttt aggattctct ttgaatagga ctcaagttgg 1620
actatatgta taaaaatgcc ttaattggag tctaaactcc acctccctct gtcttttcct 1680
tttctttttc tttccttcct tcctttttctt ttctccttta aaaatatttt gagctttgtg 1740
ctgaagaagt ttttggtggg ctttagtgac tgtgctttgc aaaagcaatt aagaacaaag 1800
ttactccttc tggctattgg gacccttttgg ccaggaaaaa ttatgcttag aatctattat 1860
ttaaagaagt atttgtgaaa tgaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1920
aaaaaaaaaa aaa 1933

<210> 21
<211> 2215
<212> DNA
<213> Homo sapiens

<400> 21
actcagtaga ccgccactgg ctgtgcacgt tatggggttt ccacctaggg ctcggcctga 60
ggcttgtaac actccgtttt cccccgagtc acaggggcag tcttgccccct cgcagctggg 120
tcgcggtgtc tctcaaaggt ccccctctac aggggcttcg tgaggcccgg gcccacaggg 180
cgctcggtcc cggaagtgac gtctcccaga ggggccggaa gtggcagtgg agggagggaa 240
gatggcggag gtgggggaga taatcgaggg ctgccgccta cccgtgctgc ggcggaacca 300
ggacaacgaa gatgagtggc ccctggccga gatcctgagc gtgaaggaca tcagtggccg 360
gaagcttttc tacgtccatt acattgactt caacaaacgt ctggatgaat gggtgacgca 420
tgagcggctg gacctaaaga agatccagtt ccccaagaaa gaggccaaga cccccactaa 480
gaacggactt cctgggtccc gtcctggctc tccagagaga gaggtgccgg cctcggcgca 540
ggccagcggg aagaccttgc caatcccggt ccagatcaca ctccgcttca acctgcccaa 600
ggagcgggag gccattcccg gtggcgagcc tgaccagccg ctctcctcca gctcctgcct 660
gcagcccaac caccgctcaa cgaaacggaa ggtggaggtg gtttcaccag caactccagt 720
gcccagcgag acagccccgg cctcggtttt tccccagaat ggagccgccc gtagggcagt 780
ggcagcccag ccaggacgga agcgaaaatc gaattgtttg ggcactgatg aggactccca 840
ggacagctct gatggaatac cgtcagcacc acgcatgact ggcagcctgg tgtctgatcg 900
aagccacgac gacatcgtca cccggatgaa gaacattgag tgcattgagc tgggccggca 960
ccgcctcaag ccgtggtact tctccccgta cccacaggaa ctcaccacat tgcctgtcct 1020
ctacctgtgc gagttctgcc tcaagtacgg ccgtagtctc aagtgtcttc agcgtcattt 1080
gaccaagtgt gacctacgac atcctccagg caatgagatt taccgcaagg gcaccatctc 1140
cttctttgag attgatggac gtaagaacaa gagttattcc cagaacctgt gtcttttggc 1200
caagtgtttc cttgaccata agacactgta ctatgacaca gacccttttcc tcttctacgt 1260
catgacagag tatgactgta agggcttcca catcgtgggc tacttctcca aggagaaaga 1320
atcaacggaa gactacaatg tggcctgcat cctaaccctg cctccctacc agcgccgggg 1380

ctaccggaag ctgctgatcg agttcagcta tgaactctcc aaagtggaag ggaaaacagg 1440
gacccctgag aagcccctct cagaccttgg cctcctatcc tatcgaagct actggtccca 1500
gaccatcctg gagatcctga tggggctgaa gtcggagagc ggggagaggc cacagatcac 1560
catcaatgag attagtgaaa tcaccagcat caagaaggag gatgtcatct ccactctgca 1620
gtacctcaat ctcatcaact actacaaggg ccagtacatc ctcacactgt cagaggacat 1680
cgtggatggc catgagcggg ccatgctcaa gcggctcctg cggatcgact ccaagtgtct 1740
gcacttcact cccaaggact ggagcaagag ggggaagtgg tgaccagaca ctgcccactg 1800
cagtgccaag acggcagcag gactggggct gatagcccac cccgccccca ctgcagctcc 1860
cacaaagcac tctaagggag atggggctga ggacagctca aaaaggagag gacaggcctg 1920
cagggcccac ttgcccagca ccaaggcgag ctccgggctc agaccaactc caaggtcagc 1980
tggccacagc ccaggcctcc tctgaagcag ggaccagagg gagccaggca gctgtgtaca 2040
gtgagaaggg atccggatgg gggagctctg tacagagggc tggtgattgt aaaaatttct 2100
tttgtaaagt agaagttggg ggtggggtgg gtgctggctg caaaaatttc tggcttctct 2160
taccctatt gccccccggca ataaattgtt tctatatgcc aaaaaaaaaa aaaaa 2215

<210> 22
<211> 721
<212> DNA
<213> Homo sapiens

<400> 22
tcgagccagc gtcgccgcga tggtgttgtt ggagagcgag cagttcctga cggagctgac 60
cagacttttc cagaagtgcc ggacgtcggg cagcgtctat atcaccttga agaagtatga 120
cggtcgaacc aaacccattc caaagaaagg tactgtggag ggctttgagc ccgcagacaa 180
caagtgtctg ttaagagcta ccgatgggaa gaagaagatc agcactgtgg tgagctccaa 240
ggaagtgaat aagtttcaga tggcttattc aaacctcctt agagctaaca tggatgggtt 300
gaagaagaga gacaaaaaga acaaaactaa gaagaccaaa gcagcagcag cagcagcagc 360
agcagcacct gccgcagcag caacagcagc aacaacagca gcaacaacag cagcaacagc 420
agcacagtaa agggcataca tttcctgctt tcaccaatta accactgaat tgctattttt 480
tccttttggc cagatagcta ggtttctggt tcccccacag taggtgtttt cacataagat 540
tagggtcctt ttggaaagaa tagttgcagt gtttatagga tagttgtggt aagaatctag 600
tttattttgc atttggctaa ttggtctgtg ctgcatggtt atatactcct ggattataga 660
ttaaaagtct ctgtagacat ctctgtgaag agcaagctat cattaaacat gtctgtttat 720
c 721

<210> 23
<211> 1685
<212> DNA
<213> Homo sapiens

<400> 23
gagtagctgc tttcggtccg ccggacacac cggacagata gacgtgcgga cggcccacca 60
ccccagcccg ccaactagtc agcctgcgcc tggcgcctcc cctctccagg tccatccgcc 120
atgtggcccc tgtggcgcct cgtgtctctg ctggccctga gccaggccct gccctttgag 180
cagagaggct tctgggactt caccctggac gatgggccat tcatgatgaa cgatgaggaa 240
gcttcgggcg ctgacacctc aggcgtcctg gacccggact ctgtcacacc cacctacagc 300
gccatgtgtc ctttcggctg ccactgccac ctgcgggtgg ttcagtgctc cgacctgggt 360
ctgaagtctg tgcccaaaga gatctccct gacaccacgc tgctggacct gcagaacaac 420
gacatctccg agctccgcaa ggatgacttc aagggtctcc agcacctcta cgccctcgtc 480
ctggtgaaca acaagatctc caagatccat gagaaggcct tcagcccact gcggaagctg 540
cagaagctct acatctccaa gaaccacctg gtggagatcc cgcccaacct acccagctcc 600
ctggtggagc tccgcatcca cgacaaccgc atccgcaagg tgcccaaggg agtgttcagc 660
gggctccgga acatgaactg catcgagatg ggcgggaacc cactggagaa cagtggcttt 720
gaacctggag ccttcgatgg cctgaagctc aactacctgc gcatctcaga ggccaagctg 780
actggcatcc ccaaagacct ccctgagacc ctgaatgaac tccacctaga ccacaacaaa 840
atccaggcca tcgaactgga ggacctgctt cgctactcca agctgtacag gctgggccta 900
ggccacaacc agatcaggat gatcgagaac gggagcctga gcttcctgcc caccctccgg 960
gagctccact tggacaacaa caagttggcc agggtgccct cagggctccc agacctcaag 1020
ctcctccagg tggtctatct gcactccaac aacatcacca aagtgggtgt caacgacttc 1080
tgtcccatgg gcttcggggt gaagcgggcc tactacaacg gcatcagcct cttcaacaac 1140
cccgtgccct actgggaggt gcagccggcc actttccgct gcgtcactga ccgcctggcc 1200
atccagtttg gcaactacaa aaagtagagg cagctgcagc caccgcgggg cctcagtggg 1260

ggtctctggg gaacacagcc agacatcctg atggggaggc agagccagga agctaagcca 1320
gggcccagct gcgtccaacc cagcccccca cctcaggtcc ctgaccccag ctcgatgccc 1380
catcaccgcc tctccctggc tcccaagggt gcaggtgggc gcaaggcccg gcccccatca 1440
catgttccct tggcctcaga gctgcccctg ctctcccacc acagccaccc agaggcaccc 1500
catgaagctt ttttctcgtt cactcccaaa cccaagtgtc caaagctcca gtcctaggag 1560
aacagtccct gggtcagcag ccaggaggcg gtccataaga atggggacag tgggctctgc 1620
cagggctgcc gcacctgtcc agaacaacat gttctgttcc tcctcctcat gcatttccag 1680
ccttg 1685

<210> 24
<211> 6428
<212> DNA
<213> Homo sapiens

<400> 24
gctagtggaa gttactgccg cgccaccgag tccggaccgg agactttggg gcctaactag 60
tgaatggtag tgtctagaaa gggtatgtcc cttcaagaga gaggtgccaa tgtccaaccg 120
gcctaataac aatccagggg ggtcactgcg acgttcacag aggaacactg ccggggccca 180
accacaagac gactcaatag gaggaagaag ctgcagttca tcatctgctg tgatagttcc 240
acaaccagag gatccagaca gagccaatac ttcagaaaga caaaaaacgg ggcaggtgcc 300
taagaaagac aattctcgag gagtgaagcg cagtgctagt ccagactaca acaggaccaa 360
ttctcctagc tctgcaaaaa aaccaaaagc acttcagcat actgaatctc cctcagaaac 420
aaataagcca catagtaagt caaagaagag acatttagac caggagcaac aactgaaatc 480
tgcacaatca ccatcaacaa gcaaggctca taccaggaag agtggggcca ctggcggttc 540
acggagtcag aaaagaaaaa ggacagagag ttcttgtgta aagagtggct ccgggtctga 600
atcaactggt gcagaagaga gatctgcgaa acctaccaag ctggcttcaa aatcagccac 660
ctcagccaaa gctgggtgta gcaccatcac tgattcttct tctgctgcct ctacttcctc 720
ctcgtcttct gctgtagcct cggcctcctc cactgtacca ccaggtgcca gagtgaaaca 780
aggaaaagat cagaacaagg ccaggcgttc ccgttcagcg tccagtccca gccccagaag 840
aagtagcagg gaaaaggaac agagtaaaac tggtggctct tcaaaatttg attgggctgc 900
tcgtttcagc cctaaagtta gccttcctaa aacaaaactg tctcttccag ggtcttctaa 960
gtcagagaca tcaaaacctg gaccttctgg attacaggcc aaattagcaa gtttaagaaa 1020
atctacgaag aaacgcagtg agtctccacc tgctgagctc cccagtttga ggcggagcac 1080
acgccaaaag accacgggct cctgtgctag taccagtcgg cgaggctctg gcctgggcaa 1140
aagaggagca gctgaagctc gtcgacagga gaaaatggca gaccctgaaa gcaaccagga 1200
ggcagtaaat tcttcagctg ctcggacaga tgaagctccc caaggagctg caggggctgt 1260
tggcatgacc acctctgggg agagtgaatc agatgattcc gagatgggac gtttgcaagc 1320
tttgttagag gcaaggggtc ttccccctca cctatttggt cctcttggtc ctcggatgtc 1380
acagcttttc catagaacaa ttggaagtgg agctagttct aaggcccagc agctactaca 1440
aggattgcaa gccagtgatg aaagtcaaca gcttcaggca gttattgaga tgtgtcagtt 1500
actggtcatg ggaaatgagg agacactggg agggtttcct gtcaagagtg ttgttccagc 1560
tttgattacg ttacttcaga tggagcacaa ttttgatatt atgaaccatg cttgtcgagc 1620
cttaacatac atgatggaag cacttcctcg atcttctgct gttgtagtag atgctattcc 1680
tgtcttttta gaaaagctgc aagttattca gtgtattgat gtggcagagc aggccttgac 1740
tgccttggag atgttgtcac ggagacatag taaagccatt ctacaggcgg gtggtttggc 1800
agactgcttg ctgtacctag aattcttcag cataaatgcc caaagaaatg cattagcaat 1860
tgcagctaat tgctgccaga gtatcacgcc agatgaattt cattttgtgg cagattcact 1920
cccattgcta acccaaaggc taacacatca ggataaaaag tcagtagaaa gcacttgcct 1980
ttgttttgca cgcctagtgg acaacttcca gcatgaggag aatttactcc agcaggttgc 2040
ttccaaagat ctgcttacaa atgttcaaca gctgttggta gtgactccac ccattttaag 2100
ttctgggatg tttataatgg tggttcgcat gttttctctg atgtgttcca actgtccaac 2160
tttagctgtt caacttatga aacaaaacat tgcagaaacg cttcactttc tcctgtgtgg 2220
tgcctccaat ggaagttgtc aggaacagat tgatcttgtt ccacgaagcc ctcaagagtt 2280
gtatgaactg acatctctga tttgtgaact tatgccatgt ttaccaaaag aaggcatttt 2340
tgcagttgat accatgttga agaagggaaa tgcacagaac acagatggtg cgatatggca 2400
gtggcgtgat gatcggggcc tctggcatcc atataacagg attgacagcc ggatcattga 2460
gcaaatcaat gaggacacgg gaacagcacg tgccattcag agaaaaccta acccgttagc 2520
caatagtaac actagtggat attcagagtc aaagaaggat gatgctcgag cacagcttat 2580
gaaagaggat ccggaactgg ctaagtcttt tattaagaca ttatttggtg ttctttatga 2640
agtgtatagt tcctcagcag gacctgcggt cagacataag tgccttagag caattcttag 2700
gataatttat tttgcggatg ctgaacttct gaaggatgtt ctgaaaaatc atgctgtttc 2760
aagtcacatt gcttccatgc tgtcaagcca agacctgaag atagtagtgg gagcacttca 2820
gatggcagaa attttaatgc agaagttacc tgatattttt agtgtttact tcagaagaga 2880

aggtgtaatg catcaagtaa aacacttagc agaatcagag tctttgttga caagtccacc 2940
aaaggcatgt acgaatggat cgggatccat gggatccaca acttcagtca gcagtgggac 3000
agccacagct gccactcatg ctgcagctga cttgggatca cccagcttgc agcacagcag 3060
ggatgattct ttagatctca gccctcaagg tcgattaagt gatgttctaa agagaaaacg 3120
actgccaaaa cgagggccaa gaaggccaaa gtactcacct ccaagagatg atgacaaagt 3180
agacaatcaa gctaaaagcc ccaccactac tcagtcacct aaatcttctt tcctggcaag 3240
cttgaatcca aaaacatggg gaaggttaag tacacagtcc aacagcaaca acattgagcc 3300
agcacggact gcgggaggta gtggccttgc cagggctgcc tcaaaggata ccatctccaa 3360
taatagagaa aaaattaaag gttggattaa ggagcaggca cataaatttg tagaacgtta 3420
tttcagttct gagaatatgg atggaagcaa ccctgcattg aatgtccttc agagactttg 3480
tgctgcaacc gaacaactca acctccaggt ggatggtgga gctgagtgcc ttgtagaaat 3540
ccgtagcata gtctcagagt cagatgtttc atcatttgaa atccaacata gtggatttgt 3600
gaagcagctg ttgctttatt tgacatctaa aagtgaaaag gatgctgtga gcagagagat 3660
cagattaaag cgatttcttc atgtattttt ttcttctcca cttcctggag aagagcccat 3720
tggaagagtg gaaccagtgg gtaatgcacc tttgttggca ttagttcaca agatgaacaa 3780
ctgcctcagc cagatggaac aatttccagt caaagtacat gatttcccta gtggaaatgg 3840
gacaggaggc agcttttctc tcaacagagg atcacaggct ttaaaatttt tcaacacaca 3900
tcaattaaaa tgccagttac aaaggcatcc agactgtgca aatgtgaagc agtggaaggg 3960
tggacctgtc aagattgacc ctctggcttt ggtacaagcc atcgagagat accttgtagt 4020
tagagggtat ggaagagtaa gagaagatga tgaagacagc gatgacgatg gatcagatga 4080
ggaaatagat gagtctctgg ctgctcagtt cctaaattca ggaaatgtaa gacacaggct 4140
gcagttttat attggagaac atttgctgcc gtataacatg actgtgtatc aggcagtacg 4200
gcagtttagt atacaggctg aagatgaaag agaatccaca gatgatgaga gcaatcctct 4260
aggcagagct ggtatttgga caaagactca tacaatatgg tataaacctg tgagagagga 4320
tgaagaaagt aataaagatt gtgttggtgg taaaagagga agagcccaaa cagctccaac 4380
gaaaacttcc cctagaaatg caaaaaagca tgatgagtta tggcacgatg gagtgtgccc 4440
atcagtatca aatcctttag aagtttacct cattcccaca ccacctgaaa atataacatt 4500
tgaagacccg tcattagatg tgatccttct tttaagagtt ttacatgcta tcagtcgata 4560
ctggtattac ttgtatgata atgcaatgtg caaggaaatt attccaacta gtgaatttat 4620
taacagtaag ttaacagcaa aagcaaatag gcaacttcaa gatcctttag taatcatgac 4680
aggaaacatc ccaacatggc ttactgagct aggaaaaacc tgcccatttt tctttccttt 4740
tgatacccgg caaatgcttt tttatgtaac tgcatttgat cgggaccgag caatgcaaag 4800
attacttgat accaacccag aaatcaacca gtctgattct caagatagca gagttgcacc 4860
tagattggat agaaaaaaac gtactgtgaa ccgagaggag ctgctgaaac aggcggagtc 4920
tgtgatgcag gacctcggca gctcacgggc catgttagaa atccagtatg aaaatgaggt 4980
tggtacaggt cttgggccta cactggagtt ttatgcgctt gtatctcagg aactacagag 5040
agctgacttg ggtctttgga gaggtgaaga agtaactctt agcaatccaa aagggagcca 5100
agaagggacc aagtatattc aaaacctcca gggcctgttt gcgcttccct ttggtaggac 5160
agcaaagcca gctcatatcg caaaggttaa gatgaagttt cgcttcttag gaaaattaat 5220
ggccaaggct atcatggatt tcagattggt ggaccttccc cttggcttac ccttttataa 5280
atggatgcta cggcaagaaa cttcactgac atcacacgat ttgtttgaca tcgacccagt 5340
tgtagccaga tcagtttatc acctagaaga cattgtcaga cagaagaaaa gacttgaaca 5400
agataaatcc cagaccaaag agagtctaca gtatgcatta gaaaccttga ctatgaatgg 5460
ctgctcagtt gaagatctag gactggattt cactctgcca gggtttccca atatcgaact 5520
gaagaaagga gggaaggata taccagtcac tatccacaat ttagaggagt atctaagact 5580
ggttatattc tgggcactaa atgaaggcgt ttctaggcaa tttgattcgt tcagagatgg 5640
atttgaatca gtcttcccac tcagtcatct tcagtacttc tacccggagg aactggatca 5700
gctcctttgt ggcagtaaag cagacacttg ggatgcaaag acactgatgg aatgctgtag 5760
gcctgatcat ggttatactc atgacagtcg ggctgtgaag tttttgtttg agattctcag 5820
tagttttgat aatgagcagc agaggttatt tctccagttt gtgactggta gcccaagatt 5880
gcctgttgga ggattccgga gtttgaatcc acctttgaca attgtccgaa agacgtttga 5940
atcaacagaa aacccagatg acttcttgcc ctctgtaatg acttgtgtga actatcttaa 6000
gttgccggac tattcaagca ttgagataat gcgtgaaaaa ctgttgatag cagcaagaga 6060
agggcagcag tcgttccatc tttcctgatt atagcaagaa atgcagtgtc tgcctgttac 6120
agcaaaagaa acaaatcatg atttcttttc taatgttatc acctgagtca aggaaacatg 6180
ttacgccttc ttgttgtagg aaaaacggct tgcagattat aaagagacat ttggttgata 6240
ttcattaatg gccccatgga cttaaagtga tcaggcccta aaacgttgtt gtgatgaggt 6300
ttctttagca agttcttgtt taaattatca tttatttgat gagtgaagtt tttaacatgc 6360
tttgctgtgt gaaatttaaa aaagggatgt ttttccaggc tggaacaata aatgtggctg 6420
tgcagttt 6428

<210> 25
<211> 1370

<212> DNA
<213> Homo sapiens

<400> 25
gcacgaggcc agccttacgg gcccgaaccc tcgtgtgaag ggtgcagtac ctaagccgga 60
gcggggtaga ggcgggccgg cacccccttc tgacctccag tgccgccggc ctcaagatca 120
gacatggccc agaacttgaa ggacttggcg ggacggctgc ccgccgggcc ccggggcatg 180
ggcacggccc tgaagctgtt gctgggggcc ggcgccgtgg cctacggtgt gcgcgaatct 240
gtgttcaccg tggaaggcgg gcacagagcc atcttcttca atcggatcgg tggagtgcag 300
caggacacta tcctggccga gggccttcac ttcaggatcc cttggttcca gtaccccatt 360
atctatgaca ttcgggccag acctcgaaaa atctcctccc ctacaggctc caaagaccta 420
cagatggtga atatctccct gcgagtgttg tctcgaccca atgctcagga gcttcctagc 480
atgtaccagc gcctagggct ggactacgag gaacgagtgt tgccgtccat tgttaatgag 540
gtgctcaaga gtgtggtggc caagttcaat gcctcacagc tgatcaccca gcgggcccag 600
gtatccctgt tgatccgccg ggagctgaca gagagggcca aggacttcag cctcatcctg 660
gatgatgtgg ccatcacaga gctgagcttt agccgagagt acacagctgc tgtagaagcc 720
aaacaagtgg cccagcagga ggcccagcgg gcccaattct tggtagaaaa agcaaagcag 780
gaacagcggc agaaaattgt gcaggccgag ggtgaggccg aggctgccaa gatgcttgga 840
gaagcactga gcaagaaccc tggctacatc aaacttcgca agattcgagc agcccagaat 900
atctccaaga cgatcgccac atcacagaat cgtatctatc tcacagctga caaccttgtg 960
ctgaacctac aggatgaaag tttcaccagg ggaagtgaca gcctcatcaa gggtaagaaa 1020
tgagcctagt caccaagaac tccacccccа gaggaagtgg atctgcttct ccagtttttg 1080
aggagccagc caggggtcca gcacagccct accccgcccc agtatcatgc gatggtcccc 1140
cacaccggtt ccctgaaccc ctcttggatt aaggaagact gaagactagc ccctttctctg 1200
gggaattact ttcctcctcc ctgtgttaac tggggctgtt ggggacagtg cgtgatttct 1260
cagtgatttc ctacagtgtt gttccctccc tcaaggctgg gaggagataa acaccaaccc 1320
aggaattctc aataaatttt tattacttaa cctgaaaaaa aaaaaaaaaa 1370

<210> 26
<211> 642
<212> DNA
<213> Homo sapiens

<400> 26
ccgcgcgtca cgtgacccca gcgcctactt gggctgagga gccgccgcgt cccctcgccg 60
agtcccctcg ccagattccc tccgtcgccg ccaagatgat gtgcggggcg ccctccgcca 120
cgcagccggc caccgccgag acccagcaca tcgccgacca ggtgaggtcc cagcttgaag 180
agaaagaaaa caagaagttc cctgtgttta aggccgtgtc attcaagagc caggtggtcg 240
cggggacaaa ctacttcatc aaggtgcacg tcggcgacga ggacttcgta cacctgcgag 300
tgttccaatc tctccctcat gaaaacaagc ccttgacctt atctaactac cagaccaaca 360
aagccaagca tgatgagctg acctatttct gatcctgact ttggacaagg cccttcagcc 420
agaagactga caaagtcatc ctccgtctac cagagcgtgc acttgtgatc ctaaaataag 480
cttcatctcc gctgtgccct tggggtggaa ggggcaggat tctgcagctg cttttgcatt 540
tctcttccta aatttcattg tgttgatttc tttccttccc aataggtgat cttaattact 600
ttcagaatat tttcaaaata gatatatttt taaaatcctt ac 642