BLADDER CANCER BIOMARKERS AND USES THEREOF

This application claims the benefit of US Provisional Application 60/676,921 filed May 2, 2005 and US Provisional Application 60/729,056 filed October 21, 2005, which are here-in incorporated by reference.

TABLES

This application includes a compact disc in duplicate (2 compact discs: Tables - Copy 1 and Tables - Copy 2), which are hereby incorporated by reference in their entirety. Each compact disc is identical and contains the following files (corresponding to Tables 1 – 7 and 11-12):

TABLE	DESCRIPTION	SIZE	CREATED	Text File Name
1	Bladder cancer biomarkers	90 KB	10/17 /2005	TABLE1.TXT
2	Bladder cancer biomarkers disclosed in PCT Application Number PCT/US04/020836	710 KB	10/17 /2005	TABLE2.TXT
3	Bladder cancer biomarker products of Table 1	121KB	10/17 /2005	TABLE3.TXT
4	Early Stage Bladder cancer biomarkers	176 KB	10/17/2005	TABLE4.TXT
5	Early Stage Bladder cancer biomarkers disclosed in PCT Application Number PCT/US04/020836	746 KB	10/17/2005	TABLE5.TXT

6	Bladder cancer biomarker products of Table 4	144 KB	10/17/2005	TABLE6.TXT
7	Biomarkers which differentiate bladder cancer from normal, whether bladder cancer be early stage or not.	70 KB	10/17/2005	TABLE7.TXT
11	Biomarker which distinguish as between bladder cancer as compared with either testicular or renal cell carcinoma.	75 KB	10/17/2005	TABLE11.TXT
12	Bladder cancer biomarker products of Table 11	72	10/17/2005	TABLE12.TXT

1. Field of the Invention

The invention encompasses the identification and selection of novel bladder cancer biomarkers including biomarkers for early stage bladder cancer as well as the identification and selection of novel biomarker combinations which are differentially expressed in individuals with bladder cancer or superficial bladder cancer as compared with individuals without bladder cancer. The measurement of expression of the products of the biomarkers and combinations of biomarkers of the invention demonstrates particular advantage in diagnosing individuals as having bladder cancer. Further encompassed are polynucleotides and proteins which specifically and/or selectively hybridize/bind to the products of the biomarkers of the invention. The invention also provides for methods of using the products of the biomarkers of the invention in the identification of compounds that bind and/or modulate the activity of the biomarker genes of the invention. The compounds identified via such methods are useful for the development of assays to study bladder cancer and bladder cancer progression. Further, the compounds identified via such methods are useful as lead compounds in the development of prophylactic and therapeutic compositions

for the prevention, treatment, management and/or amelioration of bladder cancer or a symptom thereof.

2. Background of the Invention

Bladder cancer is the fourth most commonly diagnosed cancer in men and the eighth most commonly diagnosed cancer in women (Greenlee et al. *Cancer Statistics,* 2001). Exposure to cigarette smoke, aromatic amines, coal combustion by-products, chlorinated compounds and certain aldehydes have been shown to increase a person's risk of getting bladder cancer. Bladder cancer itself has a low impact on mortality, however, metastasis of the cancer to other sites increases the risk of mortality substantially. As such, a diagnostic method which identified bladder cancer would be a benefit. In particular, a diagnostic method which allowed early detection and intervention would be of great benefit.

Currently used screening methods of bladder cancer include the use of cystoscopies; bladder wash and/or a urinary cytologic examination. Each of these methods allows a physician to look at the bladder and/or cells of the bladder to detect tumours or other bladder cell abnormalities. These methods however are expensive and are therefore not practical as screening methods for persons who are not suspected of having an increased risk of bladder cancer. Another common screening technique is hematuria analysis (i.e. looking for blood in urine). Although hematuria is the most common presenting symptom of bladder cancer, it can also indicate many other potential diagnosis (i.e. low specificity). For example, prospective analysis of patients attending a hematuria clinic in the United Kingdom only 19.2% of patients having gross hematuria were determined to have bladder cancer as determined by cytoscopy. In addition, these methods become even less specific and sensitive for those patients with earlier stages of bladder cancer.

Thus there is a need in the art for a method for screening for bladder cancer which is relatively inexpensive and sufficiently sensitive and specific so as to improve the detection of those patients at risk for bladder cancer. There is also a need for a method which has a greater ability to screen for bladder cancer at relatively early stages so as to allow for early intervention and more treatment options for those patients found to have bladder cancer.

3. Summary of the Invention

The invention encompasses the identification and selection of novel bladder cancer biomarkers and the identification and selection of superficial (early stage) bladder cancer biomarkers as well as identification of combinations of these biomarkers so as to provide a simple and relatively inexpensive test to provide improved screening for bladder cancer. The methods disclosed herein along with the products of the biomarkers and combinations of biomarkers identified using these methods demonstrate particular advantage in screening patients to identify those with an increased risk of bladder cancer. As would be understood, in order to measure the products of biomarkers of the invention, polynucleotides and proteins which specifically and/or selectively hybridize/bind to the products of the biomarkers identified, and derivatives thereof, of the invention are also encompassed within the scope of the invention as are kits containing said polynucleotides and proteins for use in diagnosing individuals as having bladder cancer. Further encompassed by the invention is the use of the polynucleotides and proteins which specifically and/or selectively hybridize to the product of the biomarkers of the invention to monitor bladder cancer progression in an individual and to monitor the efficacy of therapeutic regimens. The invention also provides for the identification of methods of using the products of the biomarkers of the invention in the identification of novel therapeutic targets for bladder cancer.

3.1 Definitions

The following definitions are provided for specific terms which are used in the following written description.

As used herein, the term "3' end" refers to the end of an mRNA up to the last 1000 nucleotides or 1/3 of the mRNA, where the 3' terminal nucleotide is that terminal nucleotide of the coding or untranslated region that adjoins the poly-A tail, if one is present. That is, the 3' end of an mRNA does not include the poly-A tail, if one is present. The "3' region" of a gene refers to a polynucleotide (double-stranded or single-stranded) located within or at the 3' end of a gene, and includes, but is not limited to, the 3' untranslated region, if that is present, and the 3' protein coding region of a gene. The 3' region is not shorter than 8 nucleotides in length and not longer than 1000 nucleotides in length. Other possible lengths

of the 3' region include but are not limited to 10, 20, 25, 50, 100, 200, 400, and 500 nucleotides.

As used herein, the term "5' end" refers to the end of an mRNA up to the first 1000 nucleotides or 1/3 of the mRNA (where the full length of the mRNA does not include the poly A tail), starting at the first nucleotide of the mRNA. The "5' region" of a gene refers to a polynucleotide (double-stranded or single-stranded) located within or at the 5' end of a gene, and includes, but is not limited to, the 5' untranslated region, if that is present, and the 5' protein coding region of a gene. The 5' region is not shorter than 8 nucleotides in length and not longer than 1000 nucleotides in length. Other possible lengths of the 5' region include but are not limited to 10, 20, 25, 50, 100, 200, 400, and 500 nucleotides.

As used herein, the term "amplified", when applied to a nucleic acid sequence, refers to a process whereby one or more copies of a particular nucleic acid sequence is generated from a template nucleic acid, preferably by the method of polymerase chain reaction (Mullis and Faloona, 1987, Methods Enzymol., 155:335). "Polymerase chain reaction" or "PCR" refers to an in vitro method for amplifying a specific nucleic acid template sequence. In some embodiments, the PCR reaction involves a repetitive series of temperature cycles and is typically performed in a volume of 50-100 μl. The reaction mix comprises dNTPs (each of the four deoxynucleotides dATP, dCTP, dGTP, and dTTP), primers, buffers, DNA polymerase, and nucleic acid template. The PCR reaction can comprise providing a set of polynucleotide primers wherein a first primer contains a sequence complementary to a region in one strand of the nucleic acid template sequence and primes the synthesis of a complementary DNA strand, and a second primer contains a sequence complementary to a region in a second strand of the target nucleic acid sequence and primes the synthesis of a complementary DNA strand, and amplifying the nucleic acid template sequence employing a nucleic acid polymerase as a template-dependent polymerizing agent under conditions which are permissive for PCR cycling steps of (i) annealing of primers required for amplification to a target nucleic acid sequence contained within the template sequence, (ii) extending the primers wherein the nucleic acid polymerase synthesizes a primer extension product. "A set of polynucleotide primers" or "a set of PCR primers" can comprise two, three, four or more primers. In one embodiment, an exo- Pfu DNA polymerase is used to amplify a nucleic acid template in PCR reaction. Other methods of

amplification include, but are not limited to, ligase chain reaction (LCR), polynucleotide-specific based amplification (NSBA), or any other method known in the art.

As used herein, the "amino terminal" region of a polypeptide refers to the polypeptide sequences encoded by polynucleotide sequences (double-stranded or single-stranded) located within or at the 5' end of an mRNA As used herein, the "amino terminal" region refers to the amino terminal end of a polypeptide up to the first 300 amino acids or 1/3 of the polypeptide, starting at the first amino acid of the polypeptide. The "amino terminal" region of a polypeptide is not shorter than 3 amino acids in length and not longer than 350 amino acids in length. Other possible lengths of the "amino terminal" region of a polypeptide include but are not limited to 5, 10, 20, 25, 50, 100 and 200 amino acids.

As used herein, the term "analog" in the context of proteinaceous agent (e.g., proteins, polypeptides, peptides, and antibodies) refers to a proteinaceous agent that possesses a similar or identical function as a second proteinaceous agent but does not necessarily comprise a similar or identical amino acid sequence of the second proteinaceous agent, or possess a similar or identical structure of the second proteinaceous agent. A proteinaceous agent that has a similar amino acid sequence refers to a second proteinaceous agent that satisfies at least one of the following: (a) a proteinaceous agent having an amino acid sequence that is at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95% or at least 99% identical to the amino acid sequence of a second proteinaceous agent; (b) a proteinaceous agent encoded by a nucleotide sequence that hybridizes under stringent conditions to a nucleotide sequence encoding a second proteinaceous agent of at least 5 contiguous amino acid residues, at least 10 contiguous amino acid residues, at least 15 contiguous amino acid residues, at least 20 contiguous amino acid residues, at least 25 contiguous amino acid residues, at least 40 contiguous amino acid residues, at least 50 contiguous amino acid residues, at least 60 contiguous amino residues, at least 70 contiguous amino acid residues, at least 80 contiguous amino acid residues, at least 90 contiguous amino acid residues, at least 100 contiguous amino acid residues, at least 125 contiguous amino acid residues, or at least 150 contiguous amino acid residues; and (c) a proteinaceous agent encoded by a nucleotide sequence that is at least 30%, at least 35%, at

least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95% or at least 99% identical to the nucleotide sequence encoding a second proteinaceous agent. A proteinaceous agent with similar structure to a second proteinaceous agent refers to a proteinaceous agent that has a similar secondary, tertiary or quaternary structure to the second proteinaceous agent. The structure of a proteinaceous agent can be determined by methods known to those skilled in the art, including but not limited to, peptide sequencing, X-ray crystallography, nuclear magnetic resonance, circular dichroism, and crystallographic electron microscopy.

To determine the percent identity of two amino acid sequences or of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino acid or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences (i.e., % identity=number of identical overlapping positions/total number of positions times 100%). In one embodiment, the two sequences are the same length.

The determination of percent identity between two sequences can also be accomplished using a mathematical algorithm. A preferred, non-limiting example of a mathematical algorithm utilised for the comparison of two sequences is the algorithm of Karlin and Altschul, 1990, Proc. Natl. Acad. Sci. U.S.A. 87:2264-2268, modified as in Karlin and Altschul, 1993, Proc. Natl. Acad. Sci. U.S.A. 90:5873-5877. Such an algorithm is incorporated into the NBLAST and XBLAST programs of Altschul et al., 1990, J. Mol. Biol. 215:403. BLAST nucleotide searches can be performed with the NBLAST nucleotide program parameters set, e.g., for score=100, wordlength=12 to obtain nucleotide sequences homologous to a nucleic acid molecules of the present invention. BLAST protein searches can be performed with the XBLAST program parameters set, e.g., to score-50, wordlength=3 to obtain amino acid sequences homologous to a protein molecule of the present invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilised as

described in Altschul et al., 1997, Nucleic Acids Res. 25:3389-3402. Alternatively, PSI-BLAST can be used to perform an iterated search which detects distant relationships between molecules (Id.). When utilising BLAST, Gapped BLAST, and PSI-Blast programs, the default parameters of the respective programs (e.g., of XBLAST and NBLAST) can be used (see, e.g., the NCBI website). Another preferred, non-limiting example of a mathematical algorithm utilised for the comparison of sequences is the algorithm of Myers and Miller, 1988, CABIOS 4:11-17. Such an algorithm is incorporated in the ALIGN program (version 2.0) which is part of the GCG sequence alignment software package. When utilising the ALIGN program for comparing amino acid sequences, a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 can be used. The percent identity between two sequences can be determined using techniques similar to those described above, with or without allowing gaps. In calculating percent identity, typically only exact matches are counted.

As used herein, the term "analog" in the context of a non-proteinaceous analog refers to a second organic or inorganic molecule which possess a similar or identical function as a first organic or inorganic molecule and is structurally similar to the first organic or inorganic molecule.

As used herein, the term "biomarker" refers to a gene that is differentially expressed in individuals having bladder cancer or a stage of bladder cancer as compared with those not having bladder cancer, or said stage of bladder cancer (although individuals may have other disease(s)) and can include a gene that is differentially expressed in individuals having superficial bladder cancer as compared with those not having bladder cancer.

The term “biomarker specific primers” as used herein refers to a set of primers which can produce double stranded DNA complementary to a portion of one or more RNA products of the biomarker of the invention. For example, the primers can include a first primer which is a sequence that can selectively hybridize to RNA, cDNA or EST complementary to a biomarker of the invention to create an extension product and a second primer capable of selectively hybridizing to the extension product, which are used to produce double stranded DNA complementary to a biomarker of the invention.

The term "biomarker specific probe" as used herein refers to a probe selectively and specifically hybridizes to RNA products of a unique biomarker. In one embodiment a biomarker specific probe can be a probe having a fluorophore and a quencher, for example a TaqMan® probe or a Molecular Beacon® probe. In another embodiment a biomarker specific probe is a probe which is attached to an array and selectively and specifically hybridizes to one or more RNA products (or cDNA, EST or PCR products corresponding to said RNA products) of a unique biomarker. A biomarker specific probe can include oligonucleotide probes and can also include longer probes (e.g. 100, 150, 200, 250, 300, 350, 400, 450, 500 Nucleotides etc.).

As used herein, the term "bladder cancer" refers to a disease in which the cells lining the urinary bladder lose the ability to regulate their growth resulting in a mass of cells that form a tumor. As used herein, the term "early bladder cancer" or "superficial bladder cancer" refer to non invasive bladder tumours (e.g. type Ta or Tia as determined in accordance with the AJCC guidelines) (Herr et al. 2001)

As used herein, the term "blood nucleic acid sample" refers to nucleic acids derived from blood and can include nucleic acids derived from whole blood, centrifuged lysed blood, serum free whole blood or fractionated blood including peripheral blood leukocytes (PBLs) or other fractions of blood as described herein. By whole blood is meant unfractionated blood and includes a drop of blood wherein a drop of blood includes volumes of 5µl, 10µl, 15µl, 20µl, 25µl, 30µl. By centrifuged lysed blood or 'lysed blood' is meant unfractionated whole blood that is mixed with lysis buffer and centrifuged as described herein (see Example 2). By serum free blood is meant unfractionated whole blood wherein the serum or plasma is removed by centrifugation as described herein (see Example 2). Preferably, a blood nucleic acid sample is whole blood or centrifuged lysed blood and is total RNA, mRNA or is a nucleic acid corresponding to mRNA, for example, cDNA isolated from said blood. A nucleic acid sample can also include a PCR product derived from total RNA, mRNA or cDNA.

As used herein, the "carboxy terminal" region of a polypeptide refers to the polypeptide sequences encoded by polynucleotide sequences (double-stranded or single-stranded) located within or at the 3' end of an mRNA, and. As used herein, the "carboxy

terminal" region refers to the carboxy terminal end of a polypeptide up to 300 amino acids or 1/3 of the polypeptide from the last amino acid of the polypeptide. The "3' end" does not include the polyA tail, if one is present. The "carboxy terminal" region of a polypeptide is not shorter than 3 amino acids in length and not longer than 350 amino acids in length. Other possible lengths of the "carboxy terminal" region of a polypeptide include, but are not limited to, 5, 10, 20, 25, 50, 100 and 200 amino acids.

As used herein, the term "tissue sample" refers to cells derived from a tissue which is not blood and can include bladder tissue, brain tissue, heart tissue, lung tissue, lymph node, and the like. A "tissue nucleic acid sample" is total RNA, mRNA or is a nucleic acid corresponding to RNA, for example, cDNA. A tissue nucleic acid sample can also include a PCR product derived from total RNA, mRNA or cDNA.

As used herein, the term "combination of the biomarkers of the invention" refers to any one or more biomarkers as disclosed in Table 1, Table 3, Table 4, Table 6, Table 7 and Table 10, Table 11 or Table 12 . Combination of biomarkers of the invention also includes and any combinations of any two or more biomarkers as disclosed in Table 1 and Table 2 so long as at least one of the biomarker of said combination is from Table 1. Combination of biomarkers of the invention also includes and any combinations of any two or more biomarkers as disclosed in Table 4 and Table 5 so long as at least one of the biomarkers of said combination is from Table 4. Combination of biomarkers of the invention also includes and any combinations of any two or more biomarkers as disclosed in Table 11 and Table 2 so long as at least one of the biomarker of said combination is from Table 11.

As used herein, the terms "compound" and "agent" are used interchangeably.

As used herein, "consisting essentially of" refers to the maximum number of biomarkers that are useful to diagnose or screen for bladder cancer and/or diagnose or screen for superficial bladder cancer. In one embodiment, a biomarker for the diagnosis of bladder cancer and/or superficial bladder cancer consists essentially of at least any of up to 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 100, 125, 150, 175, 200, 225, 250 or all of the biomarkers of Table 1. In another embodiment, a biomarker for the diagnosis of bladder cancer consists essentially of at least any of up to 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 100, 125, 150, 175,

200 or all of the biomarkers of Table 7. In another embodiment, a biomarker for the diagnosis of bladder cancer consists essentially of at least any of up to 1, 2, 3, 4, 5, 6, 7, 8, 9 or all of the biomarkers of Table 10. In another embodiment, a biomarker for the diagnosis of bladder cancer consists essentially of at least any of up to 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 13, 15, 20, 25, 30, 35, 40, 45, 50, 100, 125, 150, 175, 200 of the biomarkers of Table 1 in combination with at least any of up to 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10, 15, 20, 30, 40, 50, 100, 200, 300, 400, 500, 1000, 1500, 2000, 2500, 3000, 3500, 4000 or all of the biomarkers as disclosed in PCT Application Number PCT/US04/020836 and which are provided in Table 2. In another embodiment, a biomarker for the diagnosis of superficial or early stage bladder cancer consists essentially of at least any of up to 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500 or all of the biomarkers of Table 4. In another embodiment, a biomarker for the diagnosis of superficial bladder cancer consists essentially of at least any of up to 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500 of the biomarkers of Table 4 in combination with at least any of up to 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10, 15, 20, 30, 40, 50 100, 150, 200, 250, 300, 350, 400, 450, 500, 1000, 1500, 2000, 2500 or all of the biomarkers as disclosed in PCT Application Number PCT/US04/020836 and which are provided in Table 5. In another embodiment, a biomarker for the diagnosis of bladder cancer and/or superficial bladder cancer consists essentially of at least any of up to 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 100, 125, 150, 175, 200, 225, 250 or all of the biomarkers of Table 11. In another embodiment, a biomarker for the diagnosis of bladder cancer and/or superficial bladder cancer consists essentially of at least any of up to 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 100, 125, 150, 175, 200, 225, 250 or all of the biomarkers of Table 12.

As used herein, the term "control" or "control sample" in the context of this invention refers to one or more tissue nucleic acid samples and/or a blood nucleic acid samples and/or one or more individuals who are classified as having prostate cancer, having one or more stages of bladder cancer including superficial bladder cancer; not having bladder cancer or not having a stage of bladder cancer (e.g. superficial bladder cancer); using those techniques

known to a person skilled in the art (for example **NMP22®BladderChek®** to detect elevated levels of a nuclear matrix protein (called NMP22®) in the urine, intravenous pyloegram (IVP) imaging, CT scan, MRI scan, bone scan or ultrasound, and the like). The term control or control sample can also refer to the compilation of data derived from samples of one or more individuals who have been diagnosed as normal (not having bladder cancer), having bladder cancer, or having a stage of bladder cancer (e.g. superficial bladder cancer). As would be understood by a person skilled in the art – the term control is used in the context of the experiment and will depend upon the desired comparisons. As used herein, the term "control" in the context of screening for a prophylactic or therapeutic agent refers to a standard or reference for an assay or methodology to which other conditions can be compared.

As used herein, the term "data" or "biomarker data" generally refers to data reflective of the abundance or level of a product of a biomarker including either or both of RNA and protein encoded by the biomarker.

As used herein, the term "derivative" in the context of proteinaceous agent (e.g., proteins, polypeptides, peptides, and antibodies) refers to a proteinaceous agent that comprises an amino acid sequence which has been altered by the introduction of one or more amino acid residue substitutions, deletions, and/or additions. The term "derivative" as used herein also refers to a proteinaceous agent which has been modified, i.e., by the covalent attachment of any type of molecule to the proteinaceous agent. For example, but not by way of limitation, an antibody may be modified, e.g., by glycosylation, acetylation, pegylation, phosphorylation, amidation, derivatization by known protecting/blocking groups, proteolytic cleavage, linkage to a cellular ligand or other protein, etc. A derivative of a proteinaceous agent may be produced by chemical modifications using techniques known to those of skill in the art, including, but not limited to specific chemical cleavage, acetylation, formylation, metabolic synthesis of tunicamycin, etc. Further, a derivative of a proteinaceous agent may contain one or more non-classical amino acids. A derivative of a proteinaceous agent possesses a similar or identical function as the proteinaceous agent from which it was derived.

As used herein, the term "derivative" in the context of a non-proteinaceous derivative refers to a second organic or inorganic molecule that is formed based upon the structure of a first organic or inorganic molecule. A derivative of an organic molecule includes, but is not limited to, a molecule modified, e.g., by the addition or deletion of a hydroxyl, methyl, ethyl, carboxyl or amine group. An organic molecule may also be esterified, alkylated and/or phosphorylated.

As used herein "diagnosis of bladder cancer" or "bladder cancer diagnosis" "screening for bladder cancer" "screening for a stage of bladder cancer" according to one aspect of the invention refers to a process of determining the likelihood of a person having bladder cancer and/or having superficial bladder cancer. The terms refer to both traditional medical diagnostic techniques for diagnosing or screening for bladder cancer or a stage of bladder cancer, as well as diagnosis and/or screening methods as encompassed by the invention. Traditional medical diagnostic techniques for diagnosing bladder cancer include physical exam and history, medical evaluation including urine test, and appropriate laboratory tests which can include **NMP22®BladderChek®** to detect elevated levels of a nuclear matrix protein (called NMP22®) in the urine, intravenous pyloegram (IVP) imaging, CT scan, MRI scan, bone scan or ultrasound, and the like. In a specific embodiment, "diagnosis of prostate cancer" refers to a determination as between two options: e.g. that an individual has bladder cancer or a specific stage of bladder cancer or that an individual does not have bladder cancer or a specific stage of bladder cancer. In a specific embodiment, the term "diagnosis or screening for bladder cancer" refers to a determination as between two options, e.g. for "diagnosis of bladder cancer", a determination of the likelihood that an individual has bladder cancer or that an individual does not have bladder cancer. For "diagnosis of superficial bladder cancer", a determination of the likelihood that an individual has superficial bladder cancer or does not have bladder cancer. In another specific embodiment, "diagnosis" refers to a determination as between three options where one option is that it cannot be determined with sufficient degree of certainty whether an individual has bladder cancer or does not have bladder cancer (e.g. a determination that a result is untenable).

In another embodiment, "diagnosis of or screening for bladder cancer" can include an option that it cannot be determined with sufficient degree of certainty as to whether an individual

can be characterized as having bladder cancer or a specific stage of bladder cancer or does not have bladder cancer or a specific stage of bladder cancer. As would be understood by a person skilled in the art, in this context a "sufficient degree of certainty" depends upon the medical requirements for both the sensitivity and specificity of the diagnosis. More particularly the sufficient degree of certainty includes greater than 50% sensitivity and/or specificity, greater than 60% sensitivity and/or specificity, greater than 70% sensitivity and/or specificity, greater than 80% sensitivity and/or specificity, greater than 90% sensitivity and/or specificity and 100% sensitivity and/or specificity.

As used herein, the term "differential expression" refers to a difference in the level of expression of the RNA and/or protein products of one or more biomarkers of the invention, as measured by the amount or level of RNA or protein. In reference to RNA, it can include difference in the level of expression of mRNA, and/or one or more spliced variants of mRNA of the biomarker in one sample as compared with the level of expression of the same one or more biomarkers of the invention as measured by the amount or level of RNA, including mRNA and/or one or more spliced variants of mRNA in a second sample. "Differentially expressed" or "differential expression" can also include a measurement of the protein, or one or more protein variants encoded by the biomarker of the invention in a sample or population of samples as compared with the amount or level of protein expression, including one or more protein variants of the biomarker or biomarkers of the invention. Differential expression can be determined as described herein and as would be understood by a person skilled in the art. The term "differentially expressed" or "changes in the level of expression" refers to an increase or decrease in the measurable expression level of a given product of the biomarker as measured by the amount of RNA and/or the amount of protein in a sample as compared with the measurable expression level of a given product of the biomarker in a second sample. The first sample and second sample need not be from different patients, but can be samples from the same patient taken at different time points. The term "differentially expressed" or "changes in the level of expression" can also refer to an increase or decrease in the measurable expression level of a given biomarker in a population of samples as compared with the measurable expression level of a biomarker in a second population of samples. As used herein, "differentially expressed" when referring to a single sample can be measured using the ratio of the level of expression of a given biomarker in said

sample as compared with the mean expression level of the given biomarker of a control population wherein the ratio is not equal to 1.0. Differentially expressed can also be used to include comparing a first population of samples as compared with a second population of samples or a single sample to a population of samples using either a ratio of the level of expression or using p-value. When using p-value, a measure of the statistical significance of the differential expression, a nucleic acid transcript including hnRNA and mRNA is identified as being differentially expressed as between a first and second population when the p-value of less than 0.3, 0.2, 0.1, less than 0.05, less than 0.01, less than 0.005, less than 0.001 etc. are considered statistically significant. When determining differential expression on the basis of the ratio of the level of gene product expression, an RNA or protein gene product is differentially expressed if the ratio of the level of its RNA or protein product in a first sample as compared with that in a second sample is greater than or less than 1.0. For instance, a ratio of greater than 1, for example 1.2, 1.5, 1.7, 2, 3, 4, 10, 20, or a ratio of less than 1, for example 0.8, 0.6, 0.4, 0.2, 0.1. 0.05, of RNA or protein product of a gene would be indicative of differential expression. In another embodiment of the invention, a nucleic acid transcript including hnRNA and mRNA is differentially expressed if the ratio of the mean level of expression of a first transcript in a nucleic acid population as compared with its mean level of expression in a second population is greater than or less than 1.0. For instance, a ratio of greater than 1, for example 1.2, 1.5, 1.7, 2, 3, 4, 10, 20, or a ratio less than 1, for example 0.8, 0.6, 0.4, 0.2, 0.1. 0.05 would be indicative of differential expression. In another embodiment of the invention a nucleic acid transcript including hnRNA, and mRNA is differentially expressed if the ratio of its level of expression in a first sample as compared with the mean of the second population is greater than or less than 1.0 and includes for example, a ratio of greater than 1, for instance 1.2, 1.5, 1.7, 2, 3, 4, 10, 20, or a ratio less than 1, for example 0.8, 0.6, 0.4, 0.2, 0.1. 0.05. "Differentially increased expression" refers to 1.1 fold, 1.2 fold, 1.4 fold, 1.6 fold, 1.8 fold, or more, relative to a standard, such as the mean of the expression level of the second population. "Differentially decreased expression" refers to less than 1.0 fold, 0.8 fold, 0.6 fold, 0.4 fold, 0.2 fold, 0.1 fold or less, relative to a standard, such as the mean of the expression level of the second population.

As used herein, the term "drug efficacy" refers to the effectiveness of a drug. "Drug efficacy" is usually measured by the clinical response of the patient who has been or is

being treated with a drug. A drug is considered to have a high degree of efficacy, if it achieves desired clinical results, for example, the reduction of the symptoms of osteoarthritis or the prevention of osteoarthritis progression as described in the present specification. The amount of drug absorbed may be used to predict a patient's response. A general rule is that as the dose of a drug is increased, a greater effect is seen in the patient until a maximum desired effect is reached. If more drug is administered after the maximum point is reached, the side effects will normally increase.

As used herein, the term "effective amount" refers to the amount of a compound which is sufficient to reduce or ameliorate the progression and or severity of bladder cancer or one or more symptoms thereof, prevent the development, recurrence or onset of bladder cancer or one or more symptoms thereof, prevent the advancement of bladder cancer or one or more symptoms thereof, or enhance or improve the prophylactic or therapeutic effect(s) of another therapy.

As used herein, the term "fragment" in the context of a proteinaceous agent refers to a peptide or polypeptide comprising an amino acid sequence of at least 5 contiguous amino acid residues, at least 10 contiguous amino acid residues, at least 15 contiguous amino acid residues, at least 20 contiguous amino acid residues, at least 25 contiguous amino acid residues, at least 40 contiguous amino acid residues, at least 50 contiguous amino acid residues, at least 60 contiguous amino residues, at least 70 contiguous amino acid residues, at least contiguous 80 amino acid residues, at least contiguous 90 amino acid residues, at least contiguous 100 amino acid residues, at least contiguous 125 amino acid residues, at least 150 contiguous amino acid residues, at least contiguous 175 amino acid residues, at least contiguous 200 amino acid residues, or at least contiguous 250 amino acid residues of the amino acid sequence of a polypeptide or a protein. In a specific embodiment, a fragment of a protein or polypeptide retains at least one function of the protein or polypeptide. In another embodiment, a fragment of a protein or polypeptide retains at least one, two, three, four, or five functions of the protein or polypeptide. Preferably, a fragment of an antibody retains the ability to immunospecifically bind to an antigen.

As used herein, the term "fusion protein" refers to a polypeptide that comprises an amino acid sequence of a first protein or polypeptide or fragment thereof, or functional fragment thereof, or an analog or derivative thereof, and an amino acid sequence

of a heterologous protein, polypeptide, or peptide (i.e., a second protein or polypeptide or fragment, analog or derivative thereof different than the first protein or fragment, analog or derivative thereof). In one embodiment, a fusion protein comprises a prophylactic or therapeutic agent fused to a heterologous protein, polypeptide or peptide. In accordance with this embodiment, the heterologous protein, polypeptide or peptide may or may not be a different type of prophylactic or therapeutic agent.

As used herein, the terms "gene expression pattern", "gene expression profile" are used interchangeably and comprise the pattern of expression of the RNA or protein products of two or more biomarkers of the invention. For example, the gene expression pattern or profile having reference to the RNA products can also be termed a "nucleic acid array expression profile" and constitutes the hybridization of a plurality of target nucleic acid sequences hybridized to a plurality of nucleic acid probes on an array to differentiate as between bladder cancer or early stage bladder cancer individuals as compared with non bladder cancer individuals. "Gene expression pattern", "gene expression profile" and can also refer to a pattern of the level of abundance of proteins corresponding to two or more biomarkers of the invention as is determined by any methodology known in the art for measuring the levels of said proteins. For example, the pattern can be a mathematical representation of the pattern e.g. a mathematical equation, vector etc.

As used herein, the terms "hybridizing to" and "hybridization" refer to the sequence specific non-covalent binding interactions with a complementary nucleic acid, for example, interactions between a target nucleic acid sequence and a nucleic acid member on an array.

As used herein, the term "immunoglobulin" refers to a protein consisting of one or more polypeptides substantially encoded by immunoglobulin genes or antigen binding fragment thereof. The recognized human immunoglobulin genes include the kappa, lambda, alpha (IgA1 and IgA2), gamma (IgG1, IgG2, IgG3, IgG4), delta, epsilon and mu constant region genes, as well as the myriad immunoglobulin variable region genes. Full-length immunoglobulin "light chains" (about 25 Kd or 214 amino acids) are encoded by a variable region gene at the NH2-terminus (about 110 amino acids) and a kappa or lambda constant region gene at the COOH-terminus. Full-length immunoglobulin "heavy chains" (about 50 Kd or 446 amino acids), are similarly encoded by a variable region gene (about 116 amino

acids) and one of the other aforementioned constant region genes, e.g., gamma (encoding about 330 amino acids).

As used herein, the term "in combination" in reference to therapy refers to the use of more than one therapies (*e.g.*, more than one prophylactic agent and/or therapeutic agent). The use of the term "in combination" does not restrict the order in which therapies (*e.g.*, prophylactic and/or therapeutic agents) are administered to a subject. A first therapy (*e.g.*, a first prophylactic or therapeutic agent) can be administered prior to (*e.g.*, 5 minutes, 15 minutes, 30 minutes, 45 minutes, 1 hour, 2 hours, 4 hours, 6 hours, 12 hours, 24 hours, 48 hours, 72 hours, 96 hours, 1 week, 2 weeks, 3 weeks, 4 weeks, 5 weeks, 6 weeks, 8 weeks, or 12 weeks before), concomitantly with, or subsequent to (*e.g.*, 5 minutes, 15 minutes, 30 minutes, 45 minutes, 1 hour, 2 hours, 4 hours, 6 hours, 12 hours, 24 hours, 48 hours, 72 hours, 96 hours, 1 week, 2 weeks, 3 weeks, 4 weeks, 5 weeks, 6 weeks, 8 weeks, or 12 weeks after) the administration of a second therapy (*e.g.*, a second prophylactic or therapeutic agent) to a subject.

As used herein, the term "indicative of disease" or "indicative of bladder cancer" when referring to an expression pattern indicates an expression pattern which is diagnostic of bladder cancer and/or early stage bladder cancer and can include a pattern which increases the likelihood that an individual has bladder cancer and/or early stage bladder cancer.

As used herein, the term "internal coding region" of a gene refers to a polynucleotide (double-stranded or single-stranded) located between the 5' region and the 3' region of a gene as defined herein. The "internal coding region" is not shorter than 8 nucleotides in length and can be as long or longer than 1000 nucleotides in length. Other possible lengths of the "internal coding region" include but are not limited to 10, 20, 25, 50, 100, 200, 400, and 500 nucleotides. The 5', 3' and internal regions are non-overlapping and may, but need not be contiguous, and may, but need not, add up to the full length of the corresponding gene.

As used herein, the term "internal polypeptide region" of a polypeptide refers to the polypeptide sequences located between the amino terminal region and the carboxy terminal region of a polypeptide, as defined herein. The "internal polypeptide region" of a polypeptide is not shorter than 3 amino acids in length and can be as long as or longer than

350 amino acids in length. Other possible lengths of the "internal polypeptide region" of a polypeptide include, but are not limited to, 5, 10, 20, 25, 50, 100 and 200 amino acids.

The amino terminal, carboxy terminal and internal polypeptide regions of a polypeptide are non-overlapping and may, but need not be contiguous, and may, but need not, add up to the full length of the corresponding polypeptide.

As used herein, "isolated" or "purified" when used in reference to a nucleic acid means that a naturally occurring sequence has been removed from its normal cellular (e.g., chromosomal) environment or is synthesised in a non-natural environment (e.g., artificially synthesised). Thus, an "isolated" or "purified" sequence may be in a cell-free solution or placed in a different cellular environment. The term "purified" does not imply that the sequence is the only nucleotide present, but that it is essentially free (about 90-95% pure) of non-nucleotide material naturally associated with it, and thus is distinguished from isolated chromosomes.

-As used herein, the terms "isolated" and "purified" in the context of a proteinaceous agent (e.g., a peptide, polypeptide, protein or antibody) refer to a proteinaceous agent which is substantially free of cellular material and in some embodiments, substantially free of heterologous proteinaceous agents (i.e., contaminating proteins) from the cell or tissue source from which it is derived, or substantially free of chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of a proteinaceous agent in which the proteinaceous agent is separated from cellular components of the cells from which it is isolated or recombinantly produced. Thus, a proteinaceous agent that is substantially free of cellular material includes preparations of a proteinaceous agent having less than about 30%, 20%, 10%, or 5% (by dry weight) of heterologous proteinaceous agent (e.g., protein, polypeptide, peptide, or antibody; also referred to as a "contaminating protein"). When the proteinaceous agent is recombinantly produced, it is also preferably substantially free of culture medium, i.e., culture medium represents less than about 20%, 10%, or 5% of the volume of the protein preparation. When the proteinaceous agent is produced by chemical synthesis, it is preferably substantially free of chemical precursors or other chemicals, i.e., it is separated from chemical precursors or other chemicals which are involved in the synthesis of the proteinaceous agent. Accordingly, such preparations of a proteinaceous agent have less than

about 30%, 20%, 10%, 5% (by dry weight) of chemical precursors or compounds other than the proteinaceous agent of interest. Preferably, proteinaceous agents disclosed herein are isolated.

As used herein, the term "level of expression" refers to the determination of the quantity of a given nucleic acid or protein corresponding to a gene as determined by methods known to a person skilled. In reference to RNA, hnRNA, mRNA or spliced variants of mRNA corresponding to a biomarker of the invention, level of expression can be determined by hybridization as well as other measurements such as quantitative real-time RT PCR, which includes use of SYBR® green, TaqMan® and Molecular Beacons technology. Note that as used herein the determination of differential levels of expression can include a visual inspection of differences as between the quantity of a given nucleic acid or protein, for example by analyzing the northern blot or western blot.

As used herein, a "ligand" is a molecule that specifically binds to a polypeptide encoded by one of the genes of a biomarker of the invention. A ligand can be a nucleic acid (RNA or DNA), polypeptide, peptide or chemical compound. A ligand of the invention can be a peptide ligand, e.g., a scaffold peptide, a linear peptide, or a cyclic peptide. In a preferred embodiment, the polypeptide ligand is an antibody. The antibody can be a human antibody, a chimeric antibody, a recombinant antibody, a humanized antibody, a monoclonal antibody or antigen binding fragment thereof,, or a polyclonal antibody. The antibody can be an intact immunoglobulin, e.g., an IgA, IgG, IgE, IgD, IgM or subtypes thereof. The antibody can be conjugated to a functional moiety (e.g., a compound which has a biological or chemical function (which may be a second different polypeptide, a therapeutic drug, a cytotoxic agent, a detectable moiety, or a solid support. A polypeptide ligand e.g. antibody of the invention interacts with a polypeptide, encoded by one of the genes of a biomarker, with high affinity and specificity. For example, the polypeptide ligand binds to a polypeptide, encoded by one of the genes of a biomarker, with an affinity constant of at least 10^7 M^{-1}, preferably, at least 10^8 M^{-1}, 10^9 M^{-1}, or 10^{10} M^{-1}.

As used herein, the term "majority" refers to a number representing more than 50% (e.g., 51%, 60%, or 70%, or 80% or 90% or up to 100%) of the total members of a composition. The term "majority", when referring to an array, it means more than 50% (e.g.,

51%, 60%, or 70%, or 80% or 90% or up to 100%) of the total nucleic acid members that are stably associated with the solid substrate of the array.

As used herein, the terms "manage", "managing" and "management" refer to the beneficial effects that a subject derives from a therapy (e.g., a prophylactic or therapeutic agent) which does not result in a cure of osteoarthritis. In certain embodiments, a subject is administered one or more therapies to "manage" osteoarthritis so as to prevent the progression or worsening of the osteoarthritis.

As used herein, "mRNA integrity" refers to the quality of mRNA extracts from either tissue samples or blood samples. In one embodiment, mRNA integrity is determined by examining whether the RNA is degraded when examined by methods well known in the art, for example, by RNA agarose gel electrophoresis (e.g., Ausubel et al., John Weley & Sons, Inc., 1997, Current Protocols in Molecular Biology). Preferably, the mRNA samples have good integrity (e.g., less than 10%, preferably less than 5%, and more preferably less than 1% of the mRNA is degraded) to truly represent the gene expression levels of the cartilage or blood samples from which they are extracted.

As used herein, the terms "non-responsive" and refractory" describe patients treated with a currently available therapy which is not clinically adequate to treat and/or relieve one or more symptoms of bladder cancer.

As used herein, "normal" refers to an individual or group of individuals who have not shown any evidence of bladder cancer, or symptoms thereof including blood in urine, and have not been diagnosed with bladder cancer or the possibility that they may have bladder cancer. Preferably said "normal" rers to an individual or group of individuals who is not at an increased risk of having bladder cancer. In addition, preferably said normal individual(s) is not on medication affecting bladder cancer and has not been diagnosed with any other disease. More preferably normal individuals have similar sex, age as compared with the test samples. "Normal", according to the invention, also refers to a samples isolated from normal individuals and includes total RNA or mRNA isolated from normal individuals. A sample taken from a normal individual can include RNA isolated from a tissue sample.

As used herein, "nucleic acid(s)" is interchangeable with the term "polynucleotide(s)" and it generally refers to any polyribonucleotide or poly-deoxyribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA or

any combination thereof. "Nucleic acids" include, without limitation, single- and double-stranded nucleic acids. As used herein, the term "nucleic acid(s)" also includes DNAs or RNAs as described above that contain one or more modified bases. Thus, DNAs or RNAs with backbones modified for stability or for other reasons are "nucleic acids". The term "nucleic acids" as it is used herein embraces such chemically, enzymatically or metabolically modified forms of nucleic acids, as well as the chemical forms of DNA and RNA characteristic of viruses and cells, including for example, simple and complex cells. A "nucleic acid" or "nucleic acid sequence" may also include regions of single- or double-stranded RNA or DNA or any combinations thereof and can include expressed sequence tags (ESTs) according to some embodiments of the invention. An EST is a portion of the expressed sequence of a gene (i.e., the "tag" of a sequence), made by reverse transcribing a region of mRNA so as to make cDNA.

As defined herein, a "nucleic acid array" refers a plurality of unique nucleic acids (or "nucleic acid members") attached to a support where each of the nucleic acid members is attached to a support in a unique pre-selected region. In one embodiment, the nucleic acid member attached to the surface of the support is DNA. In a preferred embodiment, the nucleic acid member attached to the surface of the support is either cDNA or oligonucleotides. In another preferred embodiment, the nucleic acid member attached to the surface of the support is cDNA synthesised by polymerase chain reaction (PCR). The term "nucleic acid", as used herein, is interchangeable with the term "polynucleotide". In another preferred embodiment, a "nucleic acid array" refers to a plurality of unique nucleic acids attached to nitrocellulose or other membranes used in Southern and/or Northern blotting techniques.

As used herein, a "nucleic acid probe" includes nucleic acids capable of binding to a complementary sequence of a nucleic acid member on an array through sets of non-covalent bonding interactions, including complementary base pairing interactions. As used herein, a nucleic acid probe may include natural (i. e., A, G, C, or T) or modified bases (7-deazaguanosine, inosine, etc.). In addition, the bases in nucleic acid probes may be joined by a linkage other than a phosphodiester bond, so long as it does not interfere with hybridization (i.e., the nucleic acid probe still specifically binds to its complementary sequence under standard stringent or selective hybridization conditions). Thus, nucleic acid

probes may be peptide nucleic acids in which the constituent bases are joined by peptide bonds rather than phosphodiester linkages.

As used herein "nucleic acid target" or "nucleic acid member" is defined as a nucleic acid capable of binding to an array. The nucleic acid target can either be an isolated nucleic acid sequence corresponding to a gene or portion thereof, or the nucleic acid target can be total RNA or mRNA isolated from a sample. Preferably, the nucleic acid target or nucleic acid markers are derived from human cartilage, blood, or synovial fluid extracts. More preferably, the nucleic acid targets are single- or double-stranded DNA, RNA, or DNA-RNA hybrids, from human cartilage, blood, or synovial fluid total RNA extracts, and preferably from mRNA extracts.

In one embodiment, a conventional nucleic acid array of 'target' sequences bound to the array can be representative of the entire human genome, e.g. Affymetrix chip, and the isolated biomarker consisting of or comprising two or more of the genes described in Table 1 or gene probes is applied to the conventional array.

In another embodiment, sequences bound to the array can be an isolated oligonucleotide, cDNA, EST or PCR product corresponding to a biomarker of the invention total cellular RNA is applied to the array.

As used herein, the term "oligonucleotide" is defined as a molecule comprised of two or more deoxyribonucleotides and/ or ribonucleotides, and preferably more than three. Its exact size will depend upon many factors which, in turn, depend upon the ultimate function and use of the oligonucleotide. The oligonucleotides may be from about 8 to about 1,000 nucleotides long. Although oliognucleotides of 8 to 100 nucleotides are useful in the invention, preferred oligonucleotides range from about 8 to about 15 bases in length, from about 8 to about 20 bases in length, from about 8 to about 25 bases in length, from about 8 to about 30 bases in length, from about 8 to about 40 bases in length or from about 8 to about 50 bases in length.

As used herein, the phrase "pharmaceutically acceptable salt(s)," includes, but is not limited to, salts of acidic or basic groups that may be present in compounds identified using the methods of the present invention. Compounds that are basic in nature are capable of forming a wide variety of salts with various inorganic and organic acids. The acids that can be used to prepare pharmaceutically acceptable acid addition salts of such basic

compounds are those that form non-toxic acid addition salts, i.e., salts containing pharmacologically acceptable anions, including but not limited to sulfuric, citric, maleic, acetic, oxalic, hydrochloride, hydrobromide, hydroiodide, nitrate, sulfate, bisulfate, phosphate, acid phosphate, isonicotinate, acetate, lactate, salicylate, citrate, acid citrate, tartrate, oleate, tannate, pantothenate, bitartrate, ascorbate, succinate, maleate, gentisinate, fumarate, gluconate, glucaronate, saccharate, formate, benzoate, glutamate, methanesulfonate, ethanesulfonate, benzenesulfonate, p-toluenesulfonate and pamoate (i.e., 1,1'-methylene-bis-(2-hydroxy-3-naphthoate)) salts. Compounds that include an amino moiety may form pharmaceutically acceptable salts with various amino acids, in addition to the acids mentioned above. Compounds that are acidic in nature are capable of forming base salts with various pharmacologically acceptable cations. Examples of such salts include alkali metal or alkaline earth metal salts and, particularly, calcium, magnesium, sodium lithium, zinc, potassium, and iron salts.

As used herein, "polynucleotide" encompasses double-stranded DNA, single-stranded DNA and double-stranded or single-stranded RNA of more than 8 nucleotides in length.

As used herein, "polypeptide sequences encoded by" refers to the amino acid sequences obtained after translation of the protein coding region of a gene, as defined herein. The mRNA nucleotide sequence for each of the biomarkers of the invention is identified by its Genbank Accession number (see Table 3 and/or Table 6 and/or Table 12) and the corresponding polypeptide sequence is identified by a Protein Accession number (see Table 3 and/or Table 6 and/or Table 12).

When a protein or fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to a given region or three-dimensional structure on the protein; these regions or structures are referred to as epitopes or antigenic determinants. As used herein, "antigenic fragments" refers portions of a polypeptide that contains one or more epitopes. Epitopes can be linear, comprising essentially a linear sequence from the antigen, or conformational, comprising sequences which are genetically separated by other sequences but come together structurally at the binding site for the polypeptide ligand. "Antigenic fragments" may be up to

any one of 5000, 1000, 500, 400, 300, 200, 100, 50 or 25 or 20 or 10 or 5 amino acids in length.

As used herein, "pre-selected region", "predefined region", or "unique position" refers to a localised area on a substrate which is, was, or is intended to be used for the deposit of a nucleic acid and is otherwise referred to herein in the alternative as a "selected region" or simply a "region." The pre-selected region may have any convenient shape, e.g., circular, rectangular, elliptical, wedge-shaped, etc. In some embodiments, a pre-selected region is smaller than about 1 cm^2, more preferably less than 1 mm^2, still more preferably less than 0.5 mm^2, and in some embodiments less than 0.1 mm^2. A nucleic acid member at a "pre-selected region", "predefined region", or "unique position" is one whose identity (e.g., sequence) can be determined by virtue of its position at the region or unique position.

As used herein, the terms "prevent", "preventing" and "prevention" refer to the prevention of the development, recurrence or onset of mild osteoarthritis or one or more symptoms thereof resulting from the administration of one or more compounds identified in accordance the methods of the invention or the administration of a combination of such a compound and another therapy.

The term, "primer", as used herein refers to an oligonucleotide, whether occurring naturally as in a purified restriction digest or produced synthetically, which is capable of acting as a point of initiation of synthesis when placed under conditions in which synthesis of a primer extension product, which is complementary to a nucleic acid strand, is induced, i.e., in the presence of nucleotides and an inducing agent such as a DNA polymerase and at a suitable temperature and pH. The primer may be either single-stranded or double-stranded and must be sufficiently long to prime the synthesis of the desired extension product in the presence of the inducing agent. The exact length of the primer will depend upon many factors, including temperature, source of primer and the method used. For example, for diagnostic applications, depending on the complexity of the target sequence, the oligonucleotide primer typically contains 15-25 or more nucleotides, although it may contain fewer nucleotides. The factors involved in determining the appropriate length of primer are readily known to one of ordinary skill in the art. In general, the design and selection of

primers embodied by the instant invention is according to methods that are standard and well known in the art, see Dieffenbach, C.W., Lowe, T.M.J., Dveksler, G.S. (1995) General Concepts for PCR Primer Design. In: PCR Primer, A Laboratory Manual (Eds. Dieffenbach, C.W, and Dveksler, G.S.) Cold Spring Harbor Laboratory Press, New York, 133-155; Innis, M.A., and Gelfand, D.H. (1990) Optimization of PCRs. In: PCR protocols, A Guide to Methods and Applications (Eds. Innis, M.A., Gelfand, D.H., Sninsky, J.J., and White, T.J.) Academic Press, San Diego, 3-12; Sharrocks, A.D. (1994) The design of primers for PCR. In: PCR Technology, Current Innovations (Eds. Griffin, H.G., and Griffin, A.M, Ed.) CRC Press, London, 5-11.

As used herein, the term "probe" means oligonucleotides and analogs thereof and refers to a range of chemical species that recognise polynucleotide target sequences through hydrogen bonding interactions with the nucleotide bases of the target sequences. The probe or the target sequences may be single- or double-stranded RNA or single- or double-stranded DNA or a combination of DNA and RNA bases. A probe is at least 8 nucleotides in length and less than the length of a complete gene. A probe may be 10, 20, 30, 50, 75, 100, 150, 200, 250, 400, 500 and up to 2000 nucleotides in length. Probes can include oligonucleotides modified so as to have a tag which is detectable by fluorescence, chemiluminescence and the like. The probe can also be modified so as to have both a detectable tag and a quencher molecule, for example Taqman® and Molecular Beacon® probes.

The oligonucleotides and analogs thereof may be RNA or DNA, or analogs of RNA or DNA, commonly referred to as antisense oligomers or antisense oligonucleotides. Such RNA or DNA analogs comprise but are not limited to 2-'O-alkyl sugar modifications, methylphosphonate, phosphorothiate, phosphorodithioate, formacetal, 3'-thioformacetal, sulfone, sulfamate, and nitroxide backbone modifications, and analogs wherein the base moieties have been modified. In addition, analogs of oligomers may be polymers in which the sugar moiety has been modified or replaced by another suitable moiety, resulting in polymers which include, but are not limited to, morpholino analogs and peptide nucleic acid (PNA) analogs (Egholm, et al. Peptide Nucleic Acids (PNA)--Oligonucleotide Analogues with an Achiral Peptide Backbone, (1992)).

Probes may also be mixtures of any of the oligonucleotide analog types together or in combination with native DNA or RNA. At the same time, the oligonucleotides and analogs thereof may be used alone or in combination with one or more additional oliognucleotides or analogs thereof.

As used herein, the term "product of the biomarker" or "biomarker product" refers to the RNA or protein which corresponds or is encoded by the biomarker (i.e. is transcribed from the gene or genetic element or is translated from RNA which is transcribed from the gene or genetic element). For example, in some embodiments RNA resulting from the biomarker can include one or more of the following species; hnRNA, mRNA, and/or one or more spliced variants of mRNA. In some embodiments, proteins resulting from the molecular marker can include any proteins found in blood which correspond to the RNA resulting from the biomarker.

As used herein, the terms "prophylactic agent" and "prophylactic agents" refer to any compound(s) which can be used in the prevention of osteoarthritis. In certain embodiments, the term "prophylactic agent" refers to a compound identified in the screening assays described herein. In certain other embodiments, the term "prophylactic agent" refers to an agent other than a compound identified in the screening assays described herein which is known to be useful for, or has been or is currently being used to prevent or impede the onset, development and/or progression of osteoarthritis or one or more symptoms thereof.

As used herein, the phrase "prophylactically effective amount" refers to the amount of a therapy (e.g., a prophylactic agent) which is sufficient to result in the prevention of the development, recurrence or onset of osteoarthritis or one or more symptoms thereof.

As used herein, the terms "protein" and "polypeptide" and "proteinaceous agent" are used interchangeably to refer to a chain of amino acidslinked together by peptide bonds which optionally can comprise natural or non –natural amino acids. Optionally, the protein or peptide can comprise other molecules in addition to amino acids. Said chain can be of any length. In a specific embodiment, a protein is composed of less than 200, less than 175, less than 150, less than 125, less than 100, less than 50, less than 45, less than 40, less than 35, less than 30, less than 25, less than 20, less than 15, less than 10, or less than 5 amino acids linked together by peptide bonds. In another embodiment, a protein is composed

of at least 200, at least 250, at least 300, at least 350, at least 400, at least 450, at least 500 or more amino acids linked together by peptide bonds.

As used herein, "a plurality of" or "a set of" refers to more than two, for example, 3 or more, 10 or more, 100 or more, or 1000 or more, or 10,000 or more.

As used herein, the terms "RNA portion" and "a portion thereof" in context of RNA products of a biomarker of the invention refer to an RNA transcript comprising a nucleic acid sequence of at least 6, at least 9, at least 15, at least 18, at least 21, at least 24, at least 30, at least 60, at least 90, at least 99, or at least 108, or more nucleotides of a RNA product of a biomarker of the invention.

As used herein the term "product of the biomarker of the invention" refers to the RNA and/or the protein expressed by the gene corresponding to the biomarker of the invention. The "RNA product of a biomarker of the invention" includes one or more products which can include heteronuclear RNA ("hnRNA"), mRNA, and all or some of the spliced variants of mRNA whose measure of expression can be used as a biomarker in accordance with the teachings disclosed herein. The "protein product of a biomarker of the invention" includes one or more of the products of the biomarker which can include proteins, protein variants, and any post-translationally modified proteins.

As used herein, the term "selectively amplified" or "selective amplification", refers to a process whereby one or more copies of a particular target nucleic acid sequence is selectively generated from a template nucleic acid. Selective amplification or selectively amplified is to be compared with amplification in general which can be used as a method in combination with, for example, random primers and an oligodT primer to amplify a population of nucleic acid sequences (e.g. mRNA). Selective amplification is preferably done by the method of polymerase chain reaction (Mullis and Faloona, 1987, Methods Enzymol. 155:335).

As used herein, the term "selectively binds" in the context of proteins encompassed by the invention refers to the specific interaction of any two of a peptide, a protein, a polypeptide, and an antibody, wherein the interaction preferentially occurs as between any two of a peptide, protein, polypeptide and antibody preferentially as compared

with any other peptide, protein, polypeptide and antibody. For example, when the two molecules are protein molecules, a structure on the first molecule recognises and binds to a structure on the second molecule, rather than to other proteins. "Selective binding", "Selective binding", as the term is used herein, means that a molecule binds its specific binding partner with at least 2-fold greater affinity, and preferably at least 10-fold, 20-fold, 50-fold, 100-fold or higher affinity than it binds a non-specific molecule.

As used herein "selective hybridization" in the context of this invention refers to a hybridization which occurs as between a polynucleotide encompassed by the invention and an RNA , or its complement thereof, or protein product of the biomarker of the invention, wherein the hybridization is such that the polynucleotide preferentially binds to the RNA products of the biomarker of the invention relative to the RNA products of other genes in the genome in question. In a preferred embodiment a polynucleotide which "selectively hybridizes" is one which hybridizes with a selectivity of greater than 70%, greater than 80%, greater than 90% and most preferably of 100% (i.e. cross hybridization with other RNA species preferably occurs at less than 30%, less than 20%, less than 10%). As would be understood to a person skilled in the art, a polynucleotide which "selectively hybridizes" to the RNA product of a biomarker of the invention can be determined taking into account the length and composition.

As used herein, "specifically hybridizes", "specific hybridization" refers to hybridization which occurs when two nucleic acid sequences are substantially complementary (at least about 65% complementary over a stretch of at least 14 to 25 nucleotides, preferably at least about 75% complementary, more preferably at least about 90% complementary). See Kanehisa, M., 1984, Nucleic acids Res., 12:203, incorporated herein by reference. As a result, it is expected that a certain degree of mismatch is tolerated. Such mismatch may be small, such as a mono-, di- or tri-nucleotide. Alternatively, a region of mismatch can encompass loops, which are defined as regions in which there exists a mismatch in an uninterrupted series of four or more nucleotides. Numerous factors influence the efficiency and selectivity of hybridization of two nucleic acids, for example, the hybridization of a nucleic acid member on an array to a target nucleic acid sequence. These factors include nucleic acid member length, nucleotide sequence and/or composition,

hybridization temperature, buffer composition and potential for steric hindrance in the region to which the nucleic acid member is required to hybridize. A positive correlation exists between the nucleic acid length and both the efficiency and accuracy with which a nucleic acid will anneal to a target sequence. In particular, longer sequences have a higher melting temperature (T_M) than do shorter ones, and are less likely to be repeated within a given target sequence, thereby minimizing non-specific hybridization. Hybridization temperature varies inversely with nucleic acid member annealing efficiency. Similarly the concentration of organic solvents, e.g., formamide, in a hybridization mixture varies inversely with annealing efficiency, while increases in salt concentration in the hybridization mixture facilitate annealing. Under stringent annealing conditions, longer nucleic acids, hybridize more efficiently than do shorter ones, which are sufficient under more permissive conditions.

As used herein, the term "specifically binds" refers to the interaction of two molecules, e.g., a ligand and a protein or peptide, or an antibody and a protein or peptide wherein the interaction is dependent upon the presence of particular structures on the respective molecules. For example, when the two molecules are protein molecules, a structure on the first molecule recognises and binds to a structure on the second molecule, rather than to proteins in general. "Specific binding", as the term is used herein, means that a molecule binds its specific binding partner with at least 2-fold greater affinity, and preferably at least 10-fold, 20-fold, 50-fold, 100-fold or higher affinity than it binds a non-specific molecule.

As herein used, the term "standard stringent conditions" and "stringent conditions" means hybridization will occur only if there is at least 95% and preferably, at least 97% identity between the sequences, wherein the region of identity comprises at least 10 nucleotides. In one embodiment, the sequences hybridize under stringent conditions following incubation of the sequences overnight at 42°C, followed by stringent washes (0.2X SSC at 65° C). The degree of stringency of washing can be varied by changing the temperature, pH, ionic strength, divalent cation concentration, volume and duration of the washing. For example, the stringency of hybridization may be varied by conducting the hybridization at varying temperatures below the melting temperatures of the probes. The melting temperature of the probe may be calculated using the following formulas:

For oligonucleotide probes, between 14 and 70 nucleotides in length, the melting temperature (Tm) in degrees Celcius may be calculated using the formula: Tm=81.5+16.6(log [Na+]) + 0.41(fraction G+C)-(600/N) where N is the length of the oligonucleotide.

For example, the hybridization temperature may be decreased in increments of 5 °C from 68 °C to 42 °C in a hybridization buffer having a Na+ concentration of approximately 1M. Following hybridization, the filter may be washed with 2X SSC, 0.5% SDS at the temperature of hybridization. These conditions are considered to be "moderate stringency" conditions above 50°C and "low stringency" conditions below 50°C. A specific example of "moderate stringency" hybridization conditions is when the above hybridization is conducted at 55°C. A specific example of "low stringency" hybridization conditions is when the above hybridization is conducted at 45°C.

If the hybridization is carried out in a solution containing formamide, the melting temperature of the annealing nucleic acid strands may be calculated using the equation $Tm=81.5+16.6(\log [Na^+]) + 0.41(\text{fraction G + C})-(0.63\% \text{ formamide})-(600/N)$, where N is the length of the probe.

For example, the hybridization may be carried out in buffers, such as 6X SSC, containing formamide at a temperature of 42 °C. In this case, the concentration of formamide in the hybridization buffer may be reduced in 5% increments from 50% to 0% to identify clones having decreasing levels of homology to the probe. Following hybridization, the filter may be washed with 6X SSC, 0.5% SDS at 50 °C. Hybridization conditions are considered to be "moderate stringency" conditions when hybridization fluids are comprised of above 25% formamide and "low stringency" conditions when hybridization fluids are comprised of below 25% formamide. A specific example of "moderate stringency" hybridization conditions is when the above hybridization is conducted at 30% formamide. A specific example of "low stringency" hybridization conditions is when the above hybridization is conducted at 10% formamide.

As used herein, the terms "subject" and "patient" and "individual" are used interchangeably to refer to an animal (e.g., a mammal, a fish, an amphibian, a reptile, a bird

and an insect). In a specific embodiment, a subject is a mammal (e.g., a non-human mammal and a human). In another embodiment, a subject is a pet (e.g., a dog, a cat, a guinea pig, a monkey and a bird), a farm animal (e.g., a horse, a cow, a pig, a goat and a chicken) or a laboratory animal (e.g., a mouse and a rat). In another embodiment, a subject is a primate (e.g., a chimpanzee and a human). In another embodiment, a subject is a human.

As used herein, the term "synergistic" refers to a combination of a compound identified using one of the methods described herein, and another therapy (e.g., agent), which is more effective than the additive effects of the therapies. Preferably, such other therapy has been or is currently being to prevent, treat, manage or ameliorate bladder cancer or a symptom thereof. A synergistic effect of a combination of therapies (e.g., prophylactic or therapeutic agents) permits the use of lower dosages of one or more of the therapies and/or less frequent administration of said therapies to a subject with bladder cancer. The ability to utilize lower dosages of a therapy (e.g., a prophylactic or therapeutic agent) and/or to administer said therapy less frequently reduces the toxicity associated with the administration of said agent to a subject without reducing the efficacy of said therapies in the prevention, treatment, management or amelioration of bladder cancer. In addition, a synergistic effect can result in improved efficacy of therapies (e.g., agents) in the prevention, treatment, management or amelioration of bladder cancer. Finally, a synergistic effect of a combination of therapies (e.g., prophylactic or therapeutic agents) may avoid or reduce adverse or unwanted side effects associated with the use of either therapy alone.

As used herein, the terms "therapeutic agent" and "therapeutic agents" refer to any compound(s) which can be used in the treatment, management or amelioration of bladder cancer or one or more symptoms thereof.

As used herein, the term "therapeutically effective amount" refers to that amount of a therapy (e.g., a therapeutic agent) sufficient to result in the amelioration of bladder cancer or one or more symptoms thereof, prevent advancement of bladder cancer cause regression of bladder cancer, or to enhance or improve the therapeutic effect(s) of another therapy (e.g., therapeutic agent.

As used herein, the terms "treat", "treatment" and "treating" refer to the reduction or amelioration of the progression, severity and/or reoccurrence of bladder cancer

or one or more symptoms thereof resulting from the administration of one or more compounds identified in accordance the methods of the invention, or a combination of one or more compounds identified in accordance with the invention and another therapy.

As used herein, the term "up regulated" or "increased level of expression" in the context of this invention refers the product of a gene which is expressed wherein the measure of the product demonstrates an increased level of expression of the gene, as can be determined using array analysis or other similar analysis, in tissue or blood isolated from an individual having bladder cancer or an identified disease state or stage of of bladder cancer as determined by AJCC staging guidelines as compared with the same gene in tissue or blood isolated from normal individuals or from an individual with a different identified disease state or stage. An "increased level of expression" according to the present invention, is an increase in expression of at least 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%,10% or more, for example, 20%, 30%, 40%, or 50%, 60%, 70%, 80%, 90% or more, or greater than 1-fold, up to 2-fold, 3-fold, 4-fold, 5-fold, 10-fold, 50-fold, 100-fold or more as measured, for example, by the intensity of hybridization according to methods of the present invention. For example, up regulated sequences includes sequences having an increased level of expression in tissue or blood isolated from individuals characterised as having bladder cancer as compared with tissue isolated from normal individuals. Up regulated sequences can also include sequences having an increased level of expression in tissue or blood isolated from individuals characterised as having one stage of bladder cancer as compared to another stage of bladder cancer.

BRIEF DESCRIPTION OF THE DRAWINGS

The invention will now be described in relation to the drawings in which:

Figure 1 is a graphical representation of an embodiment of the invention showing all 21 possible combinations of ratios of the seven biomarkers noted in **Example 9.** The ROC for each classifier of logistic regression resulting from each possible combination of biomarkers is noted in blue. For each resulting classifier, the sensitivity is noted in red where

the specificity is set at 90%. The specificity for each classifier wherein the sensitivity is set at 90% is noted in green.

Figure 2 is a graphical representation of an embodiment of the invention showing all possible 210 of two ratios of the seven biomarkers noted in **Example 9.** The ROC for each classifier of logistic regression resulting from each possible combination of biomarkers is noted in blue. For each resulting classifier, the sensitivity is noted in red where the specificity is set at 90%. The specificity for each classifier wherein the sensitivity is set at 90% is noted in green.

Figure 3 is a graphical representation of an embodiment of the invention showing 1295 of the 1330 possible combinations of three ratios of the seven biomarkers noted in **Example 9** (note that an exhaustive search can be done, but in this case was not completed). The ROC for each classifier of logistic regression resulting from each possible combination of biomarkers is noted in blue. For each resulting classifier, the sensitivity is noted in red where the specificity is set at 90%. The specificity for each classifier wherein the sensitivity is set at 90% is noted in green.

Figure 4 is a graphical representation of an embodiment of the invention showing 5250 of the 5985 possible combinations of four ratios of the seven biomarkers noted in **Example 9** (note that an exhaustive search can be done by merely increasing the length of computer time for analysis). The ROC for each classifier of logistic regression resulting from each possible combination of biomarkers is noted in blue. For each resulting classifier, the sensitivity is noted in red where the specificity is set at 90%. The specificity for each classifier wherein the sensitivity is set at 90% is noted in green.

3.2 Embodiments

Described herein is a composition comprising a collection of two or more isolated polynucleotides, each of the polynucleotides selectively hybridizing to a biomarker. These biomarkers are genes identified in Table 1. In one embodiment, the composition can be used

to measure the level of expression of each of any of least two of these biomarkers listed in Table 1. In one embodiment, the polynucleotides of the composition are useful in quantitative RT-PCR (QRT-PCR).

Described herein is a composition comprising a collection of two or more isolated polynucleotides, each of the polynucleotides selectively hybridizing to a biomarker. These biomarkers are genes identified in Table 1 and/or in Table 2. In one embodiment, the composition can be used to measure the level of expression of each of any of least two of these biomarkers listed in Table 1 and/or in Table 2. In one embodiment, the polynucleotides of the composition are useful in quantitative RT-PCR (QRT-PCR).

Described herein is a composition comprising a collection of two or more isolated polynucleotides, each of the polynucleotides selectively hybridizing to a biomarker. These biomarkers are genes identified in Table 11. In one embodiment, the composition can be used to measure the level of expression of each of any of least two of these biomarkers listed in Table 11. In one embodiment, the polynucleotides of the composition are useful in quantitative RT-PCR (QRT-PCR).

Described herein is a composition comprising a collection of two or more isolated polynucleotides, each polynucleotide selectively hybridizing to an RNA product of a biomarker . The RNA products of the biomarkers are identified in Table 3. In one embodiment, the composition can be used to measure the level of expression of at least two of these RNA products. In one embodiment, the polynucleotides of the composition are useful in quantitative RT-PCR (QRT-PCR).

Described herein is a composition comprising a collection of two or more isolated polynucleotides, each polynucleotide selectively hybridizing to a biomarker. These biomarkers are genes identified in Table 4. In one embodiment, the composition can be used to measure the level of expression of at least two of these biomarkers. In one embodiment, the polynucleotides of the composition are useful in quantitative RT-PCR (QRT-PCR).

Described herein is a composition comprising a collection of two or more isolated polynucleotides, each the polynucleotide selectively hybridizing to a biomarker. The biomarkers are genes identified in Table 4 and Table 5, and at least one of these biomarkers is selected from Table 4. In one embodiment, the composition can be used to measure the level

of expression of at least two of the biomarkers. In one embodiment, the polynucleotides of the composition are useful in quantitative RT-PCR (QRT-PCR).

Described herein is a collection of two or more isolated polynucleotides each polynucleotide which selectively hybridizing to an RNA product of a biomarker, wherein the RNA products of the biomarkers are identified in Table 6. In one embodiment, the composition is used to measure the level of expression of at least two of the RNA products. In one embodiment, the polynucleotides of the composition are useful in quantitative RT-PCR (QRT-PCR).

Described herein is a composition comprising a collection of two or more isolated polynucleotides, each polynucleotide selectively hybridizing to a biomarker. The biomarkers are genes identified in Table 7. In one embodiment, the composition is used to measure the level of expression of at least two of the biomarkers. In one embodiment, the polynucleotides of the composition are useful in quantitative RT-PCR (QRT-PCR).

Described herein is a composition comprising a collection of two or more isolated polynucleotides, each polynucleotide selectively hybridizing to a biomarker. The biomarkers are genes identified in Table 10. In one embodiment, the composition is used to measure the level of expression of at least two of the biomarkers. In one embodiment, the polynucleotides of the composition are useful in quantitative RT-PCR (QRT-PCR).

Described herein is a composition comprising a collection of two or more isolated proteins each isolated protein which binds selectively to a protein product of a biomarker , wherein the biomarkers are selected from the group consisting of the genes as set out in Table 1. In one embodiment, the composition is used to measure the level of expression of at least two of the biomarkers.

Described herein is a composition comprising a collection of two or more isolated proteins, each isolated protein which binds selectively to a protein product of a biomarker , wherein the biomarkers are the genes as set out in Table 1 and Table 2, at least one of the biomarkers is selected from Table 1. In one embodiment, the composition is used to measure the level of expression of at least two of the biomarkers.

Described herein is a composition comprising a collection of two or more isolated proteins each isolated protein which binds selectively to a protein product of a biomarker , wherein the protein products of the biomarkers are the proteins as set out in Table 3, In one

embodiment, the composition is used to measure the level of expression of at least two of the protein products.

Described herein is a composition comprising a collection of two or more isolated proteins each isolated protein binding selectively to a protein product of a biomarker , wherein the biomarkers are the genes as set out in Table 4. In one embodiment, the composition is used to measure the level of expression of at least two of the biomarkers.

Described herein is a composition comprising a collection of two or more isolated proteins, each isolated protein which binds selectively to a protein product of a biomarker , wherein the biomarkers are selected from the group consisting of the genes as set out in Table 4 and Table 5, at least one of the biomarkers is selected from Table 4. In one embodiment, the composition is used to measure the level of expression of at least two of the biomarkers.

Described herein is a composition comprising a collection of two or more isolated proteins each isolated protein which binds selectively to a protein product of a biomarker , wherein the protein products of the biomarkers are selected from the proteins as set out in Table 6. In one embodiment, the composition is used to measure the level of expression of at least two of the protein products.

Described herein is a composition comprising a collection of two or more isolated proteins, each isolated protein binding selectively to a protein product of a biomarker, wherein the biomarkers are selected from the genes as set out in Table 7. In one embodiment, the composition is used to measure the level of expression of at least two of these biomarkers.

Described herein is a composition comprising a collection of two or more isolated proteins, each isolated protein binding selectively to a protein product of a biomarker , wherein the biomarkers include the genes as set out in Table 10. In one embodiment, the composition is used to measure the level of expression of at least two of the biomarkers.

In the compositions described herein in the above paragraphs, the isolated proteins are ligands to a protein product of a biomarker, and these ligands can be antibodies, including monoclonal antibodies.

In the compositions described herein in the above paragraphs, the isolated polynucleotides can be single or double stranded RNA, or single or double stranded DNA.

Described herein is a method of diagnosing or prognosing bladder cancer in an individual, comprising the steps of:

a) determining the level of one or more RNA transcripts expressed in blood obtained from the individual, wherein the one or more RNA transcripts corresponds to the one or more biomarkers of Table 1, and

b) comparing the level of each of the one or more RNA transcripts in the blood according to step a) with the level of each of the one or more RNA transcripts in blood from one or more individuals not having bladder cancer,

wherein detecting differential expression of each of the one or more RNA transcripts in the comparison of step b) is indicative of bladder cancer in the individual of step a).

Described herein is a method of diagnosing or prognosing bladder cancer in an individual, comprising the steps of:

a) determining the level of two or more RNA transcripts expressed in blood obtained from the individual, wherein the two or more RNA transcripts corresponds to the one or more biomarkers of Table 1 and Table 2, wherein at least one of the RNA transcripts corresponds to a biomarker of Table 1, and

b) comparing the level of each of the one or more RNA transcripts in the blood according to step a) with the level of each of the one or more RNA transcripts in blood from one or more individuals not having bladder cancer,

wherein detecting differential expression of each of the one or more RNA transcripts in the comparison of step b) is indicative of bladder cancer in the individual of step a).

Described herein is a method of diagnosing or prognosing bladder cancer in an individual, comprising the steps of:

a) determining the level of one or more RNA transcripts expressed in blood obtained from the individual, wherein thr one or more RNA transcripts corresponds to the one or more biomarkers of Table 4, and

b) comparing the level of each of the one or more RNA transcripts in the blood according to step a) with the level of each of the one or more RNA transcripts in blood from one or more individuals not having bladder cancer,

wherein detecting differential expression of each of the one or more RNA transcripts in the comparison of step b) is indicative of bladder cancer in the individual of step a).

Described herein is a method of diagnosing or prognosing bladder cancer in an individual, comprising the steps of:

a) determining the level of two or more RNA transcripts expressed in blood obtained from the individual, wherein the two or more RNA transcripts corresponds to the one or more biomarkers of Table 4 and Table 5, wherein at least one of the RNA transcripts corresponds to a biomarker of Table 4, and

b) comparing the level of each of the one or more RNA transcripts in the blood according to step a) with the level of each of the one or more RNA transcripts in blood from one or more individuals not having bladder cancer,

wherein detecting differential expression of each of the one or more RNA transcripts in the comparison of step b) is indicative of bladder cancer in the individual of step a).

Described herein is a method of diagnosing or prognosing bladder cancer in an individual, comprising the steps of:

a) determining the level of one or more RNA transcripts expressed in blood obtained from the individual, wherein the one or more RNA transcripts corresponds to the one or more biomarkers of Table 7, and

b) comparing the level of each of the one or more RNA transcripts in the blood according to step a) with the level of each of the one or more RNA transcripts in blood from one or more individuals not having bladder cancer,

wherein detecting differential expression of each of the one or more RNA transcripts in the comparison of step b) is indicative of bladder cancer in the individual of step a).

Described herein is a method of diagnosing or prognosing bladder cancer in an individual, comprising the steps of:

a) determining the level of one or more RNA transcripts expressed in blood obtained from the individual, wherein the one or more RNA transcripts corresponds to the one or more biomarkers of Table 10, and

b) comparing the level of each of the one or more RNA transcripts in the blood according to step a) with the level of each of the one or more RNA transcripts in blood from one or more individuals not having bladder cancer,

wherein detecting differential expression of each of the one or more RNA transcripts in the comparison of step b) is indicative of bladder cancer in the individual of step a).

Described herein is a method of diagnosing or prognosing bladder cancer in an individual, comprising the steps of:

a) determining the level of one or more RNA transcripts expressed in blood obtained from the individual, wherein each the one or more RNA transcripts corresponds to a biomarker selected from the biomarkers listed in Table 1 and

b) comparing the level of each of the one or more RNA transcripts in the blood according to step a) with the level of each of the one or more RNA transcripts in blood from one or more individuals having bladder cancer,

c) comparing the level of each of the one or more RNA transcripts in the blood according to step a) with the level of each of the one or more RNA transcripts in blood from one or more individuals not having bladder cancer,

d) determining whether the level of the one or more RNA transcripts of step a) classify with the levels of the transcripts in step b) as compared with levels of the transcripts in step c),

wherein the determination is indicative of the individual of step a) having bladder cancer.

Described herein is a method of diagnosing or prognosing early stage bladder cancer in an individual, comprising the steps of:

a) determining the level of one or more RNA transcripts expressed in blood obtained from the individual, wherein each the one or more RNA transcripts corresponds to a biomarker selected from the biomarkers listed in Table 4 and

b) comparing the level of each of the one or more RNA transcripts in the blood according to step a) with the level of each of the one or more RNA transcripts in blood from one or more individuals having early stage bladder cancer,

c) comparing the level of each of the one or more RNA transcripts in the blood according to step a) with the level of each of the one or more RNA transcripts in blood from one or more individuals not having bladder cancer,

d) determining whether the level of the one or more RNA transcripts of step a) classify with the levels of the transcripts in step b) as compared with levels of the transcripts in step c),

wherein the determination is indicative of the individual of step a) having early stage bladder cancer.

Described herein is a method of diagnosing or prognosing bladder cancer in an individual, comprising the steps of:

a) determining the level of one or more RNA transcripts expressed in blood obtained from the individual, wherein each the one or more RNA transcripts corresponds to a biomarker selected from the biomarkers listed in Table 7 and

b) comparing the level of each of the one or more RNA transcripts in the blood according to step a) with the level of each of the one or more RNA transcripts in blood from one or more individuals having bladder cancer,

c) comparing the level of each of the one or more RNA transcripts in the blood according to step a) with the level of each of the one or more RNA transcripts in blood from one or more individuals not having bladder cancer,

d) determining whether the level of the one or more RNA transcripts of step a) classify with the levels of the transcripts in step b) as compared with levels of the transcripts in step c),

wherein the determination is indicative of the individual of step a) having bladder cancer.

Described herein is a method of diagnosing or prognosing bladder cancer in an individual, comprising the steps of:

a) determining the level of one or more RNA transcripts expressed in blood obtained from the individual, wherein each the one or more RNA transcripts corresponds to a biomarker selected from the biomarkers listed in Table 10 and

b) comparing the level of each of the one or more RNA transcripts in the blood according to step a) with the level of each of the one or more RNA transcripts in blood from one or more individuals having bladder cancer,

c) comparing the level of each of the one or more RNA transcripts in the blood according to step a) with the level of each of the one or more RNA transcripts in blood from one or more individuals not having bladder cancer,

d) determining whether the level of the one or more RNA transcripts of step a) classify with the levels of the transcripts in step b) as compared with levels of the transcripts in step c),

wherein the determination is indicative of the individual of step a) having bladder cancer.

Described herein is are methods, including methods if diagnosing and/or prognosing bladder cancer, wherein the blood sample consists of whole blood, a drop of blood, or whole blood or a drop of whole blood that has been lysed.

Described herein is are methods, including methods if diagnosing and/or prognosing bladder cancer, which further comprising the step of isolating RNA from the blood samples,

Described herein is are methods, including methods if diagnosing and/or prognosing bladder cancer, which comprise the step of determining the level of each of the one or more RNA transcripts, which may include quantitative RT-PCR (QRT-PCR).

The QRT-PCR in some instances utilizes primers which hybridize to the one or more transcripts or the complement thereof. The primers include those that are 15-25 nucleotides in length.

Described herein is are methods, including methods if diagnosing and/or prognosing bladder cancer, which comprise the step of determining the level of each of the one or more RNA transcripts comprises hybridizing a first plurality of isolated nucleic acid molecules that correspond to the one or more transcripts, to an array comprising a second plurality of isolated nucleic acid molecules.In these methods, the first plurality of isolated nucleic acid molecules comprises RNA, DNA, cDNA, PCR products or ESTs.

Described herein is a kit for diagnosing or prognosing bladder cancer comprising:

a) two biomarker specific priming means designed to produce double stranded DNA complementary to a biomarker selected from any of Tables 1, 4, 7, or 10; 11, or 13 wherein the first priming means contains a sequence which can selectively hybridize to RNA, cDNA or an EST complementary to the biomarker to create an extension product and the second priming means capable of selectively hybridizing to the extension product;

b) an enzyme with reverse transcriptase activity,

c) an enzyme with thermostable DNA polymerase activity, and

d) a labeling means;

wherein the primers are used to detect the quantitative expression levels of the biomarker in a test subject.

Described herein is a method of diagnosing or prognosing bladder cancer in an individual, comprising the steps of:

a) determining the level of one or more proteins expressed in blood obtained from the individual, wherein the one or more proteins is encoded by one or more biomarkers listed in Table 1, and

b) comparing the level of each of the one or more proteins in the blood according to step a) with the level of each of the one or more proteins in blood from one or more individuals not having liver cancer,

wherein detecting a difference in the levels of each of the one or more proteins in the comparison of step b) is indicative of liver cancer in the individual of step a).

Described herein is a method of diagnosing or prognosing bladder cancer in an individual, comprising the steps of:

a) determining the level of one or more proteins expressed in blood obtained from the individual, wherein the one or more proteins is encoded by the one or more biomarkers of Table 1, and

b) comparing the level of each of the one or more proteins in the blood according to step a) with the level of each of the one or more proteins in blood from one or more individuals having bladder cancer,

c) comparing the level of each of the one or more proteins in the blood according to step a) with the level of each of the one or more proteins in blood from one or more individuals not having bladder cancer,

d) determining whether the level of the one or more proteins of step a) classify with the levels of the proteins in step b) as compared with levels of the proteins in step c),

wherein the determination is indicative of the individual of step a) having bladder cancer.

Described herein are methods including a method for diagnosing or prognosing bladder cancer, which comprise the step of determining the level of each of the one or more proteins which comprises the use of one or more antibodies, wherein each of the one or more antibodies is specific for a protein product of a biomarker listed in Table 3.

Described herein are methods including a method for diagnosing or prognosing bladder cancer, which comprise one or more antibodies which can be a monoclonal antibody, fv. scfv, dab, fd, fab, and fab'$_2$.

Described herein are methods of identifying biomarkers and/or combinations of biomarkers useful in the prognosis and/or diagnosis of bladder cancer, comprising comparing expression levels of biomarkers from individuals who are known to have bladder cancer with the levels from individuals who are known not to have bladder cancer, the biomarkers being selected from Table 1 and/or Table 2 or Table 7 or Table 10 or Table 11 and/or Table 12.

Described herein are methods of identifying biomarkers and/or combinations of biomarkers useful in the prognosis and/or diagnosis of early stage bladder cancer, comprising comparing expression levels of biomarkers from individuals who are known to have early stage bladder cancer with the levels from individuals who are known not to have bladder cancer, the biomarkers being selected from Table 3 and/or Table 4.

4. Tables

The objects and features of the invention can be better understood with reference to Tables 1-7 which are included after the Examples Section of the instant specification.

Table 1 is a table showing, in one embodiment of the invention, biomarkers which differentiate as between bladder cancer and normal. Biomarkers are annotated on the basis of their gene ID. Column 1 is the AffySpotID, Column 2 is the p value, Column 3 is the Fold Change (Bladder Cancer/Control), Column 4 is the GeneID, Column 5 is the Gene Symbol, and Column 6 is the Gene Description.

Table 2 is a table showing, in one embodiment of the invention, biomarkers which differentiate as between bladder cancer and normal as previously identified in PCT Patent Application PCT/US04/020836. Biomarkers are annotated on the basis of their gene ID. Number. Column 1 is the Gene ID, Column 2 is the Human RNA Accession Number,

Column 3 is the Human Protein Accession Number, Column 4 is the Gene Symbol, and Column 5 is the Gene Description.

Table 3 is a table showing, in one embodiment, products corresponding to the biomarkers identified in Table 1, including RNA products referred to by the nucleic acid reference accession numbers and protein products referred to by the protein reference accession numbers. Genes are annotated and identified on the basis of their gene ID. Column 1 is AffySpotID, Column 2 is the Gene ID, Column 3 is Human RNA Accession Number, Column 4 is the Human Protein Accession Number, Column 5 is the Gene Symbol, and Column 6 is the Gene Description.

Table 4 is a table showing, in one embodiment of the invention, biomarkers which differentiate as between early stage bladder cancer and normal. Biomarkers are annotated on the basis of their gene ID. Column 1 is AffySpotID, Column 2 is the Gene ID, Column 3 is the Gene Symbol, Column 4 is Human RNA Accession Number, Column 5 is the Human Protein Accession Number, and Column 6 is the Gene Description.

Table 5 is a table showing, in one embodiment of the invention, biomarkers which differentiate as between early stage bladder cancer and normal as previously identified in PCT Patent Application PCT/US04/020836. Biomarkers are annotated on the basis of their gene ID. Column 1 is the Gene ID, Column 2 is the Human RNA Accession Number, Column 3 is the Human Protein Accession Number, Column 4 is the Gene Symbol, and Column 5 is the Gene Description.

Table 6 is a table showing, in one embodiment, products corresponding to the biomarkers identified in Table 4 including RNA products referred to by the nucleic acid reference accession numbers and protein products referred to by the protein reference accession numbers. Genes are annotated and identified on the basis of their gene ID. Column 1 is AffySpotID, Column 2 is the Gene ID, Column 3 is the Human RNA Accession Number, Column 4 is the Human Protein Accession Number, and Column 5 is the Gene Symbol.

Table 7 is a table showing, in one embodiment, biomarkers which are found in common as between Table 1 and Table 4. Column 1 is the AffySpotID, Column 2 is the p

value, Column 3 is the GeneID, Column 4 is the Gene Symbol, Column 5 the Human RNA Accession Number, and Column 5 is the Protein Accession Number.

Table 8 is a table illustrating data that can be used to form an ROC curve based on expression data applied to a mathematical model that uses the logit.

Table 9 is a table listing possible Reporter Genes and the Properties of the Reporter Gene Products.

Table 10 is a table showing, in one embodiment, a particular selection of the biomarkers identified in Table 1 and further annotated in Table 3.

Table 11 is a table showing, in one embodiment, a selection of 349 biomarkers also identified in Table 1 which distinguish as between bladder cancer as compared with either testicular or renal cell carcinoma. Biomarkers are annotated on the basis of their gene ID. Column 1 is the AffySpotID, Column 2 is the Repr RNA Accession Number, Column 3 is the Gene ID, Column 4 is the Gene Symbol, Column 5 is the Fold Change (Bladder Cancer/Control), Column 6 is the Regulation in Disease p value Bladder Cancer/Control (BCvC), Column 7 is the Fold Change (Bladder Cancer vs. Other Cancer), and Column 8 is the Regulation in Bladder Cancer p value (Bladder Cancer vs. Other Cancer).

Table 12 is a table showing, in one embodiment, selected products corresponding to the biomarkers identified in Table 11, including RNA products referred to by the nucleic acid reference accession numbers and protein products referred to by the protein reference accession numbers. Genes are annotated and identified on the basis of their gene ID. Column 1 is the AffySpotID, Column 2 is the Gene ID, Column 3 is the Gene Symbol, Column 4 is the Human RNA Accession Number, and Column 5 is the Human Protein Accession Number.

Table 13 is a table showing, in one embodiment, a particular selection of the biomarkers identified in Table 1.

Table 14 is a table showing accession numbers to the sequences of human RNA products and human protein products corresponding to the selected biomarkers in **Table 13.**

Table 15 is selection of biomarker specific primers to the biomarkers listed in **Table 13** and which were used in accordance with Example 4.

Table 16 shows the results of quantitative real time RT-PCR (QRT-PCR) of the RNA products of the biomarkers listed in Table 13 as described in Example 4.

Table 17, in one embodiment of the invention, is a demonstration of isolated antibodies which selectively bind to the protein products of the biomarkers listed in **Table 13** and which are commercially available.

Table 18, in one embodiment of the invention, is a selection of biomarker specific primers and corresponding biomarker specific probes which selectively amplify double stranded DNA complementary to the noted RNA products which are products of the biomarkers noted in **Table 13.**

Table 19, in one embodiment of the invention, shows selected 2 gene combinations which can be used to screen for bladder cancer.

Table 20, shows one embodiment of the invention of a collection of classifiers utilizing combinations of four ratios of biomarkers where the biomarkers are selected from those biomarkers in **Table 13.**

Table 21, shows one embodiment of the invention of a collection of classifiers utilizing combinations of three ratios of biomarkers where the biomarkers are selected from those biomarkers in **Table 13.**

Table 22, shows one embodiment of the invention of a collection of classifiers utilizing combinations of two ratios of biomarkers where the biomarkers are selected from those biomarkers in **Table 13.**

Table 23, shows one embodiment of the invention of a collection of classifiers utilizing combinations of ratios of biomarkers where the biomarkers are selected from those biomarkers in **Table 13.**

5

5. Detailed Description of the Invention

The practice of the present invention employs in part conventional techniques of molecular biology, microbiology and recombinant DNA techniques, which are within the skill of the art. Such techniques are explained fully in the literature. See, e.g., Sambrook, Fritsch & Maniatis, 1989, Molecular Cloning: A Laboratory Manual, Second Edition ; Oligonucleotide Synthesis (M.J. Gait, ed., 1984); Nucleic Acid Hybridization (B.D. Harnes & S.J. Higgins, eds., 1984); A Practical Guide to Molecular Cloning (B. Perbal, 1984); and a series, Methods in Enzymology (Academic Press, Inc.); Short Protocols In Molecular Biology, (Ausubel et al., ed., 1995). All patents, patent applications, and publications mentioned herein, both supra and infra, are hereby incorporated by reference in their entireties.

The invention as disclosed herein identifies biomarkers and biomarker combinations useful in diagnosing bladder cancer and/or early stage bladder cancer. In order to use these biomarkers, the invention teaches the identification of the products of these biomarkers including the RNA products and the protein products. The invention further encompasses the oligonucleotides, cDNA, DNA, RNA, PCR products, synthetic DNA, synthetic RNA, and fragments thereof, or other combinations of naturally occurring modified nucleotides that specifically and/or selectively hybridize to the RNA products of the biomarkers of the invention. The invention further discloses proteins, peptides, antibodies, ligands, and fragments thereof including antigen binding fragments that specifically and/or selectively hybridize to the protein products of the biomarkers of the invention. The measuring of the expression of the RNA product(s) of the biomarkers and combination of biomarkers of the invention, can be done by using those polynucleotides which are specific and/or selective for the RNA product(s) of the biomarkers of the invention to quantitate the

expression of the RNA product(s). In a specific embodiment of the invention, the polynucleotides which are specific and/or selective for the RNA products are probes or primers. In one embodiment, these polynucleotides are in the form of a nucleic acid probes which can be hybridized to a manufactured array. In another embodiment, commercial arrays can be used to measure the expression of the RNA product and the invention teaches which combination of genes to analyze. In another embodiment, the polynucleotides which are specific and/or selective for the RNA products of the biomarkers of the invention are used in the form of probes and primers in techniques such as quantitative real-time RT PCR, using for example SYBR®Green, or using TaqMan® or Molecular Beacon techniques, where the polynucleotides used are used in the form of a forward primer, a reverse primer, a TaqMan labelled probe or a Molecular Beacon labelled probe. In one specific embodiment, the results generated from measuring the level of expression of the RNA products of the invention can be input into a model of the invention which is used to identify the combinations of biomarkers to determine a diagnosis as defined by the model. In a preferred embodiment, the same method is used to generate the expression data used to generate the mathematical model as is used to diagnose the test individual.

The invention further contemplates the use of proteins or polypeptides as disclosed herein and would be known by a person skilled in the art to measure the protein products of the biomarkers of the invention. Techniques known to persons skilled in the art (for example, techniques such as Western Blotting, Immunoprecipitation, protein microarray analysis and the like) can then be used to measure the level of protein products corresponding to the biomarkers of the invention. As would be understood to a person skilled in the art, the measure of the level of expression of the protein products of the biomarkers of the invention requires a protein which specifically or selectively binds to one or more of the protein products corresponding to each biomarker of the invention. Data representative of the level of expression of the protein products of the biomarker of the invention can then be input into the model generated to identify the combination in order to determine a diagnosis as defined by the model. In a preferred embodiment, the same method is used to generate the expression data used to generate the mathematical model as is used to diagnose the test individual.

5.1 Samples for Use in the Invention

Unless otherwise indicated herein, a bladder tissue sample from a subject may be used in accordance with the methods of the invention, as can a blood sample, a serum sample, and a lymph node sample. Examples of subjects from which such a sample may be obtained and utilized in accordance with the methods of the invention include, but are not limited to, asymptomatic subjects, subjects manifesting or exhibiting 1, 2, 3, 4 or more symptoms of bladder cancer, subjects clinically diagnosed as having bladder cancer and or early stage bladder cancer, subjects predisposed to bladder cancer (*e.g.*, subjects with a family history of bladder cancer, subjects with a genetic predisposition to bladder cancer, and subjects that lead a lifestyle that predisposes them to bladder cancer or increases the likelihood of developing bladder cancer), subjects suspected of having bladder cancer, subjects undergoing therapy for bladder related disorders, subjects with bladder cancer and at least one other condition (*e.g.*, subjects with 2, 3, 4, 5 or more conditions), subjects not undergoing therapy for bladder cancer, subjects determined by a medical practitioner (*e.g.*, a physician) to be healthy or bladder cancers-free (*i.e.*, normal), subjects that have been treated for bladder cancer including subjects which have been treated successfully for bladder cancer, subjects that are managing their bladder cancer and subjects that have not been diagnosed with bladder cancer.

In another embodiment, the subjects from which a sample may be obtained and utilized have early stage bladder cancer. In a further embodiment, the subject from which a sample may be obtained is a test individual wherein it is unknown whether the person has bladder cancer, and/or it is unknown what stage of bladder cancer the test individual has.

5.1.1 Tissue

In one aspect, a tissue sample is obtained is obtained using any known method from one or more subjects. In a specific embodiment, tissue is obtained from individuals suspected of having bladder cancer. Preferably, the tissue samples are stored in liquid nitrogen until needed. In a specific embodiment, a minimum of 0.05 g of tissue sample is isolated to obtain total RNA. In another embodiment, a minimum of 0.025 g tissues sample is isolated to obtain total RNA. A tissue sample that is useful according to the invention is in

an amount that is sufficient for the detection of one or more nucleic acid sequences or amino acid sequences according to the invention.

5.1.2 Blood

In one aspect of the invention, a sample of blood is obtained from a subject according to methods well known in the art. A sample of blood may be obtained from any subject as described herein. In some embodiments, a drop of blood is collected from a simple pin prick made in the skin of a subject. A drop of blood, as disclosed herein encompasses volumes of less than 1 ml. In some embodiments, the drop of blood collected from a pin prick is all that is needed. Blood may be drawn from a subject from any part of the body (*e.g.*, a finger, a hand, a wrist, an arm, a leg, a foot, an ankle, a stomach, and a neck) using techniques known to one of skill in the art, in particular methods of phlebotomy known in the art.

The amount of blood collected will vary depending upon the site of collection, the amount required for a method of the invention, and the comfort of the subject. However, an advantage of one embodiment of the present invention is that the amount of blood required to implement the methods of the present invention can be so small that more invasive procedures are not required to obtain the sample. For example, in some embodiments, all that is required is a drop of blood. This drop of blood can be obtained, for example, from a simple pinprick. In some embodiments, any amount of blood is collected that is sufficient to detect the expression of one, two, three, four, five, ten or more genes listed in any of Table 1-7, or 10-12. As such, in some embodiments, the amount of blood that is collected is 1 ml or less, 500 μl or less, 250 μl or less, 200 μl or less, 100 μl or less, 50 μl or less, 20 μl or less, 10 μl or less, 1 μl or less, 0.5 μl or less, 0.1 μl or less, or 0.01 μl or less. However, the present invention is not limited to such embodiments. In some embodiments more blood is available and in some embodiments, more blood can be used to effect the methods of the present invention. As such, in various specific embodiments, 0.001 ml, 0.005 ml, 0.01 ml, 0.05 ml, 0.1 ml, 0.15 ml, 0.2 ml, 0.25 ml, 0.5 ml, 0.75 ml, 1 ml, 1.5 ml, 2 ml, 3 ml, 4 ml, 5 ml, 10 ml, 15 ml or more of blood is collected from a subject. In another embodiment, 0.001 ml to 15 ml, 0.01 ml to 10 ml, 0.1 ml to 10 ml, 0.1 ml to 5 ml, 1 to 5 ml of blood is collected from a subject.

In some embodiments of the present invention, blood is stored within a K3/EDTA tube. In another embodiment, one can utilize tubes for storing blood which contain stabilizing agents such as disclosed in U.S. Patent No. 6,617,170 (which is incorporated herein by reference). In another embodiment the PAXgeneTM blood RNA system provided by PreAnalytiX, a Qiagen/BD company may be used to collect blood. In yet another embodiment, the TempusTM blood RNA collection tubes, offered by Applied Biosystems may be used. Tempus™ collection tubes provide a closed evacuated plastic tube containing RNA stabilizing reagent for whole blood collection.

The collected blood collected is optionally but preferably stored at refrigerated temperatures, such as 4 °C or on ice, prior to use in accordance with the methods of the invention. In some embodiments, the blood is maintained at 4 °C or on ice for only 1 hour, 2 hours, 3 hours, 4 hours 5 hours, 6 hours, 7 hours, 8 hours 10 hours or up to 24 hours. In some embodiments, in addition to storage of the blood or instead of storage of the blood, isolated nucleic acid or proteins are stored for a period of time for later use. Storage of such biomarkers can be for an hour or more, a day or more, a week or more, a month or more, a year or more, or indefinitely.

In one aspect, whole blood is obtained from a normal individual or from an individual diagnosed with, or suspected of having osteoarthritis according the methods of phlebotomy well known in the art. Whole blood includes unfractionated whole blood which includes blood wherein the serum or plasma has been removed and the RNA or mRNA from the remaining blood sample has been isolated in accordance with methods well known in the art (*e.g.*, using, preferably, gentle centrifugation at 300 to 800 xg for 5 to 10 minutes). In a specific embodiment, unfractionated whole blood also encompasses whole blood treated by mixing the blood with lysing buffer (*e.g.*, Lysis Buffer (1L): 0.6g EDTA; 1.0g $KHCO_2$, 8.2g NH_4Cl adjusted to pH 7.4 (using NaOH)), the sample is centrifuged and the cell pellet retained, and RNA or mRNA extracted in accordance with methods known in the art (e.g. "lysed whole blood") (see for example Sambrook et al.). The use of whole blood is preferred since it avoids the costly and time-consuming need to separate out the cell types within the blood (Kimoto, 1998, Mol. Gen. Genet 258:233-239; Chelly J et al., 1989, Proc. Nat. Acad. Sci. USA 86:2617-2621; Chelly J et al., 1988, Nature 333:858-860).

In some embodiments of the present invention, whole blood collected from a subject is fractionated (*i.e.*, separated into components). In specific embodiments of the present invention, blood cells are separated from whole blood collected from a subject using techniques known in the art. For example, blood collected from a subject can be subjected to Ficoll-Hypaque (Pharmacia) gradient centrifugation. Such centrifugation separates erythrocytes (red blood cells) from various types of nucleated cells and from plasma. In particular, Ficoll-Hypaque gradient centrifugation is useful to isolate peripheral blood leukocytes (PBLs) which can be used in accordance with the methods of the invention.

By way of example but not limitation, macrophages can be obtained as follows. Mononuclear cells are isolated from peripheral blood of a subject, by syringe removal of blood followed by Ficoll-Hypaque gradient centrifugation. Tissue culture dishes are pre-coated with the subject's own serum or with AB+ human serum and incubated at 37°C for one hour. Non-adherent cells are removed by pipetting. Cold (4°C) 1 mM EDTA in phosphate-buffered saline is added to the adherent cells left in the dish and the dishes are left at room temperature for fifteen minutes. The cells are harvested, washed with RPMI buffer and suspended in RPMI buffer. Increased numbers of macrophages can be obtained by incubating at 37°C with macrophage-colony stimulating factor (M-CSF). Antibodies against macrophage specific surface markers, such as Mac-1, can be labeled by conjugation of an affinity compound to such molecules to facilitate detection and separation of macrophages. Affinity compounds that can be used include but are not limited to biotin, photobiotin, fluorescein isothiocyante (FITC), or phycoerythrin (PE), or other compounds known in the art. Cells retaining labeled antibodies are then separated from cells that do not bind such antibodies by techniques known in the art such as, but not limited to, various cell sorting methods, affinity chromatography, and panning.

Blood cells can be sorted using a using a fluorescence activated cell sorter (FACS). Fluorescence activated cell sorting (FACS) is a known method for separating particles, including cells, based on the fluorescent properties of the particles. See, for example, Kamarch, 1987, Methods Enzymol 151:150-165. Laser excitation of fluorescent moieties in the individual particles results in a small electrical charge allowing electromagnetic separation of positive and negative particles from a mixture. An antibody or

ligand used to detect a blood cell antigenic determinant present on the cell surface of particular blood cells is labeled with a fluorochrome, such as FITC or phycoerythrin. The cells are incubated with the fluorescently labeled antibody or ligand for a time period sufficient to allow the labeled antibody or ligand to bind to cells. The cells are processed through the cell sorter, allowing separation of the cells of interest from other cells. FACS sorted particles can be directly deposited into individual wells of microtiter plates to facilitate separation.

Magnetic beads can be also used to separate blood cells in some embodiments of the present invention. For example, blood cells can be sorted using a using a magnetic activated cell sorting (MACS) technique, a method for separating particles based on their ability to bind magnetic beads (0.5-100 m diameter). A variety of useful modifications can be performed on the magnetic microspheres, including covalent addition of an antibody which specifically recognizes a cell-solid phase surface molecule or hapten. A magnetic field is then applied, to physically manipulate the selected beads. In a specific embodiment, antibodies to a blood cell surface marker are coupled to magnetic beads. The beads are then mixed with the blood cell culture to allow binding. Cells are then passed through a magnetic field to separate out cells having the blood cell surface markers of interest. These cells can then be isolated.

In some embodiments, the surface of a culture dish may be coated with antibodies, and used to separate blood cells by a method called panning. Separate dishes can be coated with antibody specific to particular blood cells. Cells can be added first to a dish coated with blood cell specific antibodies of interest. After thorough rinsing, the cells left bound to the dish will be cells that express the blood cell markers of interest. Examples of cell surface antigenic determinants or markers include, but are not limited to, CD2 for T lymphocytes and natural killer cells, CD3 for T lymphocytes, CD11a for leukocytes, CD28 for T lymphocytes, CD19 for B lymphocytes, CD20 for B lymphocytes, CD21 for B lymphocytes, CD22 for B lymphocytes, CD23 for B lymphocytes, CD29 for leukocytes, CD14 for monocytes, CD41 for platelets, CD61 for platelets, CD66 for granulocytes, CD67 for granulocytes and CD68 for monocytes and macrophages.

Whole blood can be fractionated into cells types such as leukocytes, platelets, erythrocytes, *etc.* and such cell types can be used in accordance with the methods of the invention. Leukocytes can be further separated into granulocytes and agranulocytes using standard techniques and such cells can be used in accordance with the methods of the invention. Granulocytes can be separated into cell types such as neutrophils, eosinophils, and basophils using standard techniques and such cells can be used in accordance with the methods of the invention. Agranulocytes can be separated into lymphocytes (*e.g.*, T lymphocytes and B lymphocytes) and monocytes using standard techniques and such cells can be used in accordance with the methods of the invention. T lymphocytes can be separated from B lymphocytes and helper T cells separated from cytotoxic T cells using standard techniques and such cells can be used in accordance with the methods of the invention. Separated blood cells (*e.g.*, leukocytes) can be frozen by standard techniques prior to use in the present methods.

A blood sample that is useful according to the invention is in an amount that is sufficient for the detection of one or more nucleic acid or amino acid sequences according to the invention. In a specific embodiment, a blood sample useful according to the invention is in an amount ranging from 1 μl to 100 ml, preferably 10 μl to 50 ml, more preferably 10 μl to 25 ml and most preferably 10 μl to 1 ml.

5.1.3 Serum

In one aspect of the invention, serum is obtained from a subject by first isolating a blood sample, and subsequently spinning the sample so as to separate the serum from the remainder of the blood in accordance with methods well known in the art. A serum sample may be obtained from any subject as described herein. The amount of serum collected will vary depending upon the site of collection of th eblood, the amount required for a method of the invention, and the comfort of the subject. However, an advantage of one embodiment of the present invention is that the amount of blood required to provide sufficient serum to implement the methods of the present invention can be so small that more invasive procedures are not required to obtain the sample. For example, in some embodiments, all that is required is a drop of blood and the amount of serum that is collected is 1 ml or less, 500 μl or less, 250 μl or less, 200 μl or less, 100 μl or less, 50 μl or less, 20 μl

or less, 10 μl or less, 1 μl or less, 0.5 μl or less, 0.1 μl or less, or 0.01 μl or less. However, the present invention is not limited to such embodiments. In some embodiments more blood is available and in some embodiments, more blood can be used to obain sufficient serum to effect the methods of the present invention.

5.2 RNA Preparation

In one aspect of the invention, RNA is isolated from an individual in order to measure the RNA products of the biomarkers of the invention. RNA is isolated from tissue samples as described herein. Samples can be from a single patient or can be pooled from multiple patients.

In another aspect, RNA is isolated directly from blood samples of persons with osteoarthritis as described herein. Samples can be from a single patient or can be pooled from multiple patients.

Total RNA is extracted from the tissue samples according to methods well known in the art. In one embodiment, RNA is purified from tissue according to the following method. Following the removal of a tissue of interest from an individual or patient, the tissue is quick frozen in liquid nitrogen, to prevent degradation of RNA. Upon the addition of a volume of tissue guanidinium solution, tissue samples are ground in a tissuemizer with two or three 10-second bursts. To prepare tissue guanidinium solution (1 L) 590.8 g guanidinium isothiocyanate is dissolved in approximately 400 ml DEPC-treated H_2O. 25 ml of 2 M Tris-Cl, pH 7.5 (0.05 M final) and 20 ml Na_2EDTA (0.01 M final) is added, the solution is stirred overnight, the volume is adjusted to 950 ml, and 50 ml 2-ME is added.

Homogenized tissue samples are subjected to centrifugation for 10 min at 12,000 x g at 12°C. The resulting supernatant is incubated for 2 min at 65°C in the presence of 0.1 volume of 20% Sarkosyl, layered over 9 ml of a 5.7M CsCl solution (0.1g CsCl/ml), and separated by centrifugation overnight at 113,000 x g at 22°C. After careful removal of the supernatant, the tube is inverted and drained. The bottom of the tube (containing the RNA pellet) is placed in a 50 ml plastic tube and incubated overnight (or longer) at 4°C in the presence of 3 ml tissue resuspension buffer (5 mM EDTA, 0.5% (v/v) Sarkosyl, 5% (v/v) 2-

ME) to allow complete resuspension of the RNA pellet. The resulting RNA solution is extracted sequentially with 25:24:1 phenol/chloroform/isoamyl alcohol, followed by 24:1 chloroform/isoamyl alcohol, precipitated by the addition of 3 M sodium acetate, pH 5.2, and 2.5 volumes of 100% ethanol, and resuspended in DEPC water (Chirgwin et al., 1979, *Biochemistry*, 18:5294).

Alternatively, RNA is isolated from tissue according to the following single step protocol. The tissue of interest is prepared by homogenization in a glass teflon homogenizer in 1 ml denaturing solution (4M guanidinium thiosulfate, 25 mM sodium citrate, pH 7.0, 0.1M 2-ME, 0.5% (w/v) N-laurylsarkosine) per 100mg tissue. Following transfer of the homogenate to a 5-ml polypropylene tube, 0.1 ml of 2 M sodium acetate, pH 4, 1 ml water-saturated phenol, and 0.2 ml of 49:1 chloroform/isoamyl alcohol are added sequentially. The sample is mixed after the addition of each component, and incubated for 15 min at 0-4°C after all components have been added. The sample is separated by centrifugation for 20 min at 10,000 x g, 4°C, precipitated by the addition of 1 ml of 100% isopropanol, incubated for 30 minutes at -20°C and pelleted by centrifugation for 10 minutes at 10,000 x g, 4°C. The resulting RNA pellet is dissolved in 0.3 ml denaturing solution, transferred to a microfuge tube, precipitated by the addition of 0.3 ml of 100% isopropanol for 30 minutes at -20°C, and centrifuged for 10 minutes at 10,000 x g at 4°C. The RNA pellet is washed in 70% ethanol, dried, and resuspended in 100-200 µl DEPC-treated water or DEPC-treated 0.5% SDS (Chomczynski and Sacchi, 1987, *Anal. Biochem.*, 162:156).

Preferably, the tissue samples are finely powdered under liquid nitrogen and total RNA is extracted using TRIzol® reagent (GIBCO/BRL).

Alternatively, RNA is isolated from blood. In one embodiment, RNA is isolated by the following protocol. Lysis Buffer is added to blood sample in a ratio of 3 parts Lysis Buffer to 1 part blood (Lysis Buffer (1L) 0.6g EDTA; 1.0g $KHCO_2$, 8.2g NH_4Cl adjusted to pH 7.4 (using NaOH)). Sample is mixed and placed on ice for 5-10 minutes until transparent. Lysed sample is centrifuged at 1000 rpm for 10 minutes at 4°C, and supernatant is aspirated. Pellet is resuspended in 5 ml Lysis Buffer, and centrifuged again at 1000 rpm for 10 minutes at 4°C. Pelleted cells are homogenized using TRIzol® (GIBCO/BRL) in a ratio of approximately 6 ml of TRIzol® for every 10 ml of the original blood sample and

vortexed well. Samples are left for 5 minutes at room temperature. RNA is extracted using 1.2 ml of chloroform per 1 ml of TRIzol®. Sample is centrifuged at 12,000 x g for 5 minutes at 4°C and upper layer is collected. To upper layer, isopropanol is added in ratio of 0.5 ml per 1 ml of TRIzol®. Sample is left overnight at -20°C or for one hour at -20°C. RNA is pelleted in accordance with known methods, RNA pellet air dried, and pellet resuspended in DEPC treated ddH_2O. RNA samples can also be stored in 75% ethanol where the samples are stable at room temperature for transportation.

Alternatively, RNA is isolated from synovial fluid using TRIzol® reagent (GIBCO/BRL) as above.

Purity and integrity of RNA is assessed by absorbance at 260/280nm and agarose gel electrophoresis followed by inspection under ultraviolet light.

5.3 Biomarkers of the Invention

In one embodiment, the invention provides biomarkers and biomarker combinations wherein the measure of the level of expression of the product or products of said biomarkers is indicative of bladder cancer. In another embodiment, the invention provides biomarkers and biomarker combinations, wherein the measure of the level of expression of the product or products of said biomarkers can be used to diagnose whether an individual has bladder cancer and/or early stage bladder cancer.

Table 1 provides a list of the gene names and the associated gene ID (formerly locus link ID)for the biomarkers of the invention wherein the measure of the level of expression of the biomarkers, either individually, or in combination, can be used to diagnose an individual as having bladder cancer; or not having bladder cancer. As would be understood by a person skilled in the art, the gene ID can be used to determine the sequence of all the RNA transcripts and all of the proteins which correspond to the biomarkers listed.

Table 2 provides biomarkers disclosed in PCT Application Number PCT/US04/020836 which can be used in combination with one or more of the biomarkers disclosed in Table 1 as taught herein to diagnose bladder cancer or to differentiate as between a person having bladder cancer or not having bladder cancer.

Table 3 in particular discloses reference accession numbers corresponding to the RNA products of the biomarkers and reference accession numbers corresponding to the protein products of the biomarkers listed in Table 1.

Table 4 provides a list of the gene names and the associated gene ID for the biomarkers of the invention wherein the measure of the level of expression of the biomarkers, either individually, or in combination, can be used to diagnose an individual as having early stage bladder cancer; or not having bladder cancer. As would be understood by a person skilled in the art, the gene ID can be used to determine the sequence of all the RNA transcripts and all of the proteins which correspond to the biomarkers listed.

Table 5 provides biomarkers disclosed in PCT Application Number PCT/US04/020836 which can be used in combination with one or more of the biomarkers disclosed in Table 4 as taught herein to diagnose early stage bladder cancer or to differentiate as between a person having early stage bladder cancer or not having bladder cancer.

Table 6 in particular discloses reference accession numbers corresponding to the RNA products of the biomarkers and reference accession numbers corresponding to the protein products of the biomarkers listed in Table 4.

Table 7 in particular discloses a list of the gene names and the associated gene ID for the biomarkers of the invention which differentiate bladder cancer from non bladder cancer.

Table 10 in particular discloses a list of the gene names and associated gene IDs for a selection of biomarkers identified in Table 1 which differentiate bladder cancer from non bladder cancer.

Table 11 provides a list of the gene names and the associated gene ID (formerly locus link ID)for a selection of biomarkers of the invention wherein the measure of the level of expression of the biomarkers, either individually, or in combination, can be used to diagnose an individual as having bladder cancer; or not having bladder cancer. As would be understood by a person skilled in the art, the gene ID can be used to determine the sequence of all the RNA transcripts and all of the proteins which correspond to the biomarkers listed.

Table 12 in particular discloses reference accession numbers corresponding to the RNA products of the biomarkers and reference accession numbers corresponding to the protein products of the biomarkers listed in Table 11.

The invention thus encompasses the use of those methods known to a person skilled in the art and outlined herein to measure the expression of these biomarkers and combinations of biomarkers for each of the purposes outlined above.

5.4 Combinations of Biomarkers

In one embodiment, combinations of biomarkers of the present invention includes any combination of any number up to 2, 3, 4, 5, 6, 7, 8, 10, 20, 30, 40, 50, 100, 125, 150, 175, 200 or all of the biomarkers listed in Table 1. In another embodiment of the invention, combinations of biomarkers of the present invention include any combination of any one or any number up to 1, 2, 3, 4, 5, 6, 7, 8, 10, 20, 30, 40, 50, 100 125, 150, 175, 200 or all of the biomarkers listed in Table 1 in combination with any one or any number up to 1, 2, 3, 4, 5, 6, 7, 8, 10, 20, 30, 40, 50, 100, 250, 500, 750, 1000, 1500, 2000, 2500, 3000, 3500, 4000 or all of the biomarkers listed in Table 2, the measurement of expression of the products of which can be used for diagnosing whether an individual has bladder cancer or does not have bladder cancer. In another embodiment, combinations of biomarkers of the present invention includes any combination of any number up to 2, 3, 4, 5, 6, 7, 8, 10, 20, 30, 40, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 1000, 1250, 1500, 1750, 2000, or all of the RNA and/or protein products listed in Table 3. In another embodiment, combinations of biomarkers of the present invention includes any combination of any number up to 2, 3, 4, 5, 6, 7, 8, 10, 20, 30, 40, 50, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500 or all of the biomarkers listed in Table 4. In another embodiment of the invention, combinations of biomarkers of the present invention include any combination of any one or any number up to 1, 2, 3, 4, 5, 6, 7, 8, 10, 20, 30, 40, 50, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500 or all of the biomarkers listed in Table 4 in combination with any one or any number up to 1, 2, 3, 4, 5, 6, 7, 8, 10, 20, 30, 40, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 1000, 1250, 1500, 1750, 2000, 2250, or all of the biomarkers listed in Table 5, the measurement of expression of the products of which can be used for diagnosing whether an individual has early bladder cancer or does not have bladder cancer. In another embodiment, combinations

of biomarkers of the present invention includes any combination of any number up to 2, 3, 4, 5, 6, 7, 8, 10, 20, 30, 40, 50, 100 150, 200, 250, 300, 350, 400, 450, 500, 1000, 1250, 1500, 1750, 2000, 2250, 2500, 2750, 3000, 3250, or all of the RNA and/or protein products listed in Table 6. In another embodiment of the invention, combinations of biomarkers of the present invention include any combination of any one or any number up to 1, 2, 3, 4, 5, 6, 7, 8, 10, 20, 30, 40, 50, 75, 100, 125, 150, 175, 200 or all of the biomarkers listed in Table 7. .In another embodiment, combinations of biomarkers of the present invention includes any combination of any number up to 2, 3, 4, 5, 6, 7, 8, 9 or all of the biomarkers listed in Table 10.

For instance, the number of possible combinations of a subset *m* of *n* genes is described in Feller, *Intro to Probability Theory*, Third Edition, volume 1, 1968, ed. J. Wiley, using the general formula:

$$m!/(n)!\ (m\text{-}n)!$$

In one embodiment, where *n* is 2 and *m* is 19, there are:

$$\frac{19!}{2!\,(19-2)!} = \frac{19 \times 18 \times 17 \times 16 \times 15 \times 14 \times 13 \times 12 \times 11 \times 10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}{(2 \times 1)\,(19 \times 18 \times 17 \times 16 \times 15 \times 14 \times 13 \times 12 \times 11 \times 10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1)}$$

$$= \frac{1.216\ 10^{17}}{7.11\ 10^{14}} = 171$$

unique two-gene combinations. The measurement of the gene expression of each of these two-gene combinations can independently be used to determine whether a patient has bladder cancer. In another specific embodiment in which *m* is 19 and *n* is three, there are 19!/3!(19-3)! unique three-gene combinations. Each of these unique three-gene combinations can independently serve as a model for determining whether a patient has bladder cancer.

5.5 Selection of Particularly Useful Combinations of Biomarkers

Although combinations of the biomarkers in any of the Table 1-7, and10-13 of the invention are useful for diagnosing bladder cancer and/or early bladder cancer respectively, the invention further provides a means of selecting and evaluating combinations

of biomarkers from any particular Table or combination of Tables which can be used to diagnose bladder cancer.

In order to identify useful combinations of biomarkers a mathematical model is used to develop a classifier which is able to differentiate as between bladder cancer and/or early bladder cancer and non bladder cancer using the data reflective of the level of expression of each biomarker of a combination of biomarkers where each biomarker is selected from those biomarkers identified in one or more of the Tables 1-7 and 10-13. Classifiers are then evaluated or scored for their ability to differentiate as between bladder cancer and/or early bladder cancer and non bladder cancer to select those classifiers (and thus those combinations of biomarkers) which are best able to diagnose an individual as having bladder cancer and/or early bladder cancer.

Classifiers are created by using data representative of the level of expression of the product of two or more selected biomarkers for each individual within a training population and applying a mathematical model. The phenotypic characteristics of the training population is important in order to ensure that the resulting classifier has the desired utility. For example, in one embodiment, the training population used is comprised of two phenotypic subgroups - individuals who are known to have bladder cancer, and individuals who are known not to have bladder cancer and the biomarkers are selected from Table 1 and/or Table 2 or Table 7 or Table 10 or Table 11 and/or Table 12. In another embodiment, the training population used is comprised of two phenotypic subgroups - individuals who are known to have early stage bladder cancer, and individuals who are known not to have bladder cancer and the biomarkers are selected from Table 3 and/or Table 4.

The data which is input into the mathematical model can be any data which is representative of the expression level of the product of a biomarker. In one embodiment, the data input into the mathematical model for each biomarker is data obtained using any method known to a person skilled in the art to measure the level of gene expression including measuring the RNA product of the biomarker and/or measuring the protein product of the biomarker. In a preferred embodiment, the data input into the mathematical model is data obtained using quantitative real-time PCR (qRT-PCR).

In order to identify particularly useful combinations of biomarkers, it is possible to develop a classifier for each possible combination of biomarkers for any one of Table 1, Table 2, Table 1 and Table 2 combined, Table 4, Table 5, Table 4 and Table 5 combined, Table 7 or Table 10 can and then score the classifiers so as to select from the scored classifiers those which are the most useful for diagnosing bladder cancer or early bladder cancer.

In order to develop classifiers, a mathematical model is used. Mathematical models which are useful in accordance with the invention can be selected from the following: a regression model, a logistic regression model, a neural network, a clustering model, principal component analysis, nearest neighbour classifier analysis, linear discriminant analysis, quadratic discriminant analysis, a support vector machine, a decision tree, a genetic algorithm, classifier optimization using bagging, classifier optimization using boosting, classifier optimization using the Random Subspace Method, a projection pursuit, and weighted voting.

The classifier generated are subsequently scored and/or ranked as outlined in section 5 so as to identify those classifiers of sufficient sensitivity and/or specificity to diagnose individuals as having bladder cancer and/or early bladder cancer. In order to score the classifiers, classifiers can be tested to determine the ability of the classifier to correctly call each individual of the training population as described herein. The classifier generated can also be evaluated or scored by determining the ability of the classifier to correctly call one or more individuals of a "scoring population". The scoring population is similar to the training population described in Section 5.7 below, however the scoring population is made up of one or more individuals not used to generate the classifier. As such the scoring population is comprised of individuals who have already been diagnosed as having e.g. bladder cancer (the first phenotypic subgroup) and individuals not having bladder cancer (the second phenotypic subgroup). In one embodiment, the scoring population includes members of the training population in addition to one or more members not used in the training population. In some embodiments, five percent or less, ten percent or less, twenty percent or less, thirty percent or less, fifty percent or less, or ninety percent or less of the members of the training population are common to the scoring population.

As would be understood by a person skilled in the art, this allows one to predict the ability of the classifiers to properly characterize an individual whose phenotypic characterization is unknown.

The resulting classifiers can be used to diagnosis an unknown or test individual to determine whether said test individual has bladder cancer and/or early stage bladder cancer. In one embodiment, two or more classifiers ("classifier group") can be used determine the diagnosis of a test individual. For instance, in some embodiments, the top 10 ranking classifiers, the top 20 ranking classifiers, or the top 100 ranking classifiers are selected. In some embodiments, any of the top 1 to top 500 ranking classifiers is selected. In instances classifier groups are used for diagnosis, in one embodiment each classifier contributes one vote to the diagnosis of the test subject such that diagnosis of the test subject is determined as a result of a combination of classifiers. In other embodiments, different weighting schema applied to each classifier. For example, weighting schema can include weighting on the basis of factors such as the original score the classifier, the logs odd ratio ("logit"), the size of the coefficients for each classifier, some combination thereof and the like.

In one embodiment, the classifier or classifier group is used directly for the diagnosing as described above. In another embodiment however, the combination identified can be used independently of the classifier which identified the genes. For example, the gene expression profile resulting from the 10 genes can be monitored to evaluate a test individual wherein the gene expression profile of the test individual is compared to the gene expression profile of the 10 genes from individuals having bladder cancer and a gene expression profile of the 10 genes in individuals not having bladder cancer.

5.6 Data for input into Mathematical Models to Select Particularly Useful Biomarker Combinations for Diagnosis of Bladder Cancer and Early Stage Bladder Cancer

For example, in order to identify those biomarkers which are useful in diagnosing an individual as having bladder cancer, or not having bladder cancer, data reflective of the level of expression of one or more of the mRNA products of the biomarkers

of Table 1 and/or Table 2 and/or Table 7 and/or Table 10 and/or Table 11 are used within a training population of individuals, the first phenotypic subgroup having bladder cancer, and a second phenotypic subgroup of individuals not having bladder cancer. For purposes of characterizing the training population into the prescribed phenotypic subgroups, any method of bladder cancer diagnosis can be used. In a preferred embodiment, a cytoscopy is utilized to determine diagnosis of bladder cancer for purposes of the training population.

In another embodiment, in order to identify those biomarkers which are useful in diagnosing an individual as having early stage bladder cancer, or not having bladder cancer, data reflective of the level of expression of one or more of the mRNA products of the biomarkers of Table 4 and/or Table 5 from individuals within a training population comprised of individuals having early stage bladder cancer and individuals not having bladder cancer.

In another embodiment data reflective of the level of expression of one or more protein products of the biomarkers of Table 1 and/or Table 2 and/or Table 7 and/or Table 10 are used in a training population of individuals, the first phenotypic subgroup having bladder cancer, and a second phenotypic subgroup of individuals not having bladder cancer. For purposes of characterizing the training population into the prescribed phenotypic subgroups, any method of bladder cancer diagnosis can be used.

In another embodiment data reflective of the level of expression of one or more protein products of the biomarkers of Table 1 and/or Table 2 and/or Table 7 and/or Table 10 and/or Table 11 are used in a training population of individuals, the first phenotypic subgroup having bladder cancer, and a second phenotypic subgroup of individuals not having bladder cancer. For purposes of characterizing the training population into the prescribed phenotypic subgroups, any method of bladder cancer diagnosis can be used.

5.7 The Training Population

In some embodiments, the reference or training population includes between 10 and 30 subjects. In another embodiment the training population contains between 30-50 subjects. In still other embodiments, the reference population includes two or more

populations each containing between 50 and 100, 100 and 500, between 500 and 1000, or more than 1000 subjects.

In some embodiments, members of each phenotypic subgroup (i.e. bladder cancer and non bladder cancer members) of the training population are preferably selected such that each phenotypic subgroup of the training population has a similar distribution with respect to at least one, two, three, four, five, six, one or more, two or more, three or more, four or more, five or more, six or more, between one and 1000 other phenotypes. For example, age, sex, genetic variation information (*e.g.*, gene SNP mutations, restriction fragment length polymorphisms, microsatellite markers, restriction fragment length polymorphisms, and presence, absence or characterization of short tandem repeats.), treatment regimens; co-morbidities; concentrations of metabolites, blood chemistry levels, and/or other indicators of health and/or wellness.

5.8 Mathematical Models

5.8.1 Regression Models

In some embodiments the expression data for each combination of biomarkers to be tested are used in within a regression model, preferably a logistic regression model. Such a regression model will determine an equation for each possible combination of biomarkers tested, each equation providing a coefficient to be multiplied by the data reflective of the expression level of each individual biomarker represented by the model.

In general, the multiple regression equation of interest can be written

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon$$

where Y, the dependent variable, is presence (when Y is positive) or absence (when Y is negative) of the first phenotypic trait of (*e.g.*, having mild osteoarthritis). This model says that the dependent variable Y depends on k explanatory variables (the measured values representative of the level of the gene product in the tissue of interest for the k select genes

from subjects in the first and second phenotypic subgroups in the training population), plus an error term that encompasses various unspecified omitted factors. In the above-identified model, the parameter β_1 gauges the effect of the first explanatory variable X_1 on the dependent variable Y, holding the other explanatory variables constant. Similarly, β_2 gives the effect of the explanatory variable X_2 on Y, holding the remaining explanatory variables constant.

The logistic regression model is a non-linear transformation of the linear regression. The logistic regression model is termed the "logit" model and can be expressed as

$$\ln[p/(1-p)] = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon \quad \text{or}$$

$$[p/(1-p)] = \exp^{\alpha} \exp^{\beta_1 X_1} \exp^{\beta_2 X_2} \times \cdots \times \exp^{\beta_k X_k} \exp^{\varepsilon}$$

where,

ln is the natural logarithm, $\log^{exp}$, where exp=2.71828...,

p is the probability that the event Y occurs, p(Y=1),

p/(1-p) is the "odds ratio",

ln[p/(1-p)] is the log odds ratio, or "logit", and

all other components of the model are the same as the general regression equation described above. It will be appreciated by those of skill in the art that the term for α and ε can be folded into the same constant. Indeed, in preferred embodiments, a single term is used to represent α and ε. The "logistic" distribution is an S-shaped distribution function. The logit distribution constrains the estimated probabilities (p) to lie between 0 and 1.

In some embodiments of the present invention, the logistic regression model is fit by maximum likelihood estimation (MLE). In other words, the coefficients (*e.g.*, α, β_1, β_2, ...) are determined by maximum likelihood. A likelihood is a conditional probability (e.g., P(Y|X), the probability of Y given X). The likelihood function (L) measures the probability

of observing the particular set of dependent variable values (Y_1, Y_2, ..., Y_n) that occur in the sample data set. It is written as the probability of the product of the dependent variables:

$$L = \text{Prob}\ (Y_1 * Y_2 *** Y_n)$$

The higher the likelihood function, the higher the probability of observing the Ys in the sample. MLE involves finding the coefficients ($\alpha, \beta_1, \beta_2, \ldots$) that makes the log of the likelihood function ($LL < 0$) as large as possible or -2 times the log of the likelihood function (-2LL) as small as possible. In MLE, some initial estimates of the parameters $\alpha, \beta_1, \beta_2, \ldots$ are made. Then the likelihood of the data given these parameter estimates is computed. The parameter estimates are improved the likelihood of the data is recalculated. This process is repeated until the parameter estimates do not change much (for example, a change of less than .01 or .001 in the probability). Examples of logistic regression and fitting logistic logistic regression models are found in Hastie, *The Elements of Statistical Learning*, Springer, New York, 2001, pp. 95-100 which is incorporated herein in its entirety.

5.8.2 Neural Networks

In another embodiment, the expression measured for each of the biomarkers of the present invention can be used to train a neural network. A neural network is a two-stage regression or classification model. A neural network has a layered structure that includes a layer of input units (and the bias) connected by a layer of weights to a layer of output units. For regression, the layer of output units typically includes just one output unit. However, neural networks can handle multiple quantitative responses in a seamless fashion. As such a neural network can be applied to allow identification of biomarkers which differentiate as between more than two populations. In one specific example, a neural network can be trained using expression data from the products of the biomarkers in Table 1 including those noted in Table 3 to identify those combinations of biomarkers which are specific for bladder cancer as compared with not having bladder cancer. As a result, the trained neural network can be used to directly identify combination of biomarkers useful as stage specific biomarkers. In some embodiments, the back-propagation neural network (see, for example

Abdi, 1994, "A neural network primer", J. Biol. System. 2, 247-283) containing a single hidden layer of ten neurons (ten hidden units) found in EasyNN-Plus version 4.0g software package (Neural Planner Software Inc.) is used.

Neural networks are described in Duda *et al.*, 2001, *Pattern Classification*, Second Edition, John Wiley & Sons, Inc., New York; and Hastie *et al.*, 2001, *The Elements of Statistical Learning*, Springer-Verlag, New York which is incorporated herein in its entirety.

5.8.3 Other Mathematical Models

The pattern classification and statistical techniques described above are merely examples of the types of models that can be used to construct a model for bladder cancer classification, for example clustering as described on pages 211-256 of Duda and Hart, *Pattern Classification and Scene Analysis*, 1973, John Wiley & Sons, Inc., New York, incorporated herein by reference in its entirety; Principal component analysis, (see for Jolliffe, 1986, *Principal Component* Analysis, Springer, New York, incorporated herein by reference); nearest neighbour classifier analysis, (see for example Duda, *Pattern Classification*, Second Edition, 2001, John Wiley & Sons, Inc; and Hastie, 2001, *The Elements of Statistical Learning*, Springer, New York); linear discriminant analysis, (see for example Duda, *Pattern Classification*, Second Edition, 2001, John Wiley & Sons, Inc; and Hastie, 2001, *The Elements of Statistical Learning*, Springer, New York; Venables & Ripley, 1997, *Modern Applied Statistics with s-plus*, Springer, New York); Support Vector Machines (see, for example, Cristianini and Shawe-Taylor, 2000, *An Introduction to Support Vector Machines*, Cambridge University Press, Cambridge, Boser *et al.*, 1992, "A training algorithm for optimal margin classifiers, in *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory,* ACM Press, Pittsburgh, PA, pp. 142-152; Vapnik, 1998, *Statistical Learning Theory*, Wiley, New York, all of which are incorporated herein by reference.)

5.9 Choice of Biomarkers Combinations to Select Classifiers

It is preferred to select all of the biomarkers for use in subsequent sections. Thus all possible combinations of biomarkers are evaluated using data representative of the

level of the products of each of the biomarkers. In some embodiments, it is not possible to choose all possible combinations, so subsets of biomarkers are first chosen and then combinations of the subset of biomarkers are used. As would be understood, a selection criteria based on the desired number of selected biomarkers will depend upon the resources available for obtaining the data representing the level of the product of the biomarker and/or the computer resources available for calculating and evaluating classifiers of all or a portion of possible combinations of the selected biomarkers. In some embodiments, the desired number of selected biomarkers which can be evaluated is 4,000; 3,000; 2,000; 1,000; 900; 800; 700; 600; 500; 400; 300; 200; 190; 180; 170; 160; 150; 140; 130; 120; 110; 100; 90; 80; 70; 60; 50; 40; 30; 20; 10.

In one embodiment of the invention, each possible combination of the biomarkers in Table 1 are evaluated. In another embodiment of the invention, each possible combination of 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, etc biomarkers of Table 1 are evaluated. In another embodiment of the invention, each possible combination of 2 and/or 3 and/or 4 and/or 5 and/or 6 and/or 7 and/or 8 and/or 9 and/or 10 of all of the biomarkers of Table 1 are evaluated. In another embodiment of the invention, each possible combination of 2 and/or 3 and/or 4 and/or 5 and/or 6 and/or 7 and/or 8 and/or 9 and/or 10 of a portion of the biomarkers of Table 1 are evaluated. In another embodiment of the invention, the portion of the biomarkers of Table 1 which are evaluated are 20, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 225, or 250 of the biomarkers in Table 1. In yet another embodiment of the invention, the portion of the biomarkers of Table 1 which are selected are ranked on the basis of individual p value wherein the p value is indicative of each biomarkers individual ability to differentiate between members having bladder cancer and members not having bladder cancer, and then the top 60, 50, 40, 30, 20, or 10 individually ranked biomarkers are selected and each possible combination of 2 and/or 3 and/or 4 and/or 5 and/or 6 and/or 7 and/or 8 and/or 9 and/or 10 of all of the selected biomarkers are utilized to create classifiers, which classifiers are evaluated for the ability to differentiate as between members having bladder cancer and members not having bladder cancer. In yet another embodiment, subsets of biomarkers of Table 1 are chosen whose p value is less than 0.5; less than 0.1, less than 0.05, less than 0.01, less than 0.005, less than 0.001, less than 0.0005, less than 0.0001, less than 0.00005, less than 0.00001, less than 0.000005, less than 0.000001 etc. In some

embodiments, subsets of biomarkers of Table 1 are chosen on the basis of the level of differential expression displayed by the biomarker products as between the two or more trait subgroups. Note that in measuring differential fold change in blood, the fold change differences can be quite small, thus in some embodiments, selection of biomarkers is based on a differential fold change where the fold change is greater than 1.2, greater than 1.3, greater than 1.4, greater than 1.5, greater than 1.6, greater than 1.7, greater than 1.8, greater than 1.9, greater than 2.0, greater than 2.1, greater than 2.2, greater than 2.3, greater than 2.4, greater than 2.5, greater than 2.6, greater than 2.7, greater than 2.8, greater than 2.9, greater than 3.0, greater than 3.1, greater than 3.2. greater than 3.3, greater than 3.4 greater than 3.5, greater than 4.0 and the like. In some embodiments, it is helpful to select biomarkers for forming combinations on the basis of both p value and fold change as would be understood by a person skilled in the art. Thus in some embodiments, biomarkers are first selected as outlined above on the basis of the p value resulting from the biomarker data and then a subselection of said biomarkers is chosen on the basis of the differential fold change determined from the biomarker data. In other embodiments, biomarkers are first selected on the basis of differential fold change, and then subselection is made on the basis of p value. In some embodiments, the use of one or more of the selection criteria and subsequent ranking permits the selection of the top 2.5%, 5%, 7.5%, 10%, 12.5%, 15%, 17.5%, 20%, 30%, 40%, 50% or more of the ranked biomarkers for use in forming classifiers. .In another embodiment of the invention, each possible combination of the biomarkers in Table 1 and Table 2 combined are evaluated with the proviso that at least one biomarker from Table 1 is included in the combinations evaluated. In another embodiment of the invention, each possible combination of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, etc biomarkers of Table 1 in combination with 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, etc biomarkers of Table 2 are evaluated. In another embodiment of the invention, an exhaustive search for combinations of 2 and/or 3 and/or 4 and/or 5 and/or 6 and/or 7 and/or 8 and/or 9 and/or 10 of all of the biomarkers of Table 1 and Table 2 combined are evaluated. In another embodiment of the invention, each possible combination of 2 and/or 3 and/or 4 and/or 5 and/or 6 and/or 7 and/or 8 and/or 9 and/or 10 of a portion of the biomarkers of Table 1 and Table 2 are evaluated. In another embodiment of the invention, the portion of the biomarkers of Table 1 and Table 2 which are evaluated are 20, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 225, or 250 of

the biomarkers in Table 1 combined with 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 75, 100, 150, 200, 300, etc of the biomarkers in Table 2 are utilized. In yet another embodiment of the invention, the portion of the biomarkers of Table 1 and Table 2 which are selected are ranked on the basis of individual p value wherein the p value is indicative of each biomarkers individual ability to differentiate between members having bladder cancer and members not having bladder cancer, In yet another embodiment, subsets of biomarkers of Table 1 and Table 2 are chosen whose p value is less than 0.5; less than 0.1, less than 0.05, less than 0.01, less than 0.005, less than 0.001, less than 0.0005, less than 0.0001, less than 0.00005, less than 0.00001, less than 0.000005, less than 0.000001 etc. In some embodiments, subsets of biomarkers of Table 1 and Table 2 are chosen on the basis of the level of differential expression displayed by the biomarker products. Note that in measuring differential fold change in blood, the fold change differences can be quite small, thus in some embodiments, selection of biomarkers is based on a differential fold change where the fold change is greater than 1.2, greater than 1.3, greater than 1.4, greater than 1.5, greater than 1.6, greater than 1.7, greater than 1.8, greater than 1.9, greater than 2.0, greater than 2.1, greater than 2.2, greater than 2.3, greater than 2.4, greater than 2.5, greater than 2.6, greater than 2.7, greater than 2.8, greater than 2.9, greater than 3.0, greater than 3.1, greater than 3.2. greater than 3.3, greater than 3.4 greater than 3.5, greater than 4.0 and the like. In some embodiments, it is helpful to select biomarkers for forming combinations on the basis of both p value and fold change as would be understood by a person skilled in the art. Thus in some embodiments, biomarkers are first selected as outlined above on the basis of the p value resulting from the biomarker data and then a subselection of said biomarkers is chosen on the basis of the differential fold change determined from the biomarker data. In other embodiments, biomarkers are first selected on the basis of differential fold change, and then subselection is made on the basis of p value. In some embodiments, the use of one or more of the selection criteria and subsequent ranking permits the selection of the top 2.5%, 5%, 7.5%, 10%, 12.5%, 15%, 17.5%, 20%, 30%, 40%, 50% or more of the ranked biomarkers are used such that each possible combination of 2 and/or 3 and/or 4 and/or 5 and/or 6 and/or 7 and/or 8 and/or 9 and/or 10 of all of the selected biomarkers are utilized to create classifiers, which classifiers are evaluated for the ability to differentiate as between members having bladder cancer and members not having bladder cancer.

In another embodiment of the invention, each possible combination of the biomarkers in Table 4 are evaluated. In another embodiment of the invention, each possible combination of 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500 etc biomarkers of Table 4 are evaluated. In another embodiment of the invention, each possible combination of 2 and/or 3 and/or 4 and/or 5 and/or 6 and/or 7 and/or 8 and/or 9 and/or 10 of all of the biomarkers of Table 4 are evaluated. In another embodiment of the invention, each possible combination of 2 and/or 3 and/or 4 and/or 5 and/or 6 and/or 7 and/or 8 and/or 9 and/or 10 of a portion of the biomarkers of Table 4 are evaluated. In another embodiment of the invention, the portion of the biomarkers of Table 4 which are evaluated are 20, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 225, 250, 300, 350, 400, 450 or 500 of the biomarkers in Table 4. In yet another embodiment of the invention, the portion of the biomarkers of Table 4 which are selected are ranked on the basis of individual p value wherein the p value is indicative of each biomarkers individual ability to differentiate between members having superficial bladder cancer and members not having bladder cancer, and then the top 60, 50, 40, 30, 20, or 10 individually ranked biomarkers are selected and each possible combination of 2 and/or 3 and/or 4 and/or 5 and/or 6 and/or 7 and/or 8 and/or 9 and/or 10 of all of the selected biomarkers are utilized to create classifiers, which classifiers are evaluated for the ability to differentiate as between members having superficial bladder cancer and members not having bladder cancer. In yet another embodiment of the invention, the portion of the biomarkers of Table 4 which are selected are those whose p value is less than 0.5; less than 0.1, less than 0.05, less than 0.01, less than 0.005, less than 0.001, less than 0.0005, less than 0.0001, less than 0.00005, less than 0.00001, less than 0.000005, less than 0.000001 etc. In some embodiments, subsets of biomarkers of Table 4 are chosen to form combinations for input into classifiers on the basis of the level of differential expression displayed by the biomarker products. Note that in measuring differential fold change in blood, the fold change differences can be quite small, thus in some embodiments, selection of biomarkers is based on a differential fold change where the fold change is greater than 1.2, greater than 1.3, greater than 1.4, greater than 1.5, greater than 1.6, greater than 1.7, greater than 1.8, greater than 1.9, greater than 2.0, greater than 2.1, greater than 2.2, greater than 2.3, greater than 2.4, greater than 2.5, greater than 2.6, greater than 2.7, greater than 2.8, greater than 2.9, greater than 3.0, greater than 3.1, greater

than 3.2. greater than 3.3, greater than 3.4 greater than 3.5, greater than 4.0 and the like. In some embodiments, it is helpful to select biomarkers for forming combinations on the basis of both p value and fold change as would be understood by a person skilled in the art. Thus in some embodiments, biomarkers are first selected as outlined above on the basis of the p value resulting from the biomarker data and then a subselection of said biomarkers is chosen on the basis of the differential fold change determined from the biomarker data. In other embodiments, biomarkers are first selected on the basis of differential fold change, and then subselection is made on the basis of p value. In some embodiments, the use of one or more of the selection criteria and subsequent ranking permits the selection of the top 2.5%, 5%, 7.5%, 10%, 12.5%, 15%, 17.5%, 20%, 30%, 40%, 50% or more of the ranked biomarkers are used such that each possible combination of 2 and/or 3 and/or 4 and/or 5 and/or 6 and/or 7 and/or 8 and/or 9 and/or 10 of all of the selected biomarkers are utilized to create classifiers, which classifiers are evaluated for the ability to differentiate as between members having bladder cancer and members not having bladder cancer. In another embodiment of the invention, each possible combination of the biomarkers in Table 4 and Table 5 combined are evaluated with the proviso that at least one biomarker from Table 4 is included in the combinations evaluated. In another embodiment of the invention, each possible combination of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, etc biomarkers of Table 4 in combination with 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, etc biomarkers of Table 5 are evaluated. In another embodiment of the invention, an exhaustive search for combinations of 2 and/or 3 and/or 4 and/or 5 and/or 6 and/or 7 and/or 8 and/or 9 and/or 10 of all of the biomarkers of Table 4 and Table 5 combined are evaluated. In another embodiment of the invention, each possible combination of 2 and/or 3 and/or 4 and/or 5 and/or 6 and/or 7 and/or 8 and/or 9 and/or 10 of a portion of the biomarkers of Table 4 and Table 5 are evaluated. In another embodiment of the invention, the portion of the biomarkers of Table 4 and Table 5 which are evaluated are 20, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 225, or 250, 300, 350, 400, 450, 500 of the biomarkers in Table 4 combined with 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 75, 100, 150, 200, 300, 500, 1000, 1250, 1500, 1750, 2000, 2250 etc of the biomarkers in Table 5 are utilized. In yet another embodiment of the invention, the portion of the biomarkers of Table 4 which are selected are ranked on the basis of individual p value wherein the p value is indicative of each biomarkers individual ability to differentiate

between members having superficial bladder cancer and members not having bladder cancer, and then the top 60, 50, 40, 30, 20, or 10 individually ranked biomarkers are selected and each possible combination of 2 and/or 3 and/or 4 and/or 5 and/or 6 and/or 7 and/or 8 and/or 9 and/or 10 of all of the selected biomarkers are utilized to create classifiers, which classifiers are evaluated for the ability to differentiate as between members having superficial bladder cancer and members not having bladder cancer. . In yet another embodiment of the invention, the portion of the biomarkers of Table 4 and Table 5 which are selected are those whose p value is less than 0.5; less than 0.1, less than 0.05, less than 0.01, less than 0.005, less than 0.001, less than 0.0005, less than 0.0001, less than 0.00005, less than 0.00001, less than 0.000005, less than 0.000001 etc. In some embodiments, subsets of biomarkers of Table 4 and/or Table 5 are chosen to form combinations for input into classifiers on the basis of the level of differential expression displayed by the biomarker products. Note that in measuring differential fold change in blood, the fold change differences can be quite small, thus in some embodiments, selection of biomarkers is based on a differential fold change where the fold change is greater than 1.2, greater than 1.3, greater than 1.4, greater than 1.5, greater than 1.6, greater than 1.7, greater than 1.8, greater than 1.9, greater than 2.0, greater than 2.1, greater than 2.2, greater than 2.3, greater than 2.4, greater than 2.5, greater than 2.6, greater than 2.7, greater than 2.8, greater than 2.9, greater than 3.0, greater than 3.1, greater than 3.2. greater than 3.3, greater than 3.4 greater than 3.5, greater than 4.0 and the like. In some embodiments, it is helpful to select biomarkers for forming combinations on the basis of both p value and fold change as would be understood by a person skilled in the art. Thus in some embodiments, biomarkers are first selected as outlined above on the basis of the p value resulting from the biomarker data and then a subselection of said biomarkers is chosen on the basis of the differential fold change determined from the biomarker data. In other embodiments, biomarkers are first selected on the basis of differential fold change, and then subselection is made on the basis of p value. In some embodiments, the use of one or more of the selection criteria and subsequent ranking permits the selection of the top 2.5%, 5%, 7.5%, 10%, 12.5%, 15%, 17.5%, 20%, 30%, 40%, 50% or more of the ranked biomarkers are used such that each possible combination of 2 and/or 3 and/or 4 and/or 5 and/or 6 and/or 7 and/or 8 and/or 9 and/or 10 of all of the selected biomarkers are utilized to create classifiers,

which classifiers are evaluated for the ability to differentiate as between members having bladder cancer and members not having bladder cancer.

In another embodiment of the invention, each possible combination of the biomarkers in Table 7 is evaluated. In another embodiment of the invention, each possible combination of any of up to 2, 3, 4, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200 etc of Table 7 are evaluated. In another embodiment of the invention, biomarkers in Table 7 are ranked on the basis of individual p value (as noted in Table 1) wherein the p value is indicative of each biomarkers individual ability to differentiate between members having bladder cancer and members not having bladder cancer, and then the top 60, 50, 40, 30, 20, or 10 individually ranked biomarkers are evaluated in all possible combinations so as to select those combinations which are most able to differentiate as between members having bladder cancer and members not having bladder cancer. In yet another embodiment, combinations of 2 and/or 3 and/or 4 and/or 5 and/or 6 and/or 7 and/or 8 and/or 9 and/or 10 of all of the biomarkers of Table 7 are utilized to create classifiers which can differentiate as between bladder cancer and non bladder cancer. In yet another embodiment combinations of 2 and/or 3 and/or 4 and/or 5 and/or 6 and/or 7 and/or 8 and/or 9 and/or 10 of up to 2, 3, 4, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, or 200 of Table 7 are used to create classifiers which are then evaluated for their ability to differentiate as between bladder cancer and non bladder cancer. In yet another embodiment combinations of 2 and/or 3 and/or 4 and/or 5 and/or 6 and/or 7 and/or 8 and/or 9 and/or 10 of the top 200, 175, 150, 125, 100, 90, 80, 70, 60, 50, 40, 30, 20 or 10 of Table 7 (on the basis of p value) are used to create classifiers which are then evaluated for their ability to differentiate as between bladder cancer and non bladder cancer. In yet another embodiment of the invention, the portion of the biomarkers of Table 7 which are selected are those whose p value is less than 0.5; less than 0.1, less than 0.05, less than 0.01, less than 0.005, less than 0.001, less than 0.0005, less than 0.0001, less than 0.00005, less than 0.00001, less than 0.000005, less than 0.000001 etc. In some embodiments, subsets of biomarkers of Table 7 are chosen to form combinations for input into classifiers on the basis of the level of differential expression displayed by the biomarker products. Note that in measuring differential fold change in blood, the fold change differences can be quite small, thus in some embodiments, selection of biomarkers is based on a differential fold change where the fold change is greater than 1.2, greater than 1.3,

greater than 1.4, greater than 1.5, greater than 1.6, greater than 1.7, greater than 1.8, greater than 1.9, greater than 2.0, greater than 2.1, greater than 2.2, greater than 2.3, greater than 2.4, greater than 2.5, greater than 2.6, greater than 2.7, greater than 2.8, greater than 2.9, greater than 3.0, greater than 3.1, greater than 3.2. greater than 3.3, greater than 3.4 greater than 3.5, greater than 4.0 and the like. In some embodiments, it is helpful to select biomarkers for forming combinations on the basis of both p value and fold change as would be understood by a person skilled in the art. Thus in some embodiments, biomarkers are first selected as outlined above on the basis of the p value resulting from the biomarker data and then a subselection of said biomarkers is chosen on the basis of the differential fold change determined from the biomarker data. In other embodiments, biomarkers are first selected on the basis of differential fold change, and then subselection is made on the basis of p value. In some embodiments, the use of one or more of the selection criteria and subsequent ranking permits the selection of the top 2.5%, 5%, 7.5%, 10%, 12.5%, 15%, 17.5%, 20%, 30%, 40%, 50% or more of the ranked biomarkers are used such that each possible combination of 2 and/or 3 and/or 4 and/or 5 and/or 6 and/or 7 and/or 8 and/or 9 and/or 10 of all of the selected biomarkers are utilized to create classifiers, which classifiers are evaluated for the ability to differentiate as between members having bladder cancer and members not having bladder cancer.

In yet another embodiment, all combinations of 2 and/or 3 and/or 4 and/or 5 and/or 6 and/or 7 and/or 8 and/or 9 and/or 10 of all of the biomarkers of Table 10 are utilized to create classifiers which can differentiate as between bladder cancer and non bladder cancer.

In another embodiment of the invention, each possible combination of 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, etc biomarkers of Table 11 are evaluated. In another embodiment of the invention, each possible combination of 2 and/or 3 and/or 4 and/or 5 and/or 6 and/or 7 and/or 8 and/or 9 and/or 10 of all of the biomarkers of Table 11 are evaluated. In another embodiment of the invention, each possible combination of 2 and/or 3 and/or 4 and/or 5 and/or 6 and/or 7 and/or 8 and/or 9 and/or 10 of a portion of the biomarkers of Table 11 are evaluated. In another embodiment of the invention, the portion of the biomarkers of Table 11 which are evaluated are 20, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150,

175, 200, 225, or 250 of the biomarkers in Table 11. In yet another embodiment of the invention, the portion of the biomarkers of Table 11 which are selected are ranked on the basis of individual p value wherein the p value is indicative of each biomarkers individual ability to differentiate between members having bladder cancer and members not having bladder cancer, and then the top 60, 50, 40, 30, 20, or 10 individually ranked biomarkers are selected and each possible combination of 2 and/or 3 and/or 4 and/or 5 and/or 6 and/or 7 and/or 8 and/or 9 and/or 10 of all of the selected biomarkers are utilized to create classifiers, which classifiers are evaluated for the ability to differentiate as between members having bladder cancer and members not having bladder cancer. In yet another embodiment of the invention, the portion of the biomarkers of Table 11 which are selected are those whose p value is less than 0.5; less than 0.1, less than 0.05, less than 0.01, less than 0.005, less than 0.001, less than 0.0005, less than 0.0001, less than 0.00005, less than 0.00001, less than 0.000005, less than 0.000001 etc. In some embodiments, subsets of biomarkers of Table 11 are chosen to form combinations for input into classifiers on the basis of the level of differential expression displayed by the biomarker products. Note that in measuring differential fold change in blood, the fold change differences can be quite small, thus in some embodiments, selection of biomarkers is based on a differential fold change where the fold change is greater than 1.2, greater than 1.3, greater than 1.4, greater than 1.5, greater than 1.6, greater than 1.7, greater than 1.8, greater than 1.9, greater than 2.0, greater than 2.1, greater than 2.2, greater than 2.3, greater than 2.4, greater than 2.5, greater than 2.6, greater than 2.7, greater than 2.8, greater than 2.9, greater than 3.0, greater than 3.1, greater than 3.2. greater than 3.3, greater than 3.4 greater than 3.5, greater than 4.0 and the like. In some embodiments, it is helpful to select biomarkers for forming combinations on the basis of both p value and fold change as would be understood by a person skilled in the art. Thus in some embodiments, biomarkers are first selected as outlined above on the basis of the p value resulting from the biomarker data and then a subselection of said biomarkers is chosen on the basis of the differential fold change determined from the biomarker data. In other embodiments, biomarkers are first selected on the basis of differential fold change, and then subselection is made on the basis of p value. In some embodiments, the use of one or more of the selection criteria and subsequent ranking permits the selection of the top 2.5%, 5%, 7.5%, 10%, 12.5%, 15%, 17.5%, 20%, 30%, 40%, 50% or more of the ranked biomarkers are

used such that each possible combination of 2 and/or 3 and/or 4 and/or 5 and/or 6 and/or 7 and/or 8 and/or 9 and/or 10 of all of the selected biomarkers are utilized to create classifiers, which classifiers are evaluated for the ability to differentiate as between members having bladder cancer and members not having bladder cancer. In another embodiment of the invention, each possible combination of the biomarkers in Table 11 and Table 2 combined are evaluated with the proviso that at least one biomarker from Table 11 is included in the combinations evaluated. In another embodiment of the invention, each possible combination of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, etc biomarkers of Table 11 in combination with 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, etc biomarkers of Table 2 are evaluated. In another embodiment of the invention, an exhaustive search for combinations of 2 and/or 3 and/or 4 and/or 5 and/or 6 and/or 7 and/or 8 and/or 9 and/or 10 of all of the biomarkers of Table 1 and Table 2 combined are evaluated. In another embodiment of the invention, each possible combination of 2 and/or 3 and/or 4 and/or 5 and/or 6 and/or 7 and/or 8 and/or 9 and/or 10 of a portion of the biomarkers of Table 11 and Table 2 are evaluated. In another embodiment of the invention, the portion of the biomarkers of Table 11 and Table 2 which are evaluated are 20, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 225, or 250 of the biomarkers in Table 11 combined with 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 75, 100, 150, 200, 300, etc of the biomarkers in Table 2 are utilized. In yet another embodiment of the invention, the portion of the biomarkers of Table 11 which are selected are ranked on the basis of individual p value wherein the p value is indicative of each biomarkers individual ability to differentiate between members having bladder cancer and members not having bladder cancer, and then the top 60, 50, 40, 30, 20, or 10 individually ranked biomarkers are selected and each possible combination of 2 and/or 3 and/or 4 and/or 5 and/or 6 and/or 7 and/or 8 and/or 9 and/or 10 of all of the selected biomarkers are utilized to create classifiers, which classifiers are evaluated for the ability to differentiate as between members having bladder cancer and members not having bladder cancer. In yet another embodiment of the invention, the portion of the biomarkers of Table 11 and Table 2 which are selected are those whose p value is less than 0.5; less than 0.1, less than 0.05, less than 0.01, less than 0.005, less than 0.001, less than 0.0005, less than 0.0001, less than 0.00005, less than 0.00001, less than 0.000005, less than 0.000001 etc. In some embodiments, subsets of biomarkers of Table 11 and/or Table 2 are chosen to form combinations for input into classifiers on the basis of

the level of differential expression displayed by the biomarker products. Note that in measuring differential fold change in blood, the fold change differences can be quite small, thus in some embodiments, selection of biomarkers is based on a differential fold change where the fold change is greater than 1.2, greater than 1.3, greater than 1.4, greater than 1.5, greater than 1.6, greater than 1.7, greater than 1.8, greater than 1.9, greater than 2.0, greater than 2.1, greater than 2.2, greater than 2.3, greater than 2.4, greater than 2.5, greater than 2.6, greater than 2.7, greater than 2.8, greater than 2.9, greater than 3.0, greater than 3.1, greater than 3.2. greater than 3.3, greater than 3.4 greater than 3.5, greater than 4.0 and the like. In some embodiments, it is helpful to select biomarkers for forming combinations on the basis of both p value and fold change as would be understood by a person skilled in the art. Thus in some embodiments, biomarkers are first selected as outlined above on the basis of the p value resulting from the biomarker data and then a subselection of said biomarkers is chosen on the basis of the differential fold change determined from the biomarker data. In other embodiments, biomarkers are first selected on the basis of differential fold change, and then subselection is made on the basis of p value. In some embodiments, the use of one or more of the selection criteria and subsequent ranking permits the selection of the top 2.5%, 5%, 7.5%, 10%, 12.5%, 15%, 17.5%, 20%, 30%, 40%, 50% or more of the ranked biomarkers are used such that each possible combination of 2 and/or 3 and/or 4 and/or 5 and/or 6 and/or 7 and/or 8 and/or 9 and/or 10 of all of the selected biomarkers are utilized to create classifiers, which classifiers are evaluated for the ability to differentiate as between members having bladder cancer and members not having bladder cancer.

5.10 Evaluation of Classifiers

Once one or more classifiers have been computed using a mathematical model, the classifiers can be evaluated to determine which of the classifiers are most effective for the desired purpose. In one preferred embodiment of the invention, each classifier is evaluated or scored for its ability to properly characterize each individual of the training population as having bladder cancer or not having bladder cancer. For example one can evaluate the classifiers using cross validation Leave One out Cross Validation, n-fold cross validation, jackknife analysis using standard statistical methods and disclosed. As used herein the process of evaluating the classifiers is termed as "scoring. In some embodiments, scoring is done using the training population. In other embodiments, scoring is done using a

"scoring population" as described herein. In yet other embodiments, scoring is done using both the "training population" and the "scoring population" In one embodiment, the scoring population includes members of the training population in addition to one or more members not used in the training population. In some embodiments, five percent or less, ten percent or less, twenty percent or less, thirty percent or less, fifty percent or less, or ninety percent or less of the members of the training population are common to the scoring population.

In some embodiments, the Percent Correct Predictions statistic is used to score each classifier. The "Percent Correct Predictions" statistic assumes that if the estimated p is greater than or equal to 0.5, then the event is expected to occur and to not occur otherwise. By assigning these probabilities zeros and ones, a comparison can be made to the values of the samples in the training population to determine what percentage of the training population was sampled correctly.

In one embodiment, the method used to evaluate the classifier for its ability to properly characterize each individual of the training population is a method which evaluates the classifiers sensitivity (TPF, true positive fraction) and 1-specificity (TNF, true negative fraction). For example, in one embodiment the Receiver Operating Characteristic ("ROC") is utilised. The ROC provides several parameters to evaluate both the sensitivity and specificity of the diagnostic result of the equation generated. For example, in one embodiment the ROC area (area under the curve) is used to evaluate the equations. In a preferred embodiment, an ROC area greater than 0.5, 0.6, 0.7, 0.8, 0.9 is preferred. A perfect ROC area score of 1.0 on the other hand indicates with both 100% sensitivity and 100% specificity.

As would be understood by those of skill in the relevant arts, area under the curve converts the two dimensional information contained in the ROC curve into one dimensional information. In other embodiments, information from the two dimensional aspect of the ROC curve is utilized directly. For example, the ROC curve also provides information with respect to the sensitivity and specificity of the classifier. In some embodiments, classifiers are selected on the basis of either sensitivity or specificity. This can be an important scoring indicator. For example, a diagnostic classifier with high specificity (i.e. smaller number of false negatives) may be important in situations where it is safer to misdiagnosis an individual as having disease rather than misdiagnosing a person as normal.

Therefore in some embodiments, a cutoff can be set for either sensitivity or specificity and the classifier ranked or scored on the basis of the remaining variable. In some embodiments, ROC curves are generated for each classifier using any known method to generate data. In some embodiments data is generated using microarray. In some embodiments data is generated utilizing quantitative RT-PCR.

In some embodiments, a classifier is a weighted logistic regression model characterized by a multicategory logit model. For example, in some embodiments, a classifier discriminates between two different trait groups. In other embodiments, a classifier discriminates between more than two different trait groups. Logit models, including multicategory logit models are described in Agresti, *An Introduction to Categorical Data Analysis*, John Wiley & Sons, Inc., 1996, New York, Chapters 7 and 8, which is hereby incorporated by reference. Table 8 illustrates the data that is used to form an ROC curve based on expression data applied to a mathematical model that uses the logit:

$$\ln[p/(1-p)]=\alpha+\beta_1 X_1+\beta_2 X_2+\varepsilon.$$

Table 8: Values for the logit $\ln[p/(1-p)]=\alpha+\beta_1 X_1+\beta_2 X_2+\varepsilon$

ln[p/(1-p)]	Presence / Absence of a Trait
0.98	Y
0.97	Y
0.95	Y
0.93	Y
0.91	N
0.11	Y

ln[p/(1-p)]	Presence / Absence of a Trait
0.07	N
0.03	N

Each row in Table 8 corresponds to a different specimen in the scoring population. The left column represents the results of the logit for the classifier being sampled. The specimens in Table 8 are ranked by the logit score listed in the left hand column. The right hand column details the presence or absence of the trait that is being considered by the regression equation. Table 8 can be used to compute a ROC curve in which each row in Table 8 is considered a threshold cut-off value in order to compute ROC curve data points. Then, the area under the ROC curve can be computed in order to assess the predictive quality of the classifier.

5.11 Products of the Biomarkers of the Invention

As would be understood by a person skilled in the art, the identification of one or more of a combination of biomarkers which are differentially expressed as between bladder cancer and non bladder cancer allows the diagnosis of bladder cancer for a test individual using data reflective of the expression of the products of the biomarkers (genes) of the combination identified.

The products of each of the biomarkers of the invention includes both RNA and protein. RNA products of the biomarkers of the invention are transcriptional products of the biomarkers of the invention and include populations of hnRNA, mRNA, and one or more spliced variants of mRNA. To practice the invention, measurement of one or more of the populations of the RNA products of the biomarkers of the invention can be used for purposes of diagnosis. More particularly, measurement of those populations of RNA products of the biomarkers which are differentially expressed as between bladder cancer and/or early bladder cancer and non bladder cancer are encompassed herein.

In one embodiment of the invention, the RNA products of the biomarkers of the invention which are measured is the population of RNA products including the hnRNA,

the mRNA, and all of the spliced variants of the mRNA. In another embodiment, the RNA products of the biomarkers of the invention which are measured are the population of mRNA. In another embodiment of the invention the RNA products of the biomarkers of the invention which are measured is the population of mRNA which is expressed in blood. In yet another embodiment of the invention, RNA products of the biomarkers of the invention which are measured are the population of one or more spliced variants of the mRNA. In yet another embodiment of the invention, RNA products of the biomarkers of the invention which are measured is the population of one or more spliced variants of the mRNA which are expressed in blood. In another embodiment the RNA products of the biomarkers of the invention are all mRNA corresponding to the locus link (Gene ID) identified in any one of Tables 1-2, 4-5, 7 and 10. In another embodiment of the invention the RNA products of the biomarkers of the invention which are measured are the RNA products corresponding to the locus link (gene ID) in any one of Tables 3 and 6. In another embodiment of the invention the RNA products of the biomarkers of the invention which are measured are those RNA products corresponding to the locus link (gene ID) which are expressed in blood.

Protein products of the biomarkers of the invention are also included within the scope of the invention and include the entire population of protein products arising from a biomarker of the invention. As would be understood by a person skilled in the art, the entire population of proteins arising from a biomarker of the invention include proteins, protein variants arising from spliced mRNA variants, and post translationally modified proteins. In one embodiment the protein products of the biomarkers of the invention are all proteins corresponding to the locus link (Gene ID) identified in any one of Tables 1-2, 4-5, 7 and 10 - 11. In another embodiment of the invention the protein products of the biomarkers of the invention which are measured are the proteins corresponding to the locus link (gene ID) in any one of Tables 3 and 6. In another embodiment of the invention the protein products of the biomarkers of the invention which are measured are those proteins corresponding to the locus link (gene ID) which are expressed in blood. To practice the invention, measurement of one or more of the populations of the protein products of the biomarkers of the invention can be used for purposes of diagnosis. More particularly, measurement of those populations of protein products of the biomarkers which are differentially expressed as between

individuals with bladder cancer and/or early stage bladder cancer and individuals without bladder cancer are useful for purposes of diagnosis and are encompassed herein.

In one embodiment of the invention the protein products of the biomarkers of the invention which are measured are the entire population of protein products translated from the RNA products of the biomarkers of the invention. In another embodiment, the protein products of the biomarkers of the invention are those protein products which are expressed in blood. In yet another embodiment of the invention, the protein products of any one or more of the biomarkers of the invention are any one or more of the protein products translated from any one or more of the mRNA spliced variants. In yet another embodiment of the invention, the protein products of the biomarkers of the invention are any one or more of the protein products translated from any one or more of the mRNA spliced variants expressed in blood.

5.12 Use of the Combinations Identified to Diagnose Bladder Cancer and/or Early Stage Bladder Cancer

As described herein, the application of a mathematical model (e.g. logistic regression etc.) using the data corresponding to the level of expression of each biomarker of the tested biomarker combination creates a classifier. Classifiers use mathematical functions to convert data representative of the level of expression of each of the biomarkers of the tested biomarker combination into a diagnostic determination as between whether an individual has bladder cancer and/or early stage bladder cancer or does not. Classifiers can be used directly to diagnose an individual as having bladder cancer and/or early bladder cancer or not having bladder cancer, by providing data for a test individual for input into a classifier or each classifier in a classifier group resulting in a diagnostic determination. For example, where the classifier is developed using logistic regression the classifier takes the form as follows:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon$$

where X1, X2, ...Xk of the equation represent the measured values representative of the level of the gene product in the tissue of interest for the *k* select biomarkers of the combination used to generate the classifier. Thus to diagnose a test individual, measurement values for

each biomarker of the equation are input and the value of Y determines the diagnosis of said test individual.

The combinations identified can also be used independently of the classifier. For example, if a classifier is chosen (e.g. has an ROC area under the curve score of 0.9 indicating high sensitivity and high specificity) which uses three biomarkers then one can measure the abundance of the products for each of the three biomarkers in a test individual and compare the measurement of the abundance of each of the three biomarkers with one or more individuals from a control population of individuals having bladder cancer and/or early bladder cancer from a control population of individuals not having bladder cancer so as to determine whether the pattern of expression of the test individual is more similar to the controls having bladder cancer and/or early bladder cancer and not having bladder cancer. In a preferred embodiment, one would use the classifier generated so as to diagnose an individual, e.g., by the measure of the level of expression of the RNA and/or protein products of the biomarkers of the combination identified in a test individual for input into the classifier. In one embodiment, the same method is used to generate the expression data used to generate the mathematical model as is used to diagnose the test individual.

5.13 Use of the Combinations Identified to Monitor Regression of Bladder Cancer

The invention teaches the ability to identify useful combinations of biomarkers and classifiers for those combinations for the purposes of diagnosing an individual as having early stage bladder cancer or not having bladder cancer. It would be understood by a person skilled in the art that combinations and classifiers which are diagnostic for early stage bladder cancer as compared with not having bladder cancer are also useful in determining whether an individual has regressed with regards to the severity of their bladder cancer, for example, in response to treatment. For example, an individual can be diagnosed as having early stage bladder cancer prior to treatment using one or more of the combinations identified. Subsequent to treatment the individual could again be tested to determine whether said individual still has early stage bladder cancer. In the event that the individual can no longer be identified the stage prior to treatment, this may in itself suggest treatment is effective. In addition, the treatment may lead to regression of the stage of bladder cancer such that the individual now is diagnosed as not having bladder cancer.

5.14 Polynucleotides Used to Measure the Products of the Biomarkers of the Invention

As a means of measuring the expression of the RNA products of the biomarkers of the invention, one can use one or more of the following as would be understood by a person skilled in the art in combination with one or more methods to measure RNA expression in a sample of the invention: oligonucleotides, cDNA, DNA, RNA, PCR products, synthetic DNA, synthetic RNA, or other combinations of naturally occurring of modified nucleotides which specifically hybridize to one or more of the RNA products of the biomarkers of the invention. In another specific embodiment, the oligonucleotides, cDNA, DNA, RNA, PCR products, synthetic DNA, synthetic RNA, or other combinations of naturally occurring of modified nucleotides oligonucleotides which selectively hybridize to one or more of the RNA products of the biomarker of the invention are used. In a preferred embodiment, the oligonucleotides, cDNA, DNA, RNA, PCR products, synthetic DNA, synthetic RNA, or other combinations of naturally occurring of modified nucleotides oligonucleotides which both specifically and selectively hybridize to one or more of the RNA products of the biomarker of the invention are used.

5.15 Techniques to Measure the RNA Products of the Biomarkers of the Invention

Array Hybridization

In one embodiment of the invention, the polynucleotide used to measure the RNA products of the invention can be used as nucleic acid members stably associated with a support to comprise an array according to one aspect of the invention. The length of a nucleic acid member can range from 8 to 1000 nucleotides in length and are chosen so as to be specific for the RNA products of the biomarkers of the invention. In one embodiment, these members are specific and/or selective for RNA products of the biomarkers of the invention. In yet another embodiment these members are specific and/or selective for the mRNA products of the biomarkers of the invention. In a preferred embodiment, these members are specific and/or selective for all of the variants of the mRNA products of the biomarkers of the invention. In yet another preferred embodiment, these members are specific and/or selective for one or more variants of the mRNA products of the biomarkers of the invention. The nucleic acid members may be single or double stranded, and/or may be oligonucleotides or PCR fragments amplified from cDNA. Preferably oligonucleotides are approximately 20-30 nucleotides in length. ESTs are preferably 100 to 600 nucleotides in

length. It will be understood to a person skilled in the art that one can utilize portions of the expressed regions of the biomarkers of the invention as a probe on the array. More particularly oligonucleotides complementary to the genes of the invention and or cDNA or ESTs derived from the genes of the invention are useful. In some embodiments of the invention the polynucleotides capable of specifically and/or selectively hybridizing to RNA products of the biomarkers of the invention can be spotted onto an array for use in the invention. For oligonucleotide based arrays, the selection of oligonucleotides corresponding to the gene of interest which are useful as probes is well understood in the art. More particularly it is important to choose regions which will permit hybridization to the target nucleic acids. Factors such as the Tm of the oligonucleotide, the percent GC content, the degree of secondary structure and the length of nucleic acid are important factors. See for example US Patent No. 6,551,784. In one embodiment, the array consists of sequences of between 10-1000 nucleotides in length capable of hybridizing to one or more of the products of each of the biomarkers of the invention as disclosed in Table 1, Table 2, Table 4, Table 5 and Table 7 and/or Table 10 and/or Table 11 including those specific RNA and/or protein products noted in Table 3 and Table 6 and/or Table 12.

The target nucleic acid samples that are hybridized to and analyzed with an array of the invention are preferably from human cartilage, blood or synovial fluid. A limitation for this procedure lies in the amount of RNA available for use as a target nucleic acid sample. Preferably, at least 1 microgram of total RNA is obtained for use according to this invention. Lesser quantities of RNA can be used in combination with PCR and primers directed to the mRNA subspecies (e.g. poly T oligonucleotides).

Target Preparation

The targets for the arrays according to the invention are preferably derived from human blood and/or human bladder tissue.

A target nucleic acid is capable of binding to a nucleic acid probe or nucleic acid member of complementary sequence through one or more types of chemical bonds, usually through complementary base pairing, usually through hydrogen bond formation.

As used herein, a "nucleic acid derived from an mRNA transcript: or a "nucleic acid corresponding to an mRNA" refers to a nucleic acid for which synthesis of the mRNA transcript or a sub-sequence thereof has ultimately served as a template. Thus, a cDNA reverse transcribed from an mRNA, an RNA transcribed from that cDNA, a DNA amplified from the cDNA, an RNA transcribed from the amplified DNA, etc., are all derived from or correspond to the mRNA transcript and detection of such derived or corresponding products is indicative of or proportional to the presence and/or abundance of the original transcript in a sample. Thus, suitable target nucleic acid samples include, but are not limited to, mRNA transcripts of a gene or genes, cDNA reverse transcribed from the mRNA, cRNA transcribed from the cDNA, DNA amplified from a gene or genes, RNA transcribed from amplified DNA, and the like. The nucleic acid targets used herein are preferably derived from human cartilage, blood or synovial fluid. Preferably, the targets are nucleic acids derived from human cartilage, blood or synovial fluid extracts. Nucleic acids can be single- or double-stranded DNA, RNA, or DNA-RNA hybrids synthesised from human cartilage, blood or synovial fluid mRNA extracts using methods known in the art, for example, reverse transcription or PCR.

In the simplest embodiment, such a nucleic acid target comprises total mRNA or a nucleic acid sample corresponding to mRNA (e.g., cDNA) isolated from tissue or blood samples. In another embodiment, total mRNA is isolated from a given sample using, for example, an acid guanidinium-phenol-chloroform extraction method and polyA+ mRNA is isolated by oligo dT column chromatography or by using (dT)n magnetic beads (see, e.g., Sambrook et al., *Molecular Cloning: A Laboratory Manual* (2nd ed.), Vols. 1-3, Cold Spring Harbor Laboratory, (1989), or Current Protocols in Molecular Biology, F. Ausubel et al., ed. Greene Publishing and Wiley-Interscience, New York (1987). In a preferred embodiment, total RNA is extracted using TRIzol® reagent (GIBCO/BRL, Invitrogen Life Technologies, Cat. No. 15596). Purity and integrity of RNA is assessed by absorbance at 260/280nm and agarose gel electrophoresis followed by inspection under ultraviolet light.

In some embodiments, it is desirable to amplify the target nucleic acid sample prior to hybridization, for example, when limited quantities of tissue are used. One of skill in the art will appreciate that whatever amplification method is used, if a quantitative result is

desired, care must be taken to use a method that maintains or controls for the relative frequencies of the amplified nucleic acids. Methods of "quantitative" amplification are well known to those of skill in the art. For example, quantitative PCR involves simultaneously co-amplifying a known quantity of a control sequence using the same primers. This provides an internal standard that may be used to calibrate the PCR reaction. The high density array may then include probes specific to the internal standard for quantification of the amplified nucleic acid. Detailed protocols for quantitative PCR are provided in *PCR Protocols, A Guide to Methods and Applications*, Innis et al., Academic Press, Inc. N.Y., (1990).

Other suitable amplification methods include, but are not limited to polymerase chain reaction (PCR) (Innis, et al., *PCR Protocols. A Guide to Methods and Application*. Academic Press, Inc. San Diego, (1990)), ligase chain reaction (LCR) (see Wu and Wallace, 1989, *Genomics*, 4:560; Landegren, et al., 1988, *Science*, 241:1077 and Barringer, et al., 1990, *Gene*, 89:117, transcription amplification (Kwoh, et al., 1989, *Proc. Natl. Acad. Sci. USA*, 86: 1173), and self-sustained sequence replication (Guatelli, et al., 1990, *Proc. Nat. Acad. Sci. USA,* 87: 1874).

In a particularly preferred embodiment, the target nucleic acid sample mRNA is reverse transcribed with a reverse transcriptase and a primer consisting of oligo dT and a sequence encoding the phage T7 promoter to provide single-stranded DNA template. The second DNA strand is polymerized using a DNA polymerase. After synthesis of double-stranded cDNA, T7 RNA polymerase is added and RNA is transcribed from the cDNA template. Successive rounds of transcription from each single cDNA template results in amplified RNA. Methods of *in vitro* transcription are well known to those of skill in the art (see, e.g., Sambrook, *supra.*) and this particular method is described in detail by Van Gelder, et al., 1990, *Proc. Natl. Acad. Sci. USA*, 87: 1663-1667 who demonstrate that *in vitro* amplification according to this method preserves the relative frequencies of the various RNA transcripts. Moreover, Eberwine et al. *Proc. Natl. Acad. Sci. USA*, 89: 3010-3014 provide a protocol that uses two rounds of amplification via *in vitro* transcription to achieve greater than 10^6 fold amplification of the original starting material thereby permitting expression monitoring even where biological samples are limited.

Labelling of Target or Nucleic Acid Probe

Either the target or the probe can be labelled.

Any analytically detectable marker that is attached to or incorporated into a molecule may be used in the invention. An analytically detectable marker refers to any molecule, moiety or atom which is analytically detected and quantified.

Detectable labels suitable for use in the present invention include any composition detectable by spectroscopic, photochemical, biochemical, immunochemical, electrical, optical or chemical means. Useful labels in the present invention include biotin for staining with labelled streptavidin conjugate, magnetic beads (e.g., Dynabeads^{TM}), fluorescent dyes (e.g., fluorescein, texas red, rhodamine, green fluorescent protein, and the like), radiolabels (e.g., ^{3}H, ^{125}I, 35S, ^{14}C, or ^{32}P), enzymes (e.g., horse radish peroxidase, alkaline phosphatase and others commonly used in an ELISA), and colorimetric labels such as colloidal gold or colored glass or plastic (e.g., polystyrene, polypropylene, latex, etc.) beads. Patents teaching the use of such labels include U.S. Pat. Nos. 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149; and 4,366,241, the entireties of which are incorporated by reference herein.

Means of detecting such labels are well known to those of skill in the art. Thus, for example, radiolabels may be detected using photographic film or scintillation counters, fluorescent markers may be detected using a photodetector to detect emitted light. Enzymatic labels are typically detected by providing the enzyme with a substrate and detecting the reaction product produced by the action of the enzyme on the substrate, and colorimetric labels are detected by simply visualizing the colored label.

The labels may be incorporated by any of a number of means well known to those of skill in the art. However, in a preferred embodiment, the label is simultaneously incorporated during the amplification step in the preparation of the sample nucleic acids. Thus, for example, polymerase chain reaction (PCR) with labelled primers or labelled nucleotides will provide a labelled amplification product. In a preferred embodiment, transcription amplification, as described above, using a labelled nucleotide (e.g. fluorescein-labelled UTP and/or CTP) incorporates a label into the transcribed nucleic acids.

Alternatively, a label may be added directly to the original nucleic acid sample (e.g., mRNA, polyA mRNA, cDNA, etc.) or to the amplification product after the amplification is completed. Means of attaching labels to nucleic acids are well known to those of skill in the art and include, for example, nick translation or end-labelling (e.g. with a labelled RNA) by kinasing of the nucleic acid and subsequent attachment (ligation) of a nucleic acid linker joining the sample nucleic acid to a label (e.g., a fluorophore).

In a preferred embodiment, the fluorescent modifications are by cyanine dyes e.g. Cy-3/Cy-5 dUTP, Cy-3/Cy-5 dCTP (Amersham Pharmacia) or alexa dyes (Khan, et al., 1998, *Cancer Res.* 58:5009-5013).

In a preferred embodiment, the two target samples used for comparison are labelled with different fluorescent dyes which produce distinguishable detection signals, for example, targets made from normal cartilage are labelled with Cy5 and targets made from mild osteoarthritis cartilage are labelled with Cy3. The differently labelled target samples are hybridized to the same microarray simultaneously. In a preferred embodiment, the labelled targets are purified using methods known in the art, e.g., by ethanol purification or column purification.

In a preferred embodiment, the target will include one or more control molecules which hybridize to control probes on the microarray to normalize signals generated from the microarray. Preferably, labelled normalization targets are nucleic acid sequences that are perfectly complementary to control oligonucleotides that are spotted onto the microarray as described above. The signals obtained from the normalization controls after hybridization provide a control for variations in hybridization conditions, label intensity, "reading" efficiency and other factors that may cause the signal of a perfect hybridization to vary between arrays. In a preferred embodiment, signals (e.g., fluorescence intensity) read from all other probes in the array are divided by the signal (e.g., fluorescence intensity) from the control probes, thereby normalizing the measurements.

Preferred normalization targets are selected to reflect the average length of the other targets present in the sample, however, they are selected to cover a range of lengths. The normalization control(s) also can be selected to reflect the (average) base composition of the other probes in the array, however, in a preferred embodiment, only one or a few

normalization probes are used and they are selected such that they hybridize well (i.e., have no secondary structure and do not self hybridize) and do not match any target molecules.

Normalization probes are localised at any position in the array or at multiple positions throughout the array to control for spatial variation in hybridization efficiency. In a preferred embodiment, normalization controls are located at the corners or edges of the array as well as in the middle.

Hybridization Conditions

Nucleic acid hybridization involves providing a denatured probe or target nucleic acid member and target nucleic acid under conditions where the probe or target nucleic acid member and its complementary target can form stable hybrid duplexes through complementary base pairing. The nucleic acids that do not form hybrid duplexes are then washed away leaving the hybridized nucleic acids to be detected, typically through detection of an attached detectable label. It is generally recognized that nucleic acids are denatured by increasing the temperature or decreasing the salt concentration of the buffer containing the nucleic acids. Under low stringency conditions (e.g., low temperature and/or high salt) hybrid duplexes (e.g., DNA:DNA, RNA:RNA, or RNA:DNA) will form even where the annealed sequences are not perfectly complementary. Thus specificity of hybridization is reduced at lower stringency. Conversely, at higher stringency (e.g., higher temperature or lower salt) successful hybridization requires fewer mismatches.

The invention provides for hybridization conditions comprising the Dig hybridization mix (Boehringer); or formamide-based hybridization solutions, for example as described in Ausubel et al., *supra* and Sambrook et al. *supra*.

Methods of optimizing hybridization conditions are well known to those of skill in the art (see, e.g., *Laboratory Techniques in Biochemistry and Molecular Biology*, Vol. 24: *Hybridization With Nucleic acid Probes*, P. Tijssen, ed. Elsevier, N.Y., (1993)).

Following hybridization, non-hybridized labelled or unlabeled nucleic acid is removed from the support surface, conveniently by washing, thereby generating a pattern of hybridized target nucleic acid on the substrate surface. A variety of wash solutions are known to those of skill in the art and may be used. The resultant hybridization patterns of

labelled, hybridized oligonucleotides and/or nucleic acids may be visualized or detected in a variety of ways, with the particular manner of detection being chosen based on the particular label of the test nucleic acid, where representative detection means include scintillation counting, autoradiography, fluorescence measurement, calorimetric measurement, light emission measurement and the like.

Image Acquisition and Data Analysis

Following hybridization and any washing step(s) and/or subsequent treatments, as described above, the resultant hybridization pattern is detected. In detecting or visualizing the hybridization pattern, the intensity or signal value of the label will be not only be detected but quantified, by which is meant that the signal from each spot of the hybridization will be measured and compared to a unit value corresponding to the signal emitted by a known number of end labelled target nucleic acids to obtain a count or absolute value of the copy number of each end-labelled target that is hybridized to a particular spot on the array in the hybridization pattern.

Methods for analyzing the data collected from hybridization to arrays are well known in the art. For example, where detection of hybridization involves a fluorescent label, data analysis can include the steps of determining fluorescent intensity as a function of substrate position from the data collected, removing outliers, i.e., data deviating from a predetermined statistical distribution, and calculating the relative binding affinity of the test nucleic acids from the remaining data. The resulting data is displayed as an image with the intensity in each region varying according to the binding affinity between associated oligonucleotides and/or nucleic acids and the test nucleic acids.

The following detection protocol is used for the simultaneous analysis of two cartilage samples to be compared, where each sample is labelled with a different fluorescent dye.

Each element of the microarray is scanned for the first fluorescent color. The intensity of the fluorescence at each array element is proportional to the expression level of that gene in the sample.

The scanning operation is repeated for the second fluorescent label. The ratio of the two fluorescent intensities provides a highly accurate and quantitative measurement of the relative gene expression level in the two tissue samples.

In a preferred embodiment, fluorescence intensities of immobilized target nucleic acid sequences were determined from images taken with a custom confocal microscope equipped with laser excitation sources and interference filters appropriate for the Cy3 and Cy5 fluors. Separate scans were taken for each fluor at a resolution of 225 μm^2 per pixel and 65,536 gray levels. Image segmentation to identify areas of hybridization, normalization of the intensities between the two fluor images, and calculation of the normalized mean fluorescent values at each target are as described (Khan, et al., 1998, *Cancer Res.* 58:5009-5013. Chen, et al., 1997, *Biomed. Optics* 2:364-374). Normalization between the images is used to adjust for the different efficiencies in labelling and detection with the two different fluors. This is achieved by equilibrating to a value of one the signal intensity ratio of a set of internal control genes spotted on the array.

In another preferred embodiment, the array is scanned in the Cy 3 and Cy5 channels and stored as separate 16-bit TIFF images. The images are incorporated and analysed using software which includes a gridding process to capture the hybridization intensity data from each spot on the array. The fluorescence intensity and background-subtracted hybridization intensity of each spot is collected and a ratio of measured mean intensities of Cy5 to Cy3 is calculated. A linear regression approach is used for normalization and assumes that a scatter plot of the measured Cy5 versus Cy3 intensities should have a slope of one. The average of the ratios is calculated and used to rescale the data and adjust the slope to one. A ratio of expression not equal to 1 is used as an indication of differential gene expression.

In a particularly preferred embodiment, where it is desired to quantify the transcription level (and thereby expression) of one or more nucleic acid sequences in a sample, the target nucleic acid sample is one in which the concentration of the mRNA transcript(s) of the gene or genes, or the concentration of the nucleic acids derived from the mRNA transcript(s), is proportional to the transcription level (and therefore expression level) of that gene. Similarly, it is preferred that the hybridization signal intensity be proportional

to the amount of hybridized nucleic acid. While it is preferred that the proportionality be relatively strict (e.g., a doubling in transcription rate results in a doubling in mRNA transcript in the sample nucleic acid pool and a doubling in hybridization signal), one of skill will appreciate that the proportionality can be more relaxed and even non-linear and still provide meaningful results. Thus, for example, an assay where a 5 fold difference in concentration of the target mRNA results in a 3- to 6-fold difference in hybridization intensity is sufficient for most purposes. Where more precise quantification is required, appropriate controls are run to correct for variations introduced in sample preparation and hybridization as described herein. In addition, serial dilutions of "standard" target mRNAs are used to prepare calibration curves according to methods well known to those of skill in the art. Of course, where simple detection of the presence or absence of a transcript is desired, no elaborate control or calibration is required.

For example, if an nucleic acid member on an array is not labelled after hybridization, this indicates that the gene comprising that nucleic acid member is not expressed in either sample. If a nucleic acid member is labelled with a single color, it indicates that a labelled gene was expressed only in one sample. The labelling of a nucleic acid member comprising an array with both colors indicates that the gene was expressed in both samples. Even genes expressed once per cell are detected (1 part in 100,000 sensitivity). A difference in expression intensity in the two samples being compared is indicative of differential expression, the ratio of the intensity in the two samples being not equal to 1.0, preferably less than 0.7 or greater than 1.2, more preferably less than 0.5 or greater than 1.5.

RT-PCR

In aspect of the invention, the level of the expression of the RNA products of the biomarkers of the invention can be measured by amplifying the RNA products of the biomarkers from a sample using reverse transcription (RT) in combination with the polymerase chain reaction (PCR). In accordance with one embodiment of the invention, the RT can be quantitative as would be understood to a person skilled in the art.

Total RNA, or mRNA from a sample is used as a template and a primer specific to the transcribed portion of a biomarker of the invention is used to initiate reverse transcription. Methods of reverse transcribing RNA into cDNA are well known and

described in Sambrook et al., 1989, *supra*. Primer design can be accomplished utilizing commercially available software (*e.g.*, Primer Designer 1.0, Scientific Sofware *etc* .). using methods that are standard and well known in the art.

One embodiment of a protocol used to design and select primers encompassed by the invention describes the principle and steps involved in the design of primers for use in real-time PCR with SYBR-Green assay. Preferably, this protocol uses The National Center for Biotechnology Information (NCBI) search engine and application of PrimerQuest primer design software. The PrimerQuest is web-base software developed for Integrated DNA Technologies, Inc. (IDT). This software is based on Primer3 developed by the Whitehead Institute for Biomedical Research.

In one embodiment guidelines used for designing primers encompassed by the invention are that the product or amplicon length be 100-150 bases, that the optimum Tm preferably be 60°C, with acceptable ranges from 58-62°C, and that the most preferable GC content be 50%, with preferable ranges from 45-55% also being acceptable. It is helpful that complementary strings of the three bases at the 3'-end of each primer to itself or the other primer be avoided in order to reduce "primer-dimer" formation. Also it is helpful that complementary sequences within a primer sequence and between the primers of a pair be avoided. Preferably, runs of 3 or more G's or C's at the 3'-end are avoided, as well as single base repeats greater than 3 bases. Unbalanced distribution of G/C- and A/T rich domains preferably are avoided, and in one particular embodiment the primer has a G or C is the 3'-end. It is helpful that the 3'-end of the primers not be a T since primers with a T at the 3'-end have a greater tolerance to mismatch. It is also helpful to avoid mismatches, especially at the 3'-end; and it is useful to position at least 7 unique bases at the 3'-end. Once can also select useful primers using a scoring scheme so as to help avoid self complementarity of the primers chosen. Thus for example, a score of +1 for a given match, -1 for a mismatch, -2 for a single pb gap and then primer selected with as low a score as possible, but in one embodiment primers having a score less than 5. Preferably, genomic amplification is avoided, and as such, it is preferable that any one primers should span an intron. Preferably, primers should be designed so that one half or at least 7 nucleotides of the primer hybridizes to the 3' end of one exon and the remaining to the 5' end of the adjacent exon. In addition in a further

selected embodiment, primers are designed across exon-exon junction so as to differentiate or prevent amplification of genomic sequences. In another selected embodiment, primers are designed so as to avoid known SNPs.

Primer Software programs can be used to aid in the design and selection of primers encompassed by the instant invention, such as " The Primer Quest software" which is available through the following web site link: biotools.idtdna.com/primerquest/.

The following website links are useful when searching and updating sequence information from the Human Genome Database for use in biomarker primer design: 1) the NCBI LocusLink Homepage: www.ncbi.nlm.nih.gov/LocusLink/, and 2) Ensemble Human Genome Browser: www.ensembl.org/Homo_sapiens, preferably using pertinent biomarker information such as Gene or Sequence Description, Accession or Sequence ID, Gene Symbol, RefSeq #, and/or UniGene #.

Once the correct target DNA Sequence has been obtained from which the primers will be generated, it is preferable to note the Exon-Intron Boundaries from links of the LocusLink or from the Ensembl Gene Browser for the Gene Interest. One preferable means to optimize primer design is to use the three options of BASIC, STANDARD and ADVANCE, in the PrimerQuest software.

A preferable use of the BASIC Function of PrimerQuest software is first, under Sequence Information, to enter the name of the primer into the [Name] box and Cut and paste the target sequence into [Sequence] box, selecting to design a PCR Primer using the parameter settings of Real-Time PCR. Under the standard sequence design, it is preferable to select 50 as the Number of Primer Set to Return and human as the Mispriming Library to use. In one embodiment, it is helpful to enter the following selections under the Advanced Function of Standard Primer Design: Optimum Primer Size:20 (nt), Optimum Primer Tm:60 (°C), Optimum Primer GC%:50 (%), Product Size Range: 100-150. Further, under the standard function, the following options preferably can be fine-tuned; the primer selection;Targets, Excluded Regions, Included Regions and Start Codon Position. In another embodiment.

Once the required parameters are entered or selected, the Primer Quest search for the possible primer selections is initiated producing a detail description on potential forward and reverse primers, including the actual sequence, its start position, length, Tm, GC%, product size penalties values, and a means to predict secondary structure – mFold. The following two criteria are most useful: preferably delta G should be greater than –3.0 kcal.mol-1, and preferably the TM should be less than 50°C and not greater than 55°C. The dot plot is a little more difficult to interpret, but in general it is preferable not to select a primer that produces a long diagonal line of black dots in the dot plot since it is most likely to form a hairpin.

Preferably, the primer should be unique to the target sequence and not match to a pseudogene, which can be verified by using [BLAST] to examine the specificity of the primer. Preferably, the OligoAnalyzer 3.0 provided by IDT BioTools can be used to examine the possibility of Self-Dimer and Hetero-Dimer formation. Preferably, the information and guidelines provided by IDT BioTools or Primer 3 can be used for the selection of the best possible primer pair(s) for the investigation of the Biomarkers of the instant invention. It is preferable that only those primers that produced a single amplicon with the size matched to the expected product, as determined by the melting curve analysis and agarose gel electrophoresis separation be used in the biomarker investigation.

The following related references are hereby incorporated by reference; Dieffenbach, C.W., Lowe, T.M.J., Dveksler, G.S. (1995) General Concepts for PCR Primer Design. In: PCR Primer, A Laboratory Manual (Eds. Dieffenbach, C.W, and Dveksler, G.S.) Cold Spring Harbor Laboratory Press, New York, 133-155, Innis, M.A., and Gelfand, D.H. (1990) Optimization of PCRs. In: PCR protocols, A Guide to Methods and Applications (Eds. Innis, M.A., Gelfand, D.H., Sninsky, J.J., and White, T.J.) Academic Press, San Diego, 3-12, Sharrocks, A.D. (1994) The design of primers for PCR. In: PCR Technology, Current Innovations (Eds. Griffin, H.G., and Griffin, A.M, Ed.) CRC Press, London, 5-11.

The product of the reverse transcription is subsequently used as a template for PCR.

PCR provides a method for rapidly amplifying a particular nucleic acid sequence by using multiple cycles of DNA replication catalyzed by a thermostable, DNA-

dependent DNA polymerase to amplify the target sequence of interest. PCR requires the presence of a nucleic acid to be amplified, two single-stranded oligonucleotide primers flanking the sequence to be amplified, a DNA polymerase, deoxyribonucleoside triphosphates, a buffer and salts.

The method of PCR is well known in the art. PCR, is performed as described in Mullis and Faloona, 1987, Methods Enzymol., 155: 335, which is incorporated herein by reference. PCR is performed using template DNA (at least 1 fg; more usefully, 1-1000 ng) and at least 25 pmol of oligonucleotide primers. A typical reaction mixture includes: 2 μl of DNA, 25 pmol of oligonucleotide primer, 2.5 μl of 10H PCR buffer 1 (Perkin-Elmer, Foster City, CA), 0.4 μl of 1.25 μM dNTP, 0.15 μl (or 2.5 units) of Taq DNA polymerase (Perkin Elmer, Foster City, CA) and deionized water to a total volume of 25 μl. Mineral oil is overlaid and the PCR is performed using a programmable thermal cycler.

The length and temperature of each step of a PCR cycle, as well as the number of cycles, are adjusted according to the stringency requirements in effect. Annealing temperature and timing are determined both by the efficiency with which a primer is expected to anneal to a template and the degree of mismatch that is to be tolerated. The ability to optimize the stringency of primer annealing conditions is well within the knowledge of one of moderate skill in the art. An annealing temperature of between 30°C and 72°C is used. Initial denaturation of the template molecules normally occurs at between 92°C and 99°C for 4 minutes, followed by 20-40 cycles consisting of denaturation (94-99°C for 15 seconds to 1 minute), annealing (temperature determined as discussed above; 1-2 minutes), and extension (72°C for 1 minute). The final extension step is generally carried out for 4 minutes at 72°C, and may be followed by an indefinite (0-24 hour) step at 4°C.

QRT-PCR, which is quantitative in nature, can also be performed to provide a quantitative measure of gene expression levels. In QRT-PCR reverse transcription and PCR can be performed in two steps, or reverse transcription combined with PCR can be performed concurrently. One of these techniques, for which there are commercially available kits such as Taqman (Perkin Elmer, Foster City, CA), is performed with a transcript-specific antisense probe. This probe is specific for the PCR product (*e.g.* a nucleic acid fragment derived from a gene) and is prepared with a quencher and fluorescent reporter probe complexed to the 5'

end of the oligonucleotide. Different fluorescent markers are attached to different reporters, allowing for measurement of two products in one reaction. When Taq DNA polymerase is activated, it cleaves off the fluorescent reporters of the probe bound to the template by virtue of its 5'-to-3' exonuclease activity. In the absence of the quenchers, the reporters now fluoresce. The color change in the reporters is proportional to the amount of each specific product and is measured by a fluorometer; therefore, the amount of each color is measured and the PCR product is quantified. The PCR reactions are performed in 96 well plates so that samples derived from many individuals are processed and measured simultaneously. The Taqman system has the additional advantage of not requiring gel electrophoresis and allows for quantification when used with a standard curve.

A second technique useful for detecting PCR products quantitatively without is to use an intercolating dye such as the commercially available QuantiTect SYBR Green PCR (Qiagen, Valencia California). RT-PCR is performed using SYBR green as a fluorescent label which is incorporated into the PCR product during the PCR stage and produces a flourescense proportional to the amount of PCR product.

Both Taqman and QuantiTect SYBR systems can be used subsequent to reverse transcription of RNA. Reverse transcription can either be performed in the same reaction mixture as the PCR step (one-step protocol) or reverse transcription can be performed first prior to amplification utilizing PCR (two-step protocol).

Additionally, other systems to quantitatively measure mRNA expression products are known including Molecular Beacons® which uses a probe having a fluorescent molecule and a quencher molecule, the probe capable of forming a hairpin structure such that when in the hairpin form, the fluorescence molecule is quenched, and when hybridized the flourescense increases giving a quantitative measurement of gene expression.

QRT-PCR using TaqMan™, SybrGreen, Molecular Beacons® and the like to can be used to determine a quantification value reflective of the level of RNA which is amplified. Flourescence is measured at a set threshold level of fluorescence (at least above background fluorescence) and the cycle at which the threshold level is reached is determined (Ct value) Thus the Ct value is the concentration-dependent PCR cycle number at which the fluorescence of the amplified product (amplicon) is at the preset threshold level. The

threshold value can be the level in which fluorescence becomes distinguishable over background or more preferably at a level reflective of exponential amplification. For the purposes of identifying values for input into the mathematical model, preferably ΔCt is used. ΔCt is the change in Ct value as determined between the product of the gene being measured (i.e. amplified) and a control gene (sometimes called a housekeeping gene). As would be understood by a person skilled in the art, the control gene is supposed to be a gene for which the level of expression does not change. Typical control genes are B-Actin, or GAPDH.

Additional techniques to quantitatively measure RNA expression include, but are not limited to, polymerase chain reaction, ligase chain reaction, Qbeta replicase (see, *e.g.*, International Application No. PCT/US87/00880), isothermal amplification method (see, *e.g.*, Walker et al. (1992) PNAS 89:382-396), strand displacement amplification (SDA), repair chain reaction, Asymmetric Quantitative PCR (see, *e.g.*, U.S. Publication No. US200330134307A1) and the multiplex microsphere bead assay described in Fuja et al., 2004, Journal of Biotechnology 108:193-205.

The level of gene expression can be measured by amplifying RNA from a sample using transcription based amplification systems (TAS), including nucleic acid sequence amplification (NASBA) and 3SR. See, *e.g.*, Kwoh et al (1989) PNAS USA 86:1173; International Publication No. WO 88/10315; and U.S. Patent No. 6,329,179. In NASBA, the nucleic acids may be prepared for amplification using conventional phenol/chloroform extraction, heat denaturation, treatment with lysis buffer and minispin columns for isolation of DNA and RNA or guanidinium chloride extraction of RNA. These amplification techniques involve annealing a primer that has target specific sequences. Following polymerization, DNA/RNA hybrids are digested with RNase H while double stranded DNA molecules are heat denatured again. In either case the single stranded DNA is made fully double stranded by addition of second target specific primer, followed by polymerization. The double-stranded DNA molecules are then multiply transcribed by a polymerase such as T7 or SP6. In an isothermal cyclic reaction, the RNA's are reverse transcribed into double stranded DNA, and transcribed once with a polymerase such as T7 or SP6. The resulting products, whether truncated or complete, indicate target specific sequences.

Several techniques may be used to separate amplification products. For example, amplification products may be separated by agarose, agarose-acrylamide or polyacrylamide gel electrophoresis using conventional methods. See Sambrook et al., 1989. Several techniques for detecting PCR products quantitatively without electrophoresis may also be used according to the invention (see for example *PCR Protocols, A Guide to Methods and Applications*, Innis et al., Academic Press, Inc. N.Y., (1990)). For example, chromatographic techniques may be employed to effect separation. There are many kinds of chromatography which may be used in the present invention: adsorption, partition, ion-exchange and molecular sieve, HPLC, and many specialized techniques for using them including column, paper, thin-layer and gas chromatography (Freifelder, Physical Biochemistry Applications to Biochemistry and Molecular Biology, 2nd ed., Wm. Freeman and Co., New York, N.Y., 1982).

Another example of a separation methodology is done by covalently labeling the oligonucleotide primers used in a PCR reaction with various types of small molecule ligands. In one such separation, a different ligand is present on each oligonucleotide. A molecule, perhaps an antibody or avidin if the ligand is biotin, that specifically binds to one of the ligands is used to coat the surface of a plate such as a 96 well ELISA plate. Upon application of the PCR reactions to the surface of such a prepared plate, the PCR products are bound with specificity to the surface. After washing the plate to remove unbound reagents, a solution containing a second molecule that binds to the first ligand is added. This second molecule is linked to some kind of reporter system. The second molecule only binds to the plate if a PCR product has been produced whereby both oligonucleotide primers are incorporated into the final PCR products. The amount of the PCR product is then detected and quantified in a commercial plate reader much as ELISA reactions are detected and quantified. An ELISA-like system such as the one described here has been developed by the Raggio Italgene company under the C-Track trade name.

Amplification products must be visualized in order to confirm amplification of the nucleic acid sequences of interest. One typical visualization method involves staining of a gel with ethidium bromide and visualization under UV light. Alternatively, if the amplification products are integrally labeled with radio- or fluorometrically-labeled

nucleotides, the amplification products may then be exposed to x-ray film or visualized under the appropriate stimulating spectra, following separation.

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

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

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

Nuclease Protection Assays

In another embodiment of the invention, Nuclease protection assays (including both ribonuclease protection assays and S1 nuclease assays) can be used to detect and quantitate the RNA products of the biomarkers of the invention. In nuclease protection assays, an antisense probe (labelled with, *e.g.*, radiolabeled or nonisotopic) hybridizes in solution to an RNA sample. Following hybridization, single-stranded, unhybridized probe and RNA are degraded by nucleases. An acrylamide gel is used to separate the remaining

protected fragments. Typically, solution hybridization is more efficient than membrane-based hybridization, and it can accommodate up to 100 μg of sample RNA, compared with the 20-30 μg maximum of blot hybridizations.

The ribonuclease protection assay, which is the most common type of nuclease protection assay, requires the use of RNA probes. Oligonucleotides and other single-stranded DNA probes can only be used in assays containing S1 nuclease. The single-stranded, antisense probe must typically be completely homologous to target RNA to prevent cleavage of the probe:target hybrid by nuclease.

Northern Blots

A standard Northern blot assay can also be used to ascertain an RNA transcript size, identify alternatively spliced RNA transcripts, and the relative amounts of RNA products of the biomarker of the invention, in accordance with conventional Northern hybridization techniques known to those persons of ordinary skill in the art. In Northern blots, RNA samples are first separated by size via electrophoresis in an agarose gel under denaturing conditions. The RNA is then transferred to a membrane, crosslinked and hybridized with a labelled probe. Nonisotopic or high specific activity radiolabeled probes can be used including random-primed, nick-translated, or PCR-generated DNA probes, in vitro transcribed RNA probes, and oligonucleotides. Additionally, sequences with only partial homology (*e.g.*, cDNA from a different species or genomic DNA fragments that might contain an exon) may be used as probes. The labelled probe, *e.g.*, a radiolabelled cDNA, either containing the full-length, single stranded DNA or a fragment of that DNA sequence may be any length up to at least 20, at least 30, at least 50, or at least 100 consecutive nucleotides in length. The probe can be labelled by any of the many different methods known to those skilled in this art. The labels most commonly employed for these studies are radioactive elements, enzymes, chemicals that fluoresce when exposed to ultraviolet light, and others. A number of fluorescent materials are known and can be utilised as labels. These include, but are not limited to, fluorescein, rhodamine, auramine, Texas Red, AMCA blue and Lucifer Yellow. A particular detecting material is anti-rabbit antibody prepared in goats and conjugated with fluorescein through an isothiocyanate. Proteins can also be labelled with a radioactive element or with an enzyme. The radioactive label can be detected by any of the

currently available counting procedures. Non-limiting examples of isotopes include ^{3}H, ^{14}C, ^{32}P, ^{35}S, ^{36}Cl, ^{51}Cr, ^{57}Co, ^{58}Co, ^{59}Fe, ^{90}Y, ^{125}I, ^{131}I, and ^{186}Re. Enzyme labels are likewise useful, and can be detected by any of the presently utilised colorimetric, spectrophotometric, fluorospectrophotometric, amperometric or gasometric techniques. The enzyme is conjugated to the selected particle by reaction with bridging molecules such as carbodiimides, diisocyanates, glutaraldehyde and the like. Any enzymes known to one of skill in the art can be utilised. Examples of such enzymes include, but are not limited to, peroxidase, beta-D-galactosidase, urease, glucose oxidase plus peroxidase and alkaline phosphatase. U.S. Patent Nos. 3,654,090, 3,850,752, and 4,016,043 are referred to by way of example for their disclosure of alternate labelling material and methods.

5.16 Techniques to Measure the Protein Products of the Biomarkers of the Invention

Protein Products

Standard techniques can also be utilised for determining the amount of the protein or proteins of interest present in a sample. For example, standard techniques can be employed using, *e.g.*, immunoassays such as, for example, Western blot, immunoprecipitation followed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), immunocytochemistry, and the like to determine the amount of the protein or proteins of interest present in a sample. A preferred agent for detecting a protein of interest is an antibody capable of binding to a protein of interest, preferably an antibody with a detectable label.

For such detection methods, protein from the sample to be analyzed can easily be isolated using techniques which are well known to those of skill in the art. Protein isolation methods can, for example, be such as those described in Harlow and Lane (Harlow, E. and Lane, D., *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1988)).

Preferred methods for the detection of the protein or proteins of interest involve their detection via interaction with a protein-specific antibody. For example, antibodies directed a protein of interest can be utilised as described herein. Antibodies can be generated utilising standard techniques well known to those of skill in the art. See, *e.g.*,

Section 5.19.1 of this application and Section 5.2 of U.S. Publication No. 20040018200 for a more detailed discussion of such antibody generation techniques, which is incorporated herein by reference. Briefly, such antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or an antigen binding antibody fragment (*e.g.*, Fab or $F(ab')_2$) can, for example, be used. Preferably, the antibody is a human or humanized antibody.

For example, antibodies, or fragments of antibodies, specific for a protein of interest can be used to quantitatively or qualitatively detect the presence of the protein. This can be accomplished, for example, by immunofluorescence techniques. Antibodies (or fragments thereof) can, additionally, be employed histologically, as in immunofluorescence or immunoelectron microscopy, for *in situ* detection of a protein of interest. *In situ* detection can be accomplished by removing a histological specimen (*e.g.*, a biopsy specimen) from a patient, and applying thereto a labelled antibody thereto that is directed to a protein. The antibody (or fragment) is preferably applied by overlaying the labelled antibody (or fragment) onto a biological sample. Through the use of such a procedure, it is possible to determine not only the presence of the protein of interest, but also its distribution, its presence in cells (*e.g.*, chondrocytes and lymphocytes) within the sample. A wide variety of well-known histological methods (such as staining procedures) can be utilised in order to achieve such *in situ* detection.

Immunoassays for a protein of interest typically comprise incubating a biological sample of a detectably labelled antibody capable of identifying a protein of interest, and detecting the bound antibody by any of a number of techniques well-known in the art. As discussed in more detail, below, the term "labelled" can refer to direct labelling of the antibody via, *e.g.*, coupling (*i.e.*, physically linking) a detectable substance to the antibody, and can also refer to indirect labelling of the antibody by reactivity with another reagent that is directly labelled. Examples of indirect labelling include detection of a primary antibody using a fluorescently labelled secondary antibody.

For example, the biological sample can be brought in contact with and immobilized onto a phase support or carrier such as nitrocellulose, or other support which is capable of immobilizing cells, cell particles or soluble proteins. The support can then be

washed with suitable buffers followed by treatment with the detectably labelled fingerprint gene-specific antibody. The phase support can then be washed with the buffer a second time to remove unbound antibody. The amount of bound label on support can then be detected by conventional means.

By "phase support or carrier " in the context of proteinaceous agents is intended any support capable of binding an antigen or an antibody. Well-known supports or carriers include glass, polystyrene, polypropylene, polyethylene, dextran, nylon, amylases, natural and modified celluloses, polyacrylamides, gabbros, and magnetite. The nature of the carrier can be either soluble to some extent or insoluble for the purposes of the present invention. The support material can have virtually any possible structural configuration so long as the coupled molecule is capable of binding to an antigen or antibody. Thus, the support configuration can be spherical, as in a bead, or cylindrical, as in the inside surface of a test tube, or the external surface of a rod. Alternatively, the surface can be flat such as a sheet, test strip, etc. Preferred supports include polystyrene beads. Those skilled in the art will know many other suitable carriers for binding antibody or antigen, or will be able to ascertain the same by use of routine experimentation.

One of the ways in which a specific antibody can be detectably labelled is by linking the same to an enzyme and use in an enzyme immunoassay (EIA) (Voller, A., "The Enzyme Linked Immunosorbent Assay (ELISA)", 1978, Diagnostic Horizons 2:1-7, Microbiological Associates Quarterly Publication, Walkersville, MD); Voller, A. et al., 1978, J. Clin. Pathol. 31:507-520; Butler, J.E., 1981, Meth. Enzymol. 73:482-523; Maggio, E. (ed.), 1980, *Enzyme Immunoassay*, CRC Press, Boca Raton, FL; Ishikawa, E. et al., (eds.), 1981, *Enzyme Immunoassay*, Kgaku Shoin, Tokyo). The enzyme which is bound to the antibody will react with an appropriate substrate, preferably a chromogenic substrate, in such a manner as to produce a chemical moiety which can be detected, for example, by spectrophotometric, fluorimetric or by visual means. Enzymes which can be used to detectably label the antibody include, but are not limited to, malate dehydrogenase, staphylococcal nuclease, delta-5-steroid isomerase, yeast alcohol dehydrogenase, alpha-glycerophosphate, dehydrogenase, triose phosphate isomerase, horseradish peroxidase, alkaline phosphatase, asparaginase, glucose oxidase, beta-galactosidase, ribonuclease, urease, catalase, glucose-6-phosphate

dehydrogenase, glucoamylase and acetylcholinesterase. The detection can be accomplished by colorimetric methods which employ a chromogenic substrate for the enzyme. Detection can also be accomplished by visual comparison of the extent of enzymatic reaction of a substrate in comparison with similarly prepared standards.

Detection can also be accomplished using any of a variety of other immunoassays. For example, by radioactively labelling the antibodies or antibody fragments, it is possible to detect a protein of interest through the use of a radioimmunoassay (RIA) (see, for example, Weintraub, B., Principles of Radioimmunoassays, Seventh Training Course on Radioligand Assay Techniques, The Endocrine Society, March, 1986, which is incorporated by reference herein). The radioactive isotope (*e.g.*, ^{125}I, ^{131}I, ^{35}S or ^{3}H) can be detected by such means as the use of a gamma counter or a scintillation counter or by autoradiography.

It is also possible to label the antibody with a fluorescent compound. When the fluorescently labelled antibody is exposed to light of the proper wavelength, its presence can then be detected due to fluorescence. Among the most commonly used fluorescent labelling compounds are fluorescein isothiocyanate, rhodamine, phycoerythrin, phycocyanin, allophycocyanin, o -phthaldehyde and fluorescamine.

The antibody can also be detectably labelled using fluorescence emitting metals such as ^{152}Eu, or others of the lanthanide series. These metals can be attached to the antibody using such metal chelating groups as diethylenetriaminepentacetic acid (DTPA) or ethylenediaminetetraacetic acid (EDTA).

The antibody also can be detectably labelled by coupling it to a chemiluminescent compound. The presence of the chemiluminescent-tagged antibody is then determined by detecting the presence of luminescence that arises during the course of a chemical reaction. Examples of particularly useful chemiluminescent labelling compounds are luminol, isoluminol, theromatic acridinium ester, imidazole, acridinium salt and oxalate ester.

Likewise, a bioluminescent compound can be used to label the antibody of the present invention. Bioluminescence is a type of chemiluminescence found in biological systems in, which a catalytic protein increases the efficiency of the chemiluminescent

reaction. The presence of a bioluminescent protein is determined by detecting the presence of luminescence. Important bioluminescent compounds for purposes of labelling are luciferin, luciferase and aequorin.

Protein Arrays

Polypeptides which specifically and/or selectively bind to the protein products of the biomarkers of the invention can be immobilized on a protein array. The protein array can be used as a diagnostic tool, *e.g.*, to screen medical samples (such as bladder tissue and/or blood) for the presence of the polypeptides protein products of the biomarkers of the invention. The protein array can also include antibodies as well as other ligands, *e.g.*, that bind to the polypeptides encoded by the biomarkers of the invention.

Methods of producing polypeptide arrays are described, *e.g.*, in De Wildt et al., 2000, Nature Biotech. 18:989-994; Lueking et al., 1999, Anal. Biochem. 270:103-111; Ge, 2000, Nuc. Acids Res. 28:e3 ; MacBeath and Schreiber, 2000, Science 289:1760-1763; International Publication Nos. WO 01/40803 and WO 99/51773A1; and U.S. Patent No. 6,406,921. Polypeptides for the array can be spotted at high speed, *e.g.*, using commercially available robotic apparatus, *e.g.*, from Genetic MicroSystems and Affymetrix (Santa Clara, California, USA) or BioRobotics (Cambridge, UK). The array substrate can be, for example, nitrocellulose, plastic, glass, *e.g.*, surface-modified glass. The array can also include a porous matrix, *e.g.*, acrylamide, agarose, or another polymer.

For example, the array can be an array of antibodies, *e.g.*, as described in De Wildt, *supra*. Cells that produce the polypeptide ligands can be grown on a filter in an arrayed format. Polypeptide production is induced, and the expressed antibodies are immobilized to the filter at the location of the cell. Information about the extent of binding at each address of the array can be stored as a profile, *e.g.*, in a computer database.

In one embodiment the array is an array of protein products comprising of any number of up to 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, all or any combination of the biomarkers of the invention. In another embodiment the array is an array of protein products consisting essentially of any number of up to 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, all or any combination of the biomarkers listed in Table 1. In

another embodiment the array is an array of protein products consisting essentially of any number of up to 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, all or any combination of the biomarkers listed in Table 1 including those noted in Table 3 and any number of up to 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50 from Table 2 wherein at least one biomarker of said combinations is selected from Table 1. In another embodiment the array is an array of protein products consisting essentially of any number of up to 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, all or any combination of the biomarkers listed in Table 4. In another embodiment the array is an array of protein products consisting essentially of any number of up to 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, all or any combination of the biomarkers listed in Table 4 including those noted in Table 6 and any number of up to 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50 from Table 5 wherein at least one biomarker of said combinations is selected from Table 4 In another embodiment the array is an array of protein products consisting essentially of any number of up to 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, all or any combination of the biomarkers listed in Table 7.

In one aspect, the invention provides for antibodies or antigen binding fragments thereof, that are bound to an array which selectively bind to the protein products of the biomarkers of the invention.

5.17 Protein Production

Standard recombinant nucleic acid methods can be used to express a polypeptide or antibody of the invention (*e.g.*, a protein product of a biomarker of the invention). Generally, a nucleic acid sequence encoding the polypeptide is cloned into a nucleic acid expression vector. Of course, if the protein includes multiple polypeptide chains, each chain must be cloned into an expression vector, *e.g.*, the same or different vectors, that are expressed in the same or different cells. If the protein is sufficiently small, *i.e.*, the protein is a peptide of less than 50 amino acids, the protein can be synthesised using automated organic synthetic methods. Polypeptides comprising the 5' region, 3' region or internal coding region of a biomarker of the invention, are expressed from nucleic acid expression vectors containing only those nucleotide sequences corresponding to the 5' region,

3' region or internal coding region of a biomarker of the invention. Methods for producing antibodies directed to protein products of a biomarker of the invention, or polypeptides encoded by the 5' region, 3' region or internal coding regions of a biomarker of the invention.

The expression vector for expressing the polypeptide can include, in addition to the segment encoding the polypeptide or fragment thereof, regulatory sequences, including for example, a promoter, operably linked to the nucleic acid(s) of interest. Large numbers of suitable vectors and promoters are known to those of skill in the art and are commercially available for generating the recombinant constructs of the present invention. The following vectors are provided by way of example. Bacterial: pBs, phagescript, PsiX174, pBluescript SK, pBs KS, pNH8a, pNH16a, pNH18a, pNH46a (Stratagene, La Jolla, California, USA); pTrc99A, pKK223-3, pKK233-3, pDR540, and pRIT5 (Pharmacia, Uppsala, Sweden). Eukaryotic: pWLneo, pSV2cat, pOG44, PXTI, pSG (Stratagene) pSVK3, pBPV, pMSG, and pSVL (Pharmacia). One preferred class of preferred libraries is the display library, which is described below.

Methods well known to those skilled in the art can be used to construct vectors containing a polynucleotide of the invention and appropriate transcriptional/translational control signals. These methods include *in vitro* recombinant DNA techniques, synthetic techniques and *in vivo* recombination/genetic recombination. See, for example, the techniques described in Sambrook & Russell, *Molecular Cloning: A Laboratory Manual*, 3rd Edition, Cold Spring Harbor Laboratory, N.Y. (2001) and Ausubel *et al.*, *Current Protocols in Molecular Biology* (Greene Publishing Associates and Wiley Interscience, N.Y. (1989). Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda P, and trc. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, mouse metallothionein-I, and various art-known tissue specific promoters. In specific embodiments, the promoter is an inducible promoter. In other embodiments, the promoter is a constitutive promoter. In yet other embodiments, the promoter is a tissue-specific promoter.

Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, *e.g.*, the ampicillin resistance gene of *E. coli* and *S. cerevisiae* auxotrophic markers (such as *URA3, LEU2, HIS3,* and *TRP1* genes), and a promoter derived from a highly expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), a-factor, acid phosphatase, or heat shock proteins, among others. The polynucleotide of the invention is assembled in appropriate phase with translation initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated protein into the periplasmic space or extracellular medium. Optionally, a nucleic acid of the invention can encode a fusion protein including an N-terminal identification peptide imparting desired characteristics, *e.g.*, stabilization or simplified purification of expressed recombinant product. Useful expression-vectors for bacteria are constructed by inserting a polynucleotide of the invention together with suitable translation initiation and termination signals, optionally in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and to, if desirable, provide amplification within the host. Suitable prokaryotic hosts for transformation include *E. coli, Bacillus subtilis, Salmonella typhimurium* and various species within the genera Pseudomonas, Streptomyces, and Staphylococcus, although others may also be employed as a matter of choice.

As a representative but nonlimiting example, useful expression vectors for bacteria can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden) and pGEM1 (Promega, Madison, Wisconsin, USA).

The present invention provides host cells genetically engineered to contain the polynucleotides of the invention. For example, such host cells may contain nucleic acids of the invention introduced into the host cell using known transformation, transfection or infection methods. The present invention also provides host cells genetically engineered to

express the polynucleotides of the invention, wherein such polynucleotides are in operative association with a regulatory sequence heterologous to the host cell which drives expression of the polynucleotides in the cell.

The present invention further provides host cells containing the vectors of the present invention, wherein the nucleic acid has been introduced into the host cell using known transformation, transfection or infection methods. The host cell can be a eukaryotic host cell, such as a mammalian cell, a lower eukaryotic host cell, such as a yeast cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Introduction of the recombinant construct into the host cell can be effected, for example, by calcium phosphate transfection, DEAE, dextran mediated transfection, or electroporation (Davis, L. *et al.*, *Basic Methods in Molecular Biology* (1986)). Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention.

Any host/vector system can be used to express one or more of the genes listed in Table 2 or splice variants. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook *et al.*, in *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor, New York (1989), the disclosure of which is incorporated herein by reference in its entirety. The most preferred host cells are those which do not normally express the particular polypeptide or which expresses the polypeptide at low natural level.

In a specific embodiment, the host cells are engineered to express an endogenous gene comprising the polynucleotides of the invention under the control of inducible regulatory elements, in which case the regulatory sequences of the endogenous gene may be replaced by homologous recombination. As described herein, gene targeting can be used to replace a gene's existing regulatory region with a regulatory sequence isolated from a different gene or a novel regulatory sequence synthesised by genetic engineering methods. Such regulatory sequences may be comprised of promoters, enhancers, scaffold-attachment regions, negative regulatory elements, transcriptional initiation sites, regulatory protein binding sites or combinations of said sequences. Alternatively, sequences which affect the structure or stability of the RNA or protein produced may be replaced, removed, added, or otherwise modified by targeting, including polyadenylation signals. mRNA

stability elements, splice sites, leader sequences for enhancing or modifying transport or secretion properties of the protein, or other sequences which alter or improve the function or stability of protein or RNA molecules.

The host of the present invention may also be a yeast or other fungi. In yeast, a number of vectors containing constitutive or inducible promoters may be used. For a review see, Ausubel et al. (eds), *Current Protocols in Molecular Biology*, Vol. 2, Greene Publish. Assoc. & Wiley Interscience, Ch. 13 (1988); Grant et al., 1987, "Expression and Secretion Vectors for Yeast", Methods Enzymol. 153:516-544; Glover, *DNA Cloning*, Vol. II, IRL Press, Wash., D.C., Ch. 3 (1986); Bitter, 1987, "Heterologous Gene Expression in Yeast", Methods Enzymol. 152:673-684; and Strathern et al. (eds), *The Molecular Biology of the Yeast Saccharomyces*, Cold Spring Harbor Press, Vols. I and II (1982).

Potentially suitable yeast strains include *Saccharomyces cerevisiae, Schizosaccharomyces pombe, Kluyveromyces* strains, *Candida*, or any yeast strain capable of expressing heterologous proteins. Potentially suitable bacterial strains include *Escherichia coli,* enterobacteriaceae such as *Serratia marescans*, bacilli such as *Bacillus subtilis, Salmonella typhimurium*, pseudomonads or any bacterial strain capable of expressing heterologous proteins. If the protein is made in yeast or bacteria, it may be necessary to modify the protein produced therein, for example by phosphorylation or glycosylation of the appropriate sites, in order to obtain the functional protein. Such covalent attachments may be accomplished using known chemical or enzymatic methods.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the monkey COS cells such as COS-7 lines of monkey kidney fibroblasts, described by Gluzman, 1981, Cell 23:175 (1981), Chinese Hamster Ovary (CHO) cells, human kidney 293 cells, human epidermal A431 cells, human Colo205 cells, 3T3 cells, CV-1 cells, normal diploid cells, cell strains derived from *in vitro* culture of primary tissue, primary explants, HeLa cells, mouse L cells, BHK, HL-60, U937, HaK, C127, 3T3, or Jurkat cells, and other cell lines capable of expressing a compatible vector. Mammalian expression vectors will comprise an origin of replication, a suitable promoter and also any necessary ribosome-binding sites,

polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences.

Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents. Recombinant polypeptides produced in bacterial culture are usually isolated by initial extraction from cell pellets, followed by one or more salting-out, aqueous ion exchange or size exclusion chromatography steps. In some embodiments, the template nucleic acid also encodes a polypeptide tag, *e.g.*, penta- or hexa-histidine.

Recombinant proteins can be isolated using an technique well-known in the art. Scopes (*Protein Purification: Principles and Practice*, Springer-Verlag, New York (1994)), for example, provides a number of general methods for purifying recombinant (and non-recombinant) proteins. The methods include, *e.g.*, ion-exchange chromatography, size-exclusion chromatography, affinity chromatography, selective precipitation, dialysis, and hydrophobic interaction chromatography.

Variations, modifications, and other implementations of what is described herein will occur to those of ordinary skill in the art without departing from the spirit and scope of the invention.

In order that the invention described herein may be more fully understood, the following example is set forth. It should be understood that this example is for illustrative purposes only and are not to be construed as limiting this invention in any manner.

5.18 Methods for Identifying Compounds for Use in the Prevention, Treatment, Management or Amelioration of Bladder Cancer or a Symptom Thereof

5.18.1 Methods for Identifying Compounds that Modulate the Expression or Activity of a Biomarker

The present invention provides methods of identifying compounds that bind to the products of the biomarkers of the invention. The present invention also provides methods for identifying compounds that modulate the expression and/or activity of the products of the biomarkers of the invention. The compounds identified via such methods are useful for the development of one or more animal models to study bladder cancer. Further, the compounds identified via such methods are useful as lead compounds in the development of prophylactic

and therapeutic compositions for prevention, treatment, management and/or amelioration of osteoarthritis or a symptom thereof. Such methods are particularly useful in that the effort and great expense involved in testing potential prophylactics and therapeutics *in vivo* is efficiently focused on those compounds identified via the *in vitro* and *ex vivo* methods described herein.

The present invention provides a method for identifying a compound to be tested for an ability to prevent, treat, manage or ameliorate bladder cancer or a symptom thereof, said method comprising: (a) contacting a cell expressing a protein product of one or more biomarkers of the invention or a fragment thereof, or a RNA product of one or more biomarkers of the invention or a fragment thereof with a test compound; and (b) determining the ability of the test compound to bind to the protein product, protein fragment, RNA product, or RNA portion so that if a compound binds to the protein product, protein fragment, RNA product, RNA portion, a compound to be tested for an ability to prevent, treat, manage or ameliorate bladder cancer or a symptom thereof is identified. The cell, for example, can be a yeast cell or a cell of mammalian origin. Determining the ability of the test compound to bind to the protein product, protein fragment, RNA product, or RNA portion can be accomplished, for example, by coupling the test compound with a radioisotope or enzymatic label such that binding of the test compound to the protein product, protein fragment, RNA product, or RNA portion can be determined by detecting the labelled compound in a complex. For example, test compounds can be labelled with ^{125}I, ^{35}S, ^{14}C, or ^{3}H, either directly or indirectly, and the radioisotope detected by direct counting of radioemmission or by scintillation counting. Alternatively, test compounds can be enzymatically labelled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product. In a specific embodiment, the assay comprises contacting a cell which expresses a protein product of one or more biomarkers of the invention or a fragment thereof, or a RNA product of one or more biomarkers of the invention or a fragment thereof, with a known compound which binds the protein product, protein fragment, RNA product, or RNA portion to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the protein product, protein fragment, RNA product, or RNA portion, wherein determining the ability of the test compound to interact with the

protein product, protein fragment, RNA product, or RNA portion comprises determining the ability of the test compound to preferentially bind to the protein product, protein fragment, RNA product, or RNA portion as compared to the known compound.

The present invention provides a method for identifying a compound to be tested for an ability to prevent, treat, manage or ameliorate bladder cancer or a symptom thereof, said method comprising: (a) contacting a protein product of one or more biomarkers of the invention or a fragment thereof, or a RNA product of one or more biomarkers of the invention or a portion thereof with a test compound; and (b) determining the ability of the test compound to bind to the protein product, protein fragment, RNA product, or RNA portion so that if a compound binds to the protein product, protein fragment, RNA product, or RNA portion, a compound to be tested for an ability to prevent, treat, manage or ameliorate bladder cancer or a symptom thereof is identified. Binding of the test compound to the protein product or protein fragment can be determined either directly or indirectly. In a specific embodiment, the assay includes contacting a protein product of one or more biomarkers of the invention or a fragment thereof, or a RNA product of one or more biomarkers of the invention or a portion thereof with a known compound which binds the protein product, protein fragment, RNA product, or RNA portion to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the protein product, protein fragment, RNA product, or RNA portion, wherein determining the ability of the test compound to interact with the protein product, protein fragment, RNA product, or RNA portion comprises determining the ability of the test compound to preferentially bind to the protein product, protein fragment, RNA product, or RNA portion as compared to the known compound. Techniques well known in the art can be used to determine the binding between a test compound and a protein product of a biomarker of the invention or a fragment thereof, or a RNA product of a biomarker of the invention or a portion thereof.

In some embodiments of the above assay methods of the present invention, it may be desirable to immobilize a RNA product of a biomarker of the invention or a portion thereof, or its target molecule to facilitate separation of complexed from uncomplexed forms of the RNA product or RNA portion, the target molecule or both, as well as to accommodate

automation of the assay. In more than one embodiment of the above assay methods of the present invention, it may be desirable to immobilize either a protein product of a biomarker of the invention or a fragment thereof, or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to a protein product of a biomarker of the invention or a fragment thereof can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows one or both of the proteins to be bound to a matrix. For example, glutathione-S-transferase (GST) fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical; St. Louis, MO) or glutathione derivatized microtiter plates, which are then combined with the test compound or the test compound and either the non-adsorbed target protein or a protein product of a biomarker of the invention or a fragment thereof, and the mixture incubated under conditions conducive to complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components and complex formation is measured either directly or indirectly, for example, as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of binding of a protein product of a biomarker of the invention or a fragment thereof can be determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either a protein product of a biomarker of the invention or a fragment thereof, or a target molecule can be immobilized utilising conjugation of biotin and streptavidin. A biotinylated protein product of a biomarker of the invention or a target molecule can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well known in the art (*e.g.*, biotinylation kit, Pierce Chemicals; Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with a protein product of a biomarker of the invention or a fragment thereof can be derivatized to the wells of the plate, and protein trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of

complexes using antibodies reactive with a protein product of a biomarker of the invention, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with a protein product of a biomarker of the invention or a fragment thereof, or target molecule.

The interaction or binding of a protein product of a biomarker of the invention or a fragment thereof to a test compound can also be determined using such proteins or protein fragments as "bait proteins" in a two-hybrid assay or three hybrid assay (*see, e.g.*, U.S. Patent No. 5,283,317; Zervos et al. (1993) Cell 72:223-232; Madura et al. (1993) J. Biol. Chem. 268:12046-12054; Bartel et al. (1993) Bio/Techniques 14:920-924; Iwabuchi et al. (1993) Oncogene 8:1693-1696; and International Publication No. WO 94/10300).

The present invention provides a method for identifying a compound to be tested for an ability to prevent, treat, manage or ameliorate osteoarthritis or a symptom thereof, said method comprising: (a) contacting a cell expressing a protein or RNA product of one or more biomarkers of the invention with a test compound; (b) after an incubation period, determining the amount of the protein or RNA product present in (a); and (c) comparing the amount in (a) to that present in a corresponding control cell that has not been contacted with the test compound, so that if the amount of the protein or RNA product is altered relative to the amount in the control, a compound to be tested for an ability to prevent, treat, manage or ameliorate osteoarthritis or a symptom thereof is identified. In a specific embodiment, the expression level(s) is altered by 5%, 10%, 15%, 25%, 30%, 40%, 50%, 5 to 25%, 10 to 30%, at least 1 fold, at least 1.5 fold, at least 2 fold, 4 fold, 5 fold, 10 fold, 25 fold, 1 to 10 fold, or 5 to 25 fold relative to the expression level in the control as determined by utilising an assay described herein (*e.g.*, a microarray or RT-PCR) or an assay well known to one of skill in the art. In alternate embodiments, such a method comprises determining the amount of the protein or RNA product of any of at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 12, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, 1 to 5, 1-10, 5-10, 5-25, or 10-40, all or any combination of the biomarkers of the invention as listed in Table 1 (including those specific products noted in Table 3), or as listed in table 1 (including those specific products noted in Table 3) in combination with any one or more of the products

of the biomarkers listed in Table 2, present in the cell and comparing the amounts to those present in the control.

In yet other alternate embodiments, such a method comprises determining the amount of the protein or RNA product of any of at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 12, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, 1 to 5, 1-10, 5-10, 5-25, or 10-40, all or any combination of the biomarkers of the invention as listed in Table 4 (including those specific products noted in Table 6), or as listed in table 4 (including those specific products noted in Table 6) in combination with any one or more of the products of the biomarkers listed in Table 5, present in the cell and comparing the amounts to those present in the control.

The cells utilised in the cell-based assays described herein can be engineered to express a biomarker of the invention utilising techniques known in the art. *See, e.g.*, Section III entitled "Recombinant Expression Vectors and Host Cells" of U.S. Patent No. 6,245,527, which is incorporated herein by reference. Alternatively, cells that endogenously express a biomarker of the invention can be used. For example, immortalized bladder tissue cells can be used.

In a specific embodiment, chondrocytes are isolated from a "normal" individual, or an individual with bladder cancer and/or early stage bladder cancer and are incubated in the presence and absence of a test compound for varying amounts of time (*i.e.*, 30 min, 1 hr, 5 hr, 24 hr, 48 hr and 96 hrs). When screening for prophylactic or therapeutic agents, a clone of the full sequence of a biomarker of the invention or functional portion thereof is used to transfect immortalized bladder tissue cells. The transfected cells are cultured for varying amounts of time (*i.e.*, 1, 2, 3, 5, 7, 10, or 14 days) in the presence or absence of test compound. Following incubation, target nucleic acid samples are prepared from the cells and hybridized to a nucleic acid probe corresponding to a nucleic acid sequence which is differentially expressed in a cell derived from normal, bladder cancer or early bladder cancer cells. The nucleic acid probe is labelled, for example, with a radioactive label, according to methods well-known in the art and described herein. Hybridization is carried out by northern blot, for example as described in Ausubel et al., *supra* or Sambrook et

al., *supra*). The differential hybridization, as defined herein, of the target to the samples on the array from normal relative to RNA from any one of bladder cancer or early bladder cancer is indicative of the level of expression of RNA corresponding to a differentially expressed cell specific nucleic acid sequence. A change in the level of expression of the target sequence as a result of the incubation step in the presence of the test compound, is indicative of a compound that increases or decreases the expression of the corresponding chondrocyte specific nucleic acid sequence.

The present invention also provides a method for identifying a compound to be tested for an ability to prevent, treat, manage or ameliorate bladder cancer or a symptom thereof, said method comprises: (a) contacting a cell-free extract with a nucleic acid sequence encoding a protein or RNA product of one or more biomarkers of the invention and a test compound; (b) determining the amount of the protein or RNA product present in (a); and (c) comparing the amount(s) in (a) to that present to a corresponding control that has not been contacted with the test compound, so that if the amount of the protein or RNA product is altered relative to the amount in the control, a compound to be tested for an ability to prevent, treat, manage or ameliorate bladder cancer or a symptom thereof is identified. In a specific embodiment, the expression level(s) is altered by 5%, 10%, 15%, 25%, 30%, 40%, 50%, 5 to 25%, 10 to 30%, at least 1 fold, at least 1.5 fold, at least 2 fold, 4 fold, 5 fold, 10 fold, 25 fold, 1 to 10 fold, or 5 to 25 fold relative to the expression level in the control sample determined by utilising an assay described herein (*e.g.*, a microarray or RT-PCR) or an assay well known to one of skill in the art. In alternate embodiments, such a method comprises determining the amount of a protein or RNA product of at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 12, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, 1 to 5, 1-10, 5-10, 5-25, or 10-40, all or any combination of the biomarkers of the invention present in the extract and comparing the amounts to those present in the control.

In certain embodiments, the amount of RNA product of a biomarker of the invention is determined, in other embodiments, the amount of protein product of a biomarker of the invention is determined, while in still other embodiments, the amount of RNA and protein product of a biomarker of the invention is determined. Standard methods and

compositions for determining the amount of RNA or protein product of a biomarker of the invention can be utilised. Such methods and compositions are described in detail above.

In specific embodiments, in a screening assay described herein, the amount of protein or RNA product of a biomarker of the invention is determined utilising kits. Such kits comprise materials and reagents required for measuring the expression of any number up to at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, or more protein or RNA products of at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, all or any combination of the biomarkers of the invention. In specific embodiments, the kits may further comprise one or more additional reagents employed in the various methods, such as: (1) reagents for purifying RNA from blood, or tissue; (2) primers for generating test nucleic acids; (3) dNTPs and/or rNTPs (either premixed or separate), optionally with one or more uniquely labeled dNTPs and/or rNTPs (*e.g.*, biotinylated or Cy3 or Cy5 tagged dNTPs); (4) post synthesis labeling reagents, such as chemically active derivatives of fluorescent dyes; (5) enzymes, such as reverse transcriptases, DNA polymerases, and the like; (6) various buffer mediums, *e.g.*, hybridization and washing buffers; (7) labeled probe purification reagents and components, like spin columns, etc.; and (8) protein purification reagents; (9) signal generation and detection reagents, *e.g.*, streptavidin-alkaline phosphatase conjugate, chemifluorescent or chemiluminescent substrate, and the like. In particular embodiments, the kits comprise prelabeled quality controlled protein and or RNA transcript (preferably, mRNA) for use as a control.

In some embodiments, the kits are qRT-PCR kits. In other embodiments, the kits are nucleic acid arrays and protein arrays. Such kits according to the subject invention will at least comprise an array having associated protein or nucleic acid members of the invention and packaging means therefore. Alternatively the protein or nucleic acid members of the invention may be prepackaged onto an array.

In a specific embodiment, kits for measuring a RNA product of a biomarker of the invention comprise materials and reagents that are necessary for measuring the expression

of the RNA product. For example, a microarray or RT-PCR kit may be used and contain only those reagents and materials necessary for measuring the levels of RNA products of any number of up to at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, all or any combination of the biomarkers of the invention. Alternatively, in some embodiments, the kits can comprise materials and reagents that are not limited to those required to measure the levels of RNA products of any number of up to 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, all or any combination of the biomarkers of the invention. For example, a microarray kit may contain reagents and materials necessary for measuring the levels of RNA products any number of up to 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, all or any combination of the biomarkers of the invention, in addition to reagents and materials necessary for measuring the levels of the RNA products of any number of up to at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50 or more genes other than the biomarkers of the invention. In a specific embodiment, a microarray or RT-PCR kit contains reagents and materials necessary for measuring the levels of RNA products of any number of up to at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, all or any combination of the biomarkers of the invention, and any number of up to 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 125, 150, 175, 200, 225, 250, 300, 350, 400, 450, 500, 1000, 5000, 15,000 20,000 or more genes that are not biomarkers of the invention, or any number of 1-10, 1-100, 1-150, 1-200, 1-300, 1-400, 1-500, 1-1000, 25-100, 25-200, 25-300, 25-400, 25-500, 25-1000, 100-150, 100-200, 100-300, 100-400, 100-500, 100-1000 or 500-1000, 1000 – 5000, 5000 – 10,000, 10,000 – 20,000 or more genes that are not biomarkers of the invention.

For nucleic acid micoarray kits, the kits generally comprise probes attached to a support surface. The probes may be labeled with a detectable label. In a specific embodiment, the probes are specific for the 5' region, the 3' region, the internal coding region, an exon(s), an intron(s), an exon junction(s), or an exon-intron junction(s), of any number of up to 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, all or any

combination of the biomarkers of the invention. The microarray kits may comprise instructions for performing the assay and methods for interpreting and analyzing the data resulting from the performance of the assay. The kits may also comprise hybridization reagents and/or reagents necessary for detecting a signal produced when a probe hybridizes to a target nucleic acid sequence. Generally, the materials and reagents for the microarray kits are in one or more containers. Each component of the kit is generally in its own a suitable container.

For RT-PCR and/or qRT-PCR kits, the kits generally comprise pre-selected primers specific for particular RNA products (*e.g.*, an exon(s), an intron(s), an exon junction(s), and an exon-intron junction(s)) of any number of up to 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, all or any combination of the biomarkers of the invention. The RT-PCR and/or qRT-PCR kits may also comprise enzymes suitable for reverse transcribing and/or amplifying nucleic acids (*e.g.*, polymerases such as Taq), and deoxynucleotides and buffers needed for the reaction mixture for reverse transcription and amplification. The RT-PCR kits may also comprise probes specific for any number of up to 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, all or any combination of the biomarkers of the invention. The probes may or may not be labeled with a detectable label (*e.g.*, a fluorescent label). Each component of the RT-PCR kit is generally in its own suitable container. Thus, these kits generally comprise distinct containers suitable for each individual reagent, enzyme, primer and probe. Further, the RT-PCR kits may comprise instructions for performing the assay and methods for interpreting and analyzing the data resulting from the performance of the assay.

For antibody based kits, the kit can comprise, for example: (1) a first antibody (which may or may not be attached to a support) which binds to protein of interest (*e.g.*, a protein product of any number of up to 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, all or any combination of the biomarkers of the invention); and, optionally, (2) a second, different antibody which binds to either the protein, or the first antibody and is conjugated to a detectable label (*e.g.*, a fluorescent label, radioactive isotope or enzyme). The antibody-based kits may also comprise beads for conducting an immunoprecipitation. Each component of the antibody-based kits is generally in its own suitable container. Thus, these kits generally comprise distinct containers suitable for each antibody. Further, the antibody-

based kits may comprise instructions for performing the assay and methods for interpreting and analyzing the data resulting from the performance of the assay.

Reporter gene-based assays may also be conducted to identify a compound to be tested for an ability to prevent, treat, manage or ameliorate osteoarthritis or a symptom thereof. In a specific embodiment, the present invention provides a method for identifying a compound to be tested for an ability to prevent, treat, manage or ameliorate osteoarthritis or a symptom thereof, said method comprising: (a) contacting a compound with a cell expressing a reporter gene construct comprising a reporter gene operably linked to a regulatory element of a biomarker of the invention (*e.g.*, a promoter/enhancer element); (b) measuring the expression of said reporter gene; and (c) comparing the amount in (a) to that present in a corresponding control cell that has not been contacted with the test compound, so that if the amount of expressed reporter gene is altered relative to the amount in the control cell, a compound to be tested for an ability to prevent, treat, manage or ameliorate osteoarthritis or a symptom thereof is identified. In accordance with this embodiment, the cell may naturally express the biomarker or be engineered to express the biomarker. In another embodiment, the present invention provides a method for identifying a compound to be tested for an ability to prevent, treat, manage or ameliorate osteoarthritis or a symptom thereof, said method comprising: (a) contacting a compound with a cell-free extract and a reporter gene construct comprising a reporter gene operably linked to a regulatory element of a biomarker of the invention (*e.g.*, a promoter/enhancer element); (b) measuring the expression of said reporter gene; and (c) comparing the amount in (a) to that present in a corresponding control that has not been contacted with the test compound, so that if the amount of expressed reporter gene is altered relative to the amount in the control, a compound to be tested for an ability to prevent, treat, manage or ameliorate osteoarthritis or a symptom thereof is identified.

Any reporter gene well-known to one of skill in the art may be used in reporter gene constructs used in accordance with the methods of the invention. Reporter genes refer to a nucleotide sequence encoding a RNA transcript or protein that is readily detectable either by its presence (by, *e.g.*, RT-PCR, Northern blot, Western Blot, ELISA, *etc.*) or activity. Non-limiting examples of reporter genes are listed in Table 9, *infra*. Reporter genes may be obtained and the nucleotide sequence of the elements determined by any method well-known

to one of skill in the art. The nucleotide sequence of a reporter gene can be obtained, *e.g.*, from the literature or a database such as GenBank. Alternatively, a polynucleotide encoding a reporter gene may be generated from nucleic acid from a suitable source. If a clone containing a nucleic acid encoding a particular reporter gene is not available, but the sequence of the reporter gene is known, a nucleic acid encoding the reporter gene may be chemically synthesised or obtained from a suitable source (*e.g.*, a cDNA library, or a cDNA library generated from, or nucleic acid, preferably poly A+ RNA, isolated from, any tissue or cells expressing the reporter gene) by PCR amplification. Once the nucleotide sequence of a reporter gene is determined, the nucleotide sequence of the reporter gene may be manipulated using methods well-known in the art for the manipulation of nucleotide sequences, *e.g.*, recombinant DNA techniques, site directed mutagenesis, PCR, etc. (see, for example, the techniques described in Sambrook et al., 1990, Molecular Cloning, A Laboratory Manual, 2d Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY and Ausubel et al., eds., 1998, Current Protocols in Molecular Biology, John Wiley & Sons, NY, which are both incorporated by reference herein in their entireties), to generate reporter genes having a different amino acid sequence, for example to create amino acid substitutions, deletions, and/or insertions.

Table 9: Reporter Genes and the Properties of the Reporter Gene Products

Reporter Gene	Protein Activity & Measurement
CAT (chloramphenicol acetyltransferase)	Transfers radioactive acetyl groups to chloramphenicol or detection by thin layer chromatography and autoradiography
GAL (beta-galactosidase)	Hydrolyzes colorless galactosides to yield colored products.
GUS (beta-glucuronidase)	Hydrolyzes colorless glucuronides to yield colored products.

LUC (luciferase)	Oxidizes luciferin, emitting photons
GFP (green fluorescent protein)	Fluorescent protein without substrate
SEAP (secreted alkaline phosphatase)	Luminescence reaction with suitable substrates or with substrates that generate chromophores
HRP (horseradish peroxidase)	In the presence of hydrogen oxide, oxidation of 3,3',5,5'-tetramethylbenzidine to form a colored complex
AP (alkaline phosphatase)	Luminescence reaction with suitable substrates or with substrates that generate chromophores

In accordance with the invention, cells that naturally or normally express one or more, all or any combination of the biomarkers of the invention can be used in the methods described herein. Alternatively, cells can be engineered to express any one or more, all or any combination of the biomarkers of the invention, or a reporter gene using techniques well-known in the art and used in the methods described herein. Examples of such techniques include, but are not to, calcium phosphate precipitation (see, *e.g.*, Graham & Van der Eb, 1978, Virol. 52:546), dextran-mediated transfection, calcium phosphate mediated transfection, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the nucleic acid in liposomes, and direct microinjection of the nucleic acid into nuclei.

In a specific embodiment, the cells used in the methods described herein are chondrocytes, lymphocytes (T or B lymphocytes), monocytes, neutrophils, macrophages, eosinophils, basophils, erythrocytes or platelets. In a preferred embodiment, the cells used in the methods described herein are bladder cells. In another preferred embodiment, the cells used in the methods described herein are lymphocytes. In another preferred embodiment, the

cells used in the methods described herein are leukocytes. In another preferred embodiment, the cells used in the methods described herein are all cells in blood. In another embodiment, the cells used in the methods described herein are immortalized cell lines derived from a source, *e.g.*, a tissue.

Any cell-free extract that permits the translation, and optionally but preferably, the transcription, of a nucleic acid can be used in accordance with the methods described herein. The cell-free extract may be isolated from cells of any species origin. For example, the cell-free translation extract may be isolated from human cells, cultured mouse cells, cultured rat cells, Chinese hamster ovary (CHO) cells, Xenopus oocytes, rabbit reticulocytes, wheat germ, or rye embryo (see, *e.g.*, Krieg & Melton, 1984, Nature 308:203 and Dignam et al., 1990 Methods Enzymol. 182:194-203). Alternatively, the cell-free translation extract, *e.g.*, rabbit reticulocyte lysates and wheat germ extract, can be purchased from, *e.g.*, Promega, (Madison, WI). In a preferred embodiment, the cell-free extract is an extract isolated from human cells. In a specific embodiment, the human cells are HeLa cells, lymphocytes, or bladder cells.

In addition to the ability to modulate the expression levels of RNA and/or protein products a biomarker of the invention, it may be desirable, at least in certain instances, that compounds modulate the activity of a protein product of a biomarker of the invention. Thus, the present invention provides methods of identifying compounds to be tested for an ability to prevent, treat, manage or ameliorate bladder cancer or a symptom thereof, comprising methods for identifying compounds that modulate the activity of a protein product of one or more biomarkers of the invention. Such methods can comprise: (a) contacting a cell expressing a protein product of one or more biomarkers of the invention with a test compound; (b) after an incubation period determining the activity level of the protein product; and (c) comparing the activity level to that in a corresponding control cell that has not been contacted with the test compound, so that if the level of activity in (a) is altered relative to the level of activity in the control cell, a compound to be tested for an ability to prevent, treat, manage or ameliorate bladder cancer or a symptom thereof is identified. In a specific embodiment, the activity level(s) is altered by up to 5%, 10%, 15%, 25%, 30%, 40%, 50%, 5 to 25%, 10 to 30%, at least 1 fold, at least 1.5 fold, at least 2 fold, 4

fold, 5 fold, 10 fold, 25 fold, 1 to 10 fold, or 5 to 25 fold relative to the activity level in the control as determined by utilising an assay described herein (*e.g.*, a microarray or RT-PCR) or an assay well known to one of skill in the art. In alternate embodiments, such a method comprises determining the activity level of a protein product of any number of up to at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 12, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, 1 to 5, 1-10, 5-10, 5-25, or 10-40, all or any combination of the biomarkers of the invention present in the cell and comparing the activity levels to those present in the control.

The present invention provides methods of identifying compounds to be tested for an ability to prevent, treat, manage or ameliorate bladder cancer or a symptom thereof, comprising: (a) contacting a cell-free extract with a nucleic acid encoding a protein product of one or more biomarkers of the invention and a test compound; (b) after an incubation period, determining the activity level of the protein product; and (c) comparing the activity level to that in a corresponding control that has not been contacted with the test compound, so that if the level of activity in (a) is altered relative to the level of activity in the control, a compound to be tested for an ability to prevent, treat, manage or ameliorate bladder cancer or a symptom thereof is identified. In a specific embodiment, the activity level(s) is altered by 1% ? 5%, 10%, 15%, 25%, 30%, 40%, 50%, 5 to 25%, 10 to 30%, at least 1 fold, at least 1.5 fold, at least 2 fold, 4 fold, 5 fold, 10 fold, 25 fold, 1 to 10 fold, or 5 to 25 fold relative to the activity level in the control as determined by utilising an assay described herein (*e.g.*, a microarray or qRT-PCR) or an assay well known to one of skill in the art. In alternate embodiments, such a method comprises determining the activity level of a protein product of any number of up to at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 12, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, 1 to 5, 1-10, 5-10, 5-25, or 10-40, all or any combination of the biomarkers of the invention present in the sample and comparing the activity levels to those present in the control.

Standard techniques can be utilised to determine the level of activity of a protein product of a biomarker of the invention.

5.18.2 Biological Activity of the Compounds

Upon identification of compounds to be tested for an ability to prevent, treat, manage or ameliorate bladder cancer or a symptom thereof (for convenience referred to herein as a "lead" compound), the compounds can be further investigated. For example, the compounds identified via the present methods can be further tested *in vivo* in accepted animal models of cancer and preferably bladder cancer. Further, the compounds identified via the methods can be analyzed with respect to their specificity. .

In one embodiment, the effect of a lead compound can be assayed by measuring the cell growth or viability of the target cell. Such assays can be carried out with representative cells of cell types involved in bladder cancer (*e.g.*, bladder lining cells). Alternatively, instead of culturing cells from a patient, a lead compound may be screened using cells of a cell line.

Many assays well-known in the art can be used to assess the survival and/or growth of a patient cell or cell line following exposure to a lead compound; for example, cell proliferation can be assayed by measuring Bromodeoxyuridine (BrdU) incorporation (see, *e.g.*, Hoshino et al., 1986, Int. J. Cancer 38, 369; Campana et al., 1988, J. Immunol. Meth. 107:79) or (^{3}H)-thymidine incorporation (see, *e.g.*, Chen, J., 1996, Oncogene 13:1395-403; Jeoung, J., 1995, J. Biol. Chem. 270:18367-73), by direct cell count, by detecting changes in transcription, translation or activity of known genes such as proto-oncogenes (*e.g.*, *fos, myc*) or cell cycle markers (Rb, cdc2, cyclin A, D1, D2, D3, E, etc). The levels of such protein and RNA (*e.g.*, mRNA) and activity can be determined by any method well known in the art. For example, protein can be quantitated by known immunodiagnostic methods such as Western blotting or immunoprecipitation using commercially available antibodies. mRNA can be quantitated using methods that are well known and routine in the art, for example, using northern analysis, RNase protection, the polymerase chain reaction in connection with the reverse transcription. Cell viability can be assessed by using trypan-blue staining or other cell death or viability markers known in the art. In a specific embodiment, the level of cellular ATP is measured to determined cell viability. Differentiation can be assessed, for example, visually based on changes in morphology.

Animal Models

Compounds can be tested in suitable animal model systems prior to use in humans. Such animal model systems include but are not limited to rats, mice, chicken, cows, monkeys, pigs, dogs, rabbits, etc. Any animal system well-known in the art may be used. In certain embodiments, compounds are tested in a mouse model. Compounds can be administered repeatedly.

Accepted animal models can be utilised to determine the efficacy of the compounds identified via the methods described above for the prevention, treatment, management and/or amelioration of bladder cancer or a symptom thereof.

Toxicity

The toxicity and/or efficacy of a compound identified in accordance with the invention can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, for determining the LD_{50} (the dose lethal to 50% of the population) and the ED_{50} (the dose therapeutically effective in 50% of the population). Cells and cell lines that can be used to assess the cytotoxicity of a compound identified in accordance with the invention include, but are not limited to, peripheral blood mononuclear cells (PBMCs), Caco-2 cells, and Huh7 cells. The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD_{50}/ED_{50}. A compound identified in accordance with the invention that exhibits large therapeutic indices is preferred. While a compound identified in accordance with the invention that exhibits toxic side effects may be used, care should be taken to design a delivery system that targets such agents to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage of a compound identified in accordance with the invention for use in humans. The dosage of such agents lies preferably within a range of circulating concentrations that include the ED_{50} with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilised. For any agent used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose may be formulated in animal models to

achieve a circulating plasma concentration range that includes the IC_{50} (*i.e.*, the concentration of the compound that achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography.

Design of Congeners or Analogs

The compounds which display the desired biological activity can be used as lead compounds for the development or design of congeners or analogs having useful pharmacological activity. For example, once a lead compound is identified, molecular modeling techniques can be used to design variants of the compound that can be more effective. Examples of molecular modeling systems are the CHARM and QUANTA programs (Polygen Corporation, Waltham, MA). CHARM performs the energy minimization and molecular dynamics functions. QUANTA performs the construction, graphic modelling and analysis of molecular structure. QUANTA allows interactive construction, modification, visualization, and analysis of the behavior of molecules with each other.

A number of articles review computer modeling of drugs interactive with specific proteins, such as Rotivinen et al., 1988, Acta Pharmaceutical Fennica 97:159-166; Ripka, 1998, New Scientist 54-57; McKinaly & Rossmann, 1989, Annu. Rev. Pharmacol. Toxiciol. 29:111-122; Perry & Davies, OSAR: Quantitative Structure-Activity Relationships in Drug Design pp. 189-193 (Alan R. Liss, Inc. 1989); Lewis & Dean, 1989, Proc. R. Soc. Lond. 236:125-140 and 141-162; Askew et al., 1989, J. Am. Chem. Soc. 111:1082-1090. Other computer programs that screen and graphically depict chemicals are available from companies such as BioDesign, Inc. (Pasadena, California), Allelix, Inc. (Mississauga, Ontario, Canada), and Hypercube, Inc. (Cambridge, Ontario). Although these are primarily designed for application to drugs specific to particular proteins, they can be adapted to design of drugs specific to any identified region. The analogs and congeners can be tested for binding to the proteins of interest (*i.e.*, the protein products of a biomarker of the invention) using the above-described screens for biologic activity. Alternatively, lead compounds with

little or no biologic activity, as ascertained in the screen, can also be used to design analogs and congeners of the compound that have biologic activity.

5.18.3 Compounds

Compounds that can be tested and identified methods described herein can include, but are not limited to, compounds obtained from any commercial source, including Aldrich (1001 West St. Paul Ave., Milwaukee, WI 53233), Sigma Chemical (P.O. Box 14508, St. Louis, MO 63178), Fluka Chemie AG (Industriestrasse 25, CH-9471 Buchs, Switzerland (Fluka Chemical Corp. 980 South 2nd Street, Ronkonkoma, NY 11779)), Eastman Chemical Company, Fine Chemicals (P.O Box 431, Kingsport, TN 37662), Boehringer Mannheim GmbH (Sandhofer Strasse 116, D-68298 Mannheim), Takasago (4 Volvo Drive, Rockleigh, NJ 07647), SST Corporation (635 Brighton Road, Clifton, NJ 07012), Ferro (111 West Irene Road, Zachary, LA 70791), Riedel-deHaen Aktiengesellschaft (P.O. Box D-30918, Seelze, Germany), PPG Industries Inc., Fine Chemicals (One PPG Place, 34th Floor, Pittsburgh, PA 15272). Further any kind of natural products may be screened using the methods of the invention, including microbial, fungal, plant or animal extracts.

Compounds from large libraries of synthetic or natural compounds can be screened. Numerous means are currently used for random and directed synthesis of saccharide, peptide, and nucleic acid-based compounds. Synthetic compound libraries are commercially available from a number of companies including Maybridge Chemical Co. (Trevillet, Cornwall, UK), Comgenex (Princeton, NJ), Brandon Associates (Merrimack, NH), and Microsource (New Milford, CT). A rare chemical library is available from Aldrich (Milwaukee, WI). Combinatorial libraries are available and are prepared. Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant and animal extracts are available from e.g., Pan Laboratories (Bothell, WA) or MycoSearch (NC), or are readily producible by methods well known in the art. Additionally, natural and synthetically produced libraries and compounds are readily modified through conventional chemical, physical, and biochemical means.

Furthermore, diversity libraries of test compounds, including small molecule test compounds, may be utilised. Libraries screened using the methods of the present invention can comprise a variety of types of compounds. Examples of libraries that can be

screened in accordance with the methods of the invention include, but are not limited to, peptoids; random biooligomers; diversomers such as hydantoins, benzodiazepines and dipeptides; vinylogous polypeptides; nonpeptidal peptidomimetics; oligocarbamates; peptidyl phosphonates; peptide nucleic acid libraries; antibody libraries; carbohydrate libraries; and small molecule libraries (preferably, small organic molecule libraries). In some embodiments, the compounds in the libraries screened are nucleic acid or peptide molecules. In a non-limiting example, peptide molecules can exist in a phage display library. In other embodiments, the types of compounds include, but are not limited to, peptide analogs including peptides comprising non-naturally occurring amino acids, *e.g.*, D-amino acids, phosphorous analogs of amino acids, such as α-amino phosphoric acids and α-amino phosphoric acids, or amino acids having non-peptide linkages, nucleic acid analogs such as phosphorothioates and PNAs, hormones, antigens, synthetic or naturally occurring drugs, opiates, dopamine, serotonin, catecholamines, thrombin, acetylcholine, prostaglandins, organic molecules, pheromones, adenosine, sucrose, glucose, lactose and galactose. Libraries of polypeptides or proteins can also be used in the assays of the invention.

In a specific embodiment, the combinatorial libraries are small organic molecule libraries including, but not limited to, benzodiazepines, isoprenoids, thiazolidinones, metathiazanones, pyrrolidines, morpholino compounds, and benzodiazepines. In another embodiment, the combinatorial libraries comprise peptoids; random bio-oligomers; benzodiazepines; diversomers such as hydantoins, benzodiazepines and dipeptides;, vinylogous polypeptides; nonpeptidal peptidomimetics; oligocarbamates; peptidyl phosphonates; peptide nucleic acid libraries; antibody libraries; or carbohydrate libraries. Combinatorial libraries are themselves commercially available For example, libraries may be commercially obtained from, *e.g.*, Specs and BioSpecs B.V. (Rijswijk, The Netherlands), Chembridge Corporation (San Diego, CA), Contract Service Company (Dolgoprudny, Moscow Region, Russia), Comgenex USA Inc. (Princeton, NJ), Maybridge Chemicals Ltd. (Cornwall PL34 OHW, United Kingdom), Asinex (Moscow, Russia), ComGenex (Princeton, New Jersey), Ru, Tripos, Inc. (St. Louis, Missouri), ChemStar, Ltd (Moscow, Russia), 3D Pharmaceuticals (Exton, Pennsylvania), and Martek Biosciences (Columbia, Maryland).

In a preferred embodiment, the library is preselected so that the compounds of the library are more amenable for cellular uptake. For example, compounds are selected based on specific parameters such as, but not limited to, size, lipophilicity, hydrophilicity, and hydrogen bonding, which enhance the likelihood of compounds getting into the cells. In another embodiment, the compounds are analyzed by three-dimensional or four-dimensional computer computation programs.

The combinatorial compound library for use in accordance with the methods of the present invention may be synthesised. There is a great interest in synthetic methods directed toward the creation of large collections of small organic compounds, or libraries, which could be screened for pharmacological, biological or other activity. The synthetic methods applied to create vast combinatorial libraries are performed in solution or in the phase, *i.e.*, on a support. Solid-phase synthesis makes it easier to conduct multi-step reactions and to drive reactions to completion with high yields because excess reagents can be easily added and washed away after each reaction step. Solid-phase combinatorial synthesis also tends to improve isolation, purification and screening. However, the more traditional solution phase chemistry supports a wider variety of organic reactions than solid-phase chemistry.

Combinatorial compound libraries of the present invention may be synthesised using the apparatus described in U.S. Patent No. 6,190,619 to Kilcoin et al., which is hereby incorporated by reference in its entirety. U.S. Patent No. 6,190,619 discloses a synthesis apparatus capable of holding a plurality of reaction vessels for parallel synthesis of multiple discrete compounds or for combinatorial libraries of compounds.

In one embodiment, the combinatorial compound library can be synthesised in solution. The method disclosed in U.S. Patent No. 6,194,612 to Boger et al., which is hereby incorporated by reference in its entirety, features compounds useful as templates for solution phase synthesis of combinatorial libraries. The template is designed to permit reaction products to be easily purified from unreacted reactants using liquid/liquid or solid/liquid extractions. The compounds produced by combinatorial synthesis using the template will preferably be small organic molecules. Some compounds in the library may mimic the effects of non-peptides or peptides. In contrast to solid phase synthesize of combinatorial

compound libraries, liquid phase synthesis does not require the use of specialized protocols for monitoring the individual steps of a multistep solid phase synthesis (Egner et al., 1995, J.Org. Chem. 60:2652; Anderson et al., 1995, J. Org. Chem. 60:2650; Fitch et al., 1994, J. Org. Chem. 59:7955; Look et al., 1994, J. Org. Chem. 49:7588; Metzger et al., 1993, Angew. Chem., Int. Ed. Engl. 32:894; Youngquist et al., 1994, Rapid Commun. Mass Spect. 8:77; Chu et al., 1995, J. Am. Chem. Soc. 117:5419; Brummel et al., 1994, Science 264:399; and Stevanovic et al., 1993, Bioorg. Med. Chem. Lett. 3:431).

Combinatorial compound libraries useful for the methods of the present invention can be synthesised on solid supports. In one embodiment, a split synthesis method, a protocol of separating and mixing supports during the synthesis, is used to synthesize a library of compounds on solid supports (see *e.g.* , Lam et al., 1997, Chem. Rev. 97:41-448; Ohlmeyer et al., 1993, Proc. Natl. Acad. Sci. USA 90:10922-10926 and references cited therein). Each solid support in the final library has substantially one type of compound attached to its surface. Other methods for synthesizing combinatorial libraries on solid supports, wherein one product is attached to each support, will be known to those of skill in the art (*see, e.g.*, Nefzi et al., 1997, Chem. Rev. 97:449-472).

In some embodiments of the present invention, compounds can be attached to solid supports via linkers. Linkers can be integral and part of the solid support, or they may be nonintegral that are either synthesised on the solid support or attached thereto after synthesis. Linkers are useful not only for providing points of compound attachment to the solid support, but also for allowing different groups of molecules to be cleaved from the solid support under different conditions, depending on the nature of the linker. For example, linkers can be, *inter alia*, electrophilically cleaved, nucleophilically cleaved, photocleavable, enzymatically cleaved, cleaved by metals, cleaved under reductive conditions or cleaved under oxidative conditions. In a preferred embodiment, the compounds are cleaved from the solid support prior to high throughput screening of the compounds.

If the library comprises arrays or microarrays of compounds, wherein each compound has an address or identifier, the compound can be deconvoluted, *e.g.*, by cross-referencing the positive sample to original compound list that was applied to the individual test assays.

If the library is a peptide or nucleic acid library, the sequence of the compound can be determined by direct sequencing of the peptide or nucleic acid. Such methods are well known to one of skill in the art.

A number of physico-chemical techniques can be used for the *de novo* characterization of compounds. Examples of such techniques include, but are not limited to, mass spectrometry, NMR spectroscopy, X-ray crytallography and vibrational spectroscopy.

5.19 Use of Identified Compounds to Prevent, Treat, Manage or Ameliorate Bladder Cancer or Symptoms Thereof

The present invention provides methods of preventing, treating, managing or ameliorating bladder cancer or a symptom thereof, said methods comprising administering to a subject in need thereof one or more compounds identified in accordance with the methods of the invention. In a preferred embodiment, the subject is human.

In one embodiment, the invention provides a method of preventing, treating, managing or ameliorating bladder cancer or a symptom thereof, said method comprising administering to a subject in need thereof a dose of a prophylactically or therapeutically effective amount of one or more compounds identified in accordance with the methods of the invention. In a specific embodiment, a compound identified in accordance with the methods of the invention is not administered to prevent, treat, or ameliorate bladder cancer or a symptom thereof, if such compound has been used previously to prevent, treat, manage or ameliorate bladder cancer or a symptom thereof. In another embodiment, a compound identified in accordance with the methods of the invention is not administered to prevent, treat, or ameliorate bladder cancer or a symptom thereof, if such compound has suggested to be used to prevent, treat, manage or ameliorate bladder cancer or a symptom thereof. In another embodiment, a compound identified in accordance with the methods of the invention specifically binds to and/or alters the expression and/or activity level of a protein or RNA product of only one biomarker of the invention. In yet another embodiment, a compound identified in accordance with the methods of the invention binds to and/or alters the expression and/or activity level of a protein or RNA product of any number of up to at least 2, at least 3, at least 4, at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40 or more biomarkers of the invention.

In a specific embodiment, a compound identified in accordance with the methods of the invention increases or decreases the anabolic and/or the catabolic activity of a bladder cell or immortalized bladder cancer cell. Preferably, such a compound increases or decreases the anabolic and/or catabolic activity of a chondrocyte by greater than 1.0-fold, more preferably, 1.5-5-fold, and most preferably, 5-100-fold, as compared to an untreated bladder cell or immortalized bladder cell. In another embodiment, a compound identified in accordance with the methods of the invention ameliorates at least one of the symptoms and/or changes associated with bladder cancer including symptoms of bladder cancer. In a particular embodiment, the prophylactic or therapeutic agent administered to prevent, treat, manage or ameliorate bladder cancer or a symptom thereof is a synthetic compound or a natural product (*e.g.* a plant extract or culture supernatant), or a mixture of compounds.

The invention also provides methods of preventing, treating, managing or ameliorating ameliorate bladder cancer or a symptom thereof, said methods comprising administering to a subject in need thereof one or more of the compounds identified utilising the screening methods described herein, and one or more other therapies (*e.g.*, prophylactic or therapeutic agents and surgery). In a specific embodiment, such therapies are currently being used, have been used or are known to be useful in the prevention, treatment, management or amelioration of o ameliorate bladder cancer or a symptom thereof (including, but not limited to the prophylactic or therapeutic agents listed in sections herein below). The therapies (*e.g.*, prophylactic or therapeutic agents) of the combination therapies of the invention can be administered sequentially or concurrently. In a specific embodiment, the combination therapies of the invention comprise a compound identified in accordance with the invention and at least one other therapy that has the same mechanism of action as said compound. In another specific embodiment, the combination therapies of the invention comprise a compound identified in accordance with the methods of the invention and at least one other therapy (*e.g.*, prophylactic or therapeutic agent) which has a different mechanism of action than said compound. The combination therapies of the present invention improve the prophylactic or therapeutic effect of a compound of the invention by functioning together with the compound to have an additive or synergistic effect. The combination therapies of the present invention reduce the side effects associated with the therapies (*e.g.*, prophylactic or therapeutic agents).

The prophylactic or therapeutic agents of the combination therapies can be administered to a subject in the same pharmaceutical composition. Alternatively, the prophylactic or therapeutic agents of the combination therapies can be administered concurrently to a subject in separate pharmaceutical compositions. The prophylactic or therapeutic agents may be administered to a subject by the same or different routes of administration.

In specific embodiment, a pharmaceutical composition comprising one or more compounds identified in an assay described herein is administered to a subject, preferably a human, to prevent, treat, manage or ameliorate bladder cancer or a symptom thereof. In accordance with the invention, the pharmaceutical composition may also comprise one or more prophylactic or therapeutic agents. Preferably, such agents are currently being used, have been used or are known to be useful in the prevention, treatment, management or amelioration of bladder cancer or a symptom thereof.

A compound identified in accordance with the methods of the invention may be used as a first, second, third, fourth or fifth line of therapy for bladder cancer. The invention provides methods for treating, managing or ameliorating bladder cancer or a symptom thereof in a subject refractory to conventional therapies for bladder cancer, said methods comprising administering to said subject a dose of a prophylactically or therapeutically effective amount of a compound identified in accordance with the methods of the invention.

The invention provides methods for treating, managing or ameliorating bladder cancer or a symptom thereof in a subject refractory to existing single agent therapies for bladder cancer, said methods comprising administering to said subject a dose of a prophylactically or therapeutically effective amount of a compound identified in accordance with the methods of the invention and a dose of a prophylactically or therapeutically effective amount of one or more other therapies (*e.g.*, prophylactic or therapeutic agents). The invention also provides methods for treating or managing a bladder cancer by administering a compound identified in accordance with the methods of the invention in combination with any other therapy (*e.g.*, surgery) to patients who have proven refractory to other therapies but are no longer on these therapies. The invention also provides methods for the treatment or

management of a patient having bladder cancer and immunosuppressed by reason of having previously undergone other therapies. The invention also provides alternative methods for the treatment or management of bladder cancer where hormonal therapy and/or biological therapy/immunotherapy has proven or may prove too toxic, *i.e.*, results in unacceptable or unbearable side effects, for the subject being treated or managed.

5.19.1 Compounds for Use in Preventing, Treating, Managing or Ameliorating Bladder Cancer or a Symptom Thereof

Representative, non-limiting examples of compounds that can used in accordance with the methods of the invention to prevent, treat, manage and/or ameliorate bladder cancer or a symptom thereof are described in detail below.

First, such compounds can include, for example, antisense, ribozyme, or triple helix compounds that can downregulate the expression or activity of a protein or RNA product of a biomarker of the invention. Such compounds are described in detail in the subsection below.

Second, such compounds can include, for example, antibody compositions that can modulate the expression of a protein or RNA product of a biomarker of the invention, or the activity of a protein product of a biomarker of the invention. In a specific embodiment, the antibody compositions downregulate the expression a protein or RNA product of a biomarker of the invention, or the activity of a protein product of a biomarker of the invention. Such compounds are described in detail in the subsection below.

Third, such compounds can include, for example, protein products of a biomarker of the invention. The invention encompasses the use of peptides or peptide mimetics selected to mimic a protein product of a biomarker of the invention to prevent, treat, manage or ameliorate bladder cancer or a symptom thereof. Further, such compounds can include, for example, dominant-negative polypeptides that can modulate the expression a protein or RNA product of a biomarker of the invention, or the activity of a protein product of a biomarker of the invention.

The methods also encompass the use derivatives, analogs and fragments of a protein product of a biomarker of the invention to prevent, treat, manage or ameliorate

bladder cancer or a symptom thereof. In particular, the invention encompasses the use of fragments of a protein product of a biomarker of the invention comprising one or more domains of such a protein(s) to prevent, treat, manage or ameliorate bladder cancer or a symptom thereof. In another specific embodiment, the invention encompasses the use of a protein product of a biomarker of the invention, or an analog, derivative or fragment of such a protein which is expressed as a fusion, or chimeric protein product (comprising the protein, fragment, analog, or derivative joined via a peptide bond to a heterologous protein sequence).

In specific embodiments, an antisense oligonucleotide of at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, or more of biomarkers of the invention are administered to prevent, treat, manage or ameliorate bladder cancer or a symptom thereof. In other embodiments, one or more of protein products of a biomarker of the invention or a fragment, analog, or derivative thereof are administered to prevent, treat, manage or ameliorate bladder cancer or a symptom thereof. In other embodiment, one or more antibodies that specifically bind to a protein product of the invention are administered to prevent, treat, manage or ameliorate bladder cancer or a symptom thereof. In other embodiments, one or more dominant-negative polypeptides are administered to prevent, treat, manage or ameliorate bladder cancer or a symptom thereof.

Antisense, Ribozyme, Triple-Helix Compositions

Standard techniques can be utilised to produce antisense, triple helix, or ribozyme molecules for use as part of the methods described herein. First, standard techniques can be utilised for the production of antisense nucleic acid molecules, *i.e.*, molecules which are complementary to a sense nucleic acid encoding a polypeptide of interest, *e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. Accordingly, an antisense nucleic acid can hydrogen bond to a sense nucleic acid. The antisense nucleic acid can be complementary to an entire coding strand, or to only a portion thereof, *e.g.*, all or part of the protein coding region (or open reading frame). An antisense nucleic acid molecule can be antisense to all or part of a non-coding region of the coding strand of a nucleotide sequence encoding a polypeptide of

interest. The non-coding regions ("5' and 3' untranslated regions") are the 5' and 3' sequences that flank the coding region and are not translated into amino acids.

An antisense oligonucleotide can be, for example, any number of up to about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides or more in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesised using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest).

Antisense nucleic acid molecules administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA encoding the polypeptide of interest to thereby inhibit expression, *e.g.*, by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or,

for example, in the case of an antisense nucleic acid molecule which binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue, *e.g.*, a joint (*e.g.*, a knee, hip, elbow, and knuckle), site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell, *e.g.*, a T cell or chondrocyte, surface, *e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies which bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using vectors, *e.g.*, gene therapy vectors, described below. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

An antisense nucleic acid molecule of interest can be an α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual α-units, the strands run parallel to each other (Gaultier et al., 1987, Nucleic Acids Res. 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue et al., 1987, Nucleic Acids Res. 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue et al., 1987, FEBS Lett. 215:327-330).

Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region, and can also be generated using standard techniques. Thus, ribozymes (*e.g.*, hammerhead ribozymes (described in Haselhoff and Gerlach, 1988, Nature 334:585-591)) can be used to catalytically cleave mRNA transcripts to thereby inhibit translation of the protein encoded by the mRNA. A ribozyme having specificity for a nucleic acid molecule encoding a polypeptide of interest can be designed based upon the nucleotide sequence of a cDNA disclosed herein. For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a Cech et al. U.S. Patent No. 4,987,071; and Cech

et al. U.S. Patent No. 5,116,742. Alternatively, an mRNA encoding a polypeptide of interest can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. *See, e.g.*, Bartel and Szostak, 1993, Science 261:1411-1418.

Triple helical structures can also be generated using well known techniques. For example, expression of a polypeptide of interest can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the gene encoding the polypeptide (*e.g.*, the promoter and/or enhancer) to form triple helical structures that prevent transcription of the gene in target cells. *See generally* Helene, 1991, Anticancer Drug Des. 6(6):569-84; Helene, 1992, Ann. N.Y. Acad. Sci. 660:27-36; and Maher, 1992, Bioassays 14(12):807-15.

In various embodiments, nucleic acid compositions can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (*see* Hyrup et al., 1996, Bioorganic & Medicinal Chemistry 4(1): 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, *e.g.*, DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup et al.,1996, *supra*; Perry-O'Keefe et al., 1996, Proc. Natl. Acad. Sci. USA 93: 14670-675.

PNAs can, for example, be modified, *e.g.*, to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras can be generated which may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, *e.g.*, RNAse H and DNA polymerases, to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup, 1996, *supra*). The synthesis of PNA-DNA chimeras

can be performed as described in Hyrup, 1996, *supra,* and Finn et al., 1996, Nucleic Acids Res. 24(17):3357-63. For example, a DNA chain can be synthesised on a solid support using standard phosphoramidite coupling chemistry and modified nucleoside analogs. Compounds such as 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite can be used as a link between the PNA and the 5' end of DNA (Mag et al., 1989, Nucleic Acids Res. 17:5973-88). PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn et al., 1996, Nucleic Acids Res. 24(17):3357-63). Alternatively, chimeric molecules can be synthesised with a 5' DNA segment and a 3' PNA segment (Peterser et al., 1975, Bioorganic Med. Chem. Lett. 5:1119-11124).

In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (*see, e.g.*, Letsinger et al., 1989, Proc. Natl. Acad. Sci. USA 86:6553-6556; Lemaitre et al., 1987, Proc. Natl. Acad. Sci. USA 84:648-652; International Publication No. W0 88/09810) or the blood-brain barrier (*see, e.g.*, International Publication No. W0 89/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (*see, e.g.*, Krol et al., 1988, Bio/Techniques 6:958-976) or intercalating agents (*see, e.g.*, Zon, 1988, Pharm. Res. 5:539-549). To this end, the oligonucleotide may be conjugated to another molecule, *e.g.*, a peptide, hybridization triggered cross-linking agent, transport agent, hybridization-triggered cleavage agent, etc.

Antibody Compositions

In one embodiment, antibodies that specifically bind to one or more protein products of one or more biomarkers of the invention are administered to a subject, preferably a human, to prevent, treat, manage or ameliorate bladder cancer or a symptom thereof. In another embodiment, any combination of antibodies that specifically bind to one or more protein products of one or more biomarkers of the invention are administered to a subject, preferably a human, to prevent, treat, manage or ameliorate bladder cancer or a symptom thereof. In a specific embodiment, one or more antibodies that specifically bind to one or more protein products of one or more biomarkers of the invention are administered to a subject, preferably a human, in combination with other types of therapies to prevent, treat,

manage or ameliorate bladder cancer or a symptom thereof. In certain embodiments, antibodies known in the art that specifically bind to one or more protein products of one or more biomarkers of the invention are administered to a subject, preferably a human, alone or in combination with other types of therapies to prevent, treat, manage or ameliorate bladder cancer or a symptom thereof. In other embodiments, antibodies known in the art that specifically bind to one or more protein products of one or more biomarkers of the invention are not administered to a subject, preferably a human, alone or in combination with other types of therapies to prevent, treat, manage or ameliorate bladder cancer or a symptom thereof.

One or more antibodies that specifically bind to one or more protein products of one or more biomarkers of the invention can be administered to a subject, preferably a human, using various delivery systems are known to those of skill in the art. For example, such antibodies can be administered by encapsulation in liposomes, microparticles or microcapsules. See, *e.g.*, U.S. Patent No. 5,762,904, U.S. Patent No. 6,004,534, and International Publication No. WO 99/52563. In addition, such antibodies can be administered using recombinant cells capable of expressing the antibodies, or retroviral, other viral vectors or non-viral vectors capable of expressing the antibodies.

Antibodies that specifically bind one or more protein products of one or more biomarkers of the invention can be obtained from any known source. Alternatively, antibodies that specifically bind to one or more protein products of one or more biomarkers of the invention can be produced by any method known in the art for the synthesis of antibodies, in particular, by chemical synthesis or preferably, by recombinant expression techniques.

Antibodies include, but are not limited to, polyclonal antibodies, monoclonal antibodies, bispecific antibodies, multispecific antibodies, human antibodies, humanized antibodies, camelised antibodies, chimeric antibodies, single-chain Fvs (scFv) (see *e.g.*, Bird *et al.* (1988) *Science* 242:423-426; and Huston *et al.* (1988) *Proc. Natl. Acad. Sci. USA* 85:5879-5883), single chain antibodies, single domain antibodies, Fab fragments, F(ab') fragments, disulfide-linked Fvs (sdFv), and anti-idiotypic (anti-Id) antibodies (including, *e.g.*, anti-Id antibodies to antibodies of the invention), and antigen binding and/or epitope-binding

fragments of any of the above. The term "antibody", as used herein, refers to immunoglobulin molecules and immunologically active fragments of immunoglobulin molecules, *i.e.*, molecules that contain an antigen binding site. Immunoglobulin molecules can be of any type (*e.g.*, IgG, IgE, IgM, IgD, IgA and IgY), class (*e.g.*, IgG_1, IgG_2, IgG_3, IgG_4, IgA_1 and IgA_2) or subclass. Examples of immunologically active fragments of immunoglobulin molecules include F(ab) fragments (a monovalent fragment consisting of the VL, VH, CL and CH1 domains) and F(ab')2 fragments (a bivalent fragment comprising two Fab fragments linked by a disulfide bridge at the hinge region) which can be generated by treating the antibody with an enzyme such as pepsin or papain. Immunologically active fragments also include, but are not limited to, Fd fragments (consisting of the VH and CH1 domains), Fv fragments (consisting of the VL and VH domains of a single arm of an antibody), dAb fragments (consisting of a VH domain; Ward *et al.*, (1989) *Nature* 341:544-546), and isolated complementarity determining regions (CDRs). Antibodies that specifically bind to an antigen can be produced by any method known in the art for the synthesis of antibodies, in particular, by chemical synthesis or preferably, by recombinant expression techniques.

Polyclonal antibodies that specifically bind to an antigen can be produced by various procedures well-known in the art. For example, a human antigen can be administered to various host animals including, but not limited to, rabbits, mice, rats, etc. to induce the production of sera containing polyclonal antibodies specific for the human antigen. Various adjuvants may be used to increase the immunological response, depending on the host species, and include but are not limited to, Freund's (complete and incomplete), mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanins, dinitrophenol, and potentially useful human adjuvants such as BCG (bacille Calmette-Guerin) and corynebacterium parvum. Such adjuvants are also well known in the art.

The term "monospecific antibody" refers to an antibody that displays a single binding specificity and affinity for a particular target, *e.g.*, epitope. This term includes monoclonal antibodies. Monoclonal antibodies can be prepared using a wide variety of techniques known in the art including the use of hybridoma, recombinant, and phage display

technologies, or a combination thereof. See, *e.g.*, U.S. Pat. Nos. RE 32,011, 4,902,614, 4,543,439, 4,411,993 and 4,196,265; Kennett et al (eds.), *Monoclonal Antibodies, Hybridomas: A New Dimension in Biological Analyses*, Plenum Press (1980); and Harlow and Lane (eds.), *Antibodies. A Laboratory Manual*, Cold Spring Harbor Laboratory Press (1988), which are incorporated herein by reference. For example, monoclonal antibodies can be produced using hybridoma techniques including those known in the art and taught, for example, in Harlow et al., *Antibodies: A Laboratory Manual*, (Cold Spring Harbor Laboratory Press, 2nd ed. 1988); Hammerling, et al., in: *Monoclonal Antibodies and T-Cell Hybridomas* 563-681 (Elsevier, N.Y., 1981) (said references incorporated by reference in their entireties). Other techniques that enable the production of antibodies through recombinant techniques (*e.g.*, techniques described by William D. Huse et al., 1989, Science, 246: 1275-1281; L. Sastry et al., 1989, Proc. Natl. Acad. Sci. USA, 86: 5728-5732; and Michelle Alting-Mees et al., Strategies in Molecular Biology, 3: 1-9 (1990) involving a commercial system available from Stratacyte, La Jolla, Calif.) may also be utilised to construct monoclonal antibodies. The term "monoclonal antibody" as used herein is not limited to antibodies produced through hybridoma technology. The term "monoclonal antibody" refers to an antibody that is derived from a single clone, including any eukaryotic, prokaryotic, or phage clone, and not the method by which it is produced.

Methods for producing and screening for specific antibodies using hybridoma technology are routine and well known in the art. Briefly, mice can be immunized with a protein product of a biomarker of the invention, and once an immune response is detected, *e.g.*, antibodies specific for the protein are detected in the mouse serum, the mouse spleen is harvested and splenocytes isolated. The splenocytes are then fused by well known techniques to any suitable myeloma cells, for example cells from cell line SP20 available from the ATCC. Hybridomas are selected and cloned by limited dilution. Additionally, a RIMMS (repetitive immunization multiple sites) technique can be used to immunize an animal (Kilptrack et al., 1997, Hybridoma 16:381-9, incorporated by reference in its entirety). The hybridoma clones are then assayed by methods known in the art for cells that secrete antibodies capable of binding a polypeptide of the invention. Ascites fluid, which generally contains high levels of antibodies, can be generated by immunizing mice with positive hybridoma clones.

Accordingly, the present invention provides methods of generating antibodies by culturing a hybridoma cell secreting an antibody of the invention wherein, preferably, the hybridoma is generated by fusing splenocytes isolated from a mouse immunized with a protein product of a biomarker of the invention, with myeloma cells and then screening the hybridomas resulting from the fusion for hybridoma clones that secrete an antibody able to bind to the protein or protein fragment.

Antibody fragments which recognise specific epitopes of a protein product of a biomarker of the invention may be generated by any technique known to those of skill in the art. For example, Fab and F(ab')2 fragments of the invention may be produced by proteolytic cleavage of immunoglobulin molecules, using enzymes such as papain (to produce Fab fragments) or pepsin (to produce F(ab')2 fragments). F(ab')2 fragments contain the variable region, the light chain constant region and the CH1 domain of the heavy chain. Further, the antibodies of the present invention can also be generated using various phage display methods known in the art.

In phage display methods, functional antibody domains are displayed on the surface of phage particles which carry the polynucleotide sequences encoding them. In particular, DNA sequences encoding VH and VL domains are amplified from animal cDNA libraries (*e.g.*, human or murine cDNA libraries of affected tissues). The DNA encoding the VH and VL domains are recombined together with an scFv linker by PCR and cloned into a phagemid vector. The vector is electroporated in *E. coli* and the *E. coli* is infected with helper phage. Phage used in these methods are typically filamentous phage including fd and M13 and the VH and VL domains are usually recombinantly fused to either the phage gene III or gene VIII. Phage expressing an antigen binding domain that binds to a particular antigen can be selected or identified with antigen, *e.g.*, using labelled antigen or antigen bound or captured to a solid surface or bead. Examples of phage display methods that can be used to make the antibodies of the present invention include those disclosed in Brinkman et al., 1995, J. Immunol. Methods 182:41-50; Ames et al., 1995, J. Immunol. Methods 184:177-186; Kettleborough et al. , 1994, Eur. J. Immunol. 24:952-958; Persic et al., 1997, Gene 187:9-18; Burton et al., 1994, Advances in Immunology 57:191-280; PCT Application No. PCT/GB91/O1 134; International Publication Nos. WO 90/02809, WO 91/10737, WO

92/01047, WO 92/18619, WO 93/1 1236, WO 95/15982, WO 95/20401, and WO97/13844; and U.S. Patent Nos. 5,698,426, 5,223,409, 5,403,484, 5,580,717, 5,427,908, 5,750,753, 5,821,047, 5,571,698, 5,427,908, 5,516,637, 5,780,225, 5,658,727, 5,733,743 and 5,969,108; each of which is incorporated herein by reference in its entirety.

As described in the above references, after phage selection, the antibody coding regions from the phage can be isolated and used to generate whole antibodies, including human antibodies, or any other desired antigen binding fragment, and expressed in any desired host, including mammalian cells, insect cells, plant cells, yeast, and bacteria, *e.g.*, as described below. Techniques to recombinantly produce Fab, Fab' and F(ab')2 fragments can also be employed using methods known in the art such as those disclosed in International Publication No. WO 92/22324; Mullinax et al., 1992, BioTechniques 12(6):864-869; Sawai et al., 1995, AJRI 34:26-34; and Better et al., 1988, Science 240:1041-1043 (said references incorporated by reference in their entireties).

To generate whole antibodies, PCR primers including VH or VL nucleotide sequences, a restriction site, and a flanking sequence to protect the restriction site can be used to amplify the VH or VL sequences in scFv clones. Utilising cloning techniques known to those of skill in the art, the PCR amplified VH domains can be cloned into vectors expressing a VH constant region, *e.g.*, the human gamma 4 constant region, and the PCR amplified VL domains can be cloned into vectors expressing a VL constant region, *e.g.*, human kappa or lambda constant regions. Preferably, the vectors for expressing the VH or VL domains comprise an EF-1α promoter, a secretion signal, a cloning site for the variable domain, constant domains, and a selection marker such as neomycin. The VH and VL domains may also cloned into one vector expressing the necessary constant regions. The heavy chain conversion vectors and light chain conversion vectors are then co-transfected into cell lines to generate stable or transient cell lines that express full-length antibodies, *e.g.*, IgG, using techniques known to those of skill in the art.

For some uses, including *in vivo* use of antibodies in humans and *in vitro* detection assays, it may be preferable to use human or chimeric antibodies. Completely human antibodies are particularly desirable for therapeutic treatment of human subjects. Human antibodies can be made by a variety of methods known in the art including phage

display methods described above using antibody libraries derived from human immunoglobulin sequences. See also U.S. Patent Nos. 4,444,887 and 4,716,111; and International Publication Nos. WO 98/46645, WO 98/50433, WO 98/24893, WO98/16654, WO 96/34096, WO 96/33735, and WO 91/10741; each of which is incorporated herein by reference in its entirety.

Antibodies can also be produced by a transgenic animal. In particular, human antibodies can be produced using transgenic mice which are incapable of expressing functional endogenous immunoglobulins, but which can express human immunoglobulin genes. For example, the human heavy and light chain immunoglobulin gene complexes may be introduced randomly or by homologous recombination into mouse embryonic stem cells. Alternatively, the human variable region, constant region, and diversity region may be introduced into mouse embryonic stem cells in addition to the human heavy and light chain genes. The mouse heavy and light chain immunoglobulin genes may be rendered non-functional separately or simultaneously with the introduction of human immunoglobulin loci by homologous recombination. In particular, homozygous deletion of the J_H region prevents endogenous antibody production. The modified embryonic stem cells are expanded and microinjected into blastocysts to produce chimeric mice. The chimeric mice are then be bred to produce homozygous offspring which express human antibodies. The transgenic mice are immunized in the normal fashion with a selected antigen, *e.g.*, all or a portion of a polypeptide of the invention. Monoclonal antibodies directed against the antigen can be obtained from the immunized, transgenic mice using conventional hybridoma technology. The human immunoglobulin transgenes harbored by the transgenic mice rearrange during B cell differentiation, and subsequently undergo class switching and somatic mutation. Thus, using such a technique, it is possible to produce therapeutically useful IgG, IgA, IgM and IgE antibodies. For an overview of this technology for producing human antibodies, see Lonberg and Huszar (1995, Int. Rev. Immunol. 13:65-93). For a detailed discussion of this technology for producing human antibodies and human monoclonal antibodies and protocols for producing such antibodies, *see, e.g.*, International Publication Nos. WO 98/24893, WO 96/34096, and WO 96/33735; and U.S. Patent Nos. 5,413,923, 5,625,126, 5,633,425, 5,569,825, 5,661,016, 5,545,806, 5,814,318, and 5,939,598, which are incorporated by reference herein in their entirety. In addition, companies such as Abgenix, Inc. (Freemont,

CA) and Genpharm (San Jose, CA) can be engaged to provide human antibodies directed against a selected antigen using technology similar to that described above.

U.S. Patent No. 5,849,992, for example, describes a method of expressing an antibody in the mammary gland of a transgenic mammal. A transgene is constructed that includes a milk-specific promoter and nucleic acids encoding the antibody of interest and a signal sequence for secretion. The milk produced by females of such transgenic mammals includes, secreted-therein, the antibody of interest. The antibody can be purified from the milk, or for some applications, used directly.

A chimeric antibody is a molecule in which different portions of the antibody are derived from different immunoglobulin molecules. Methods for producing chimeric antibodies are known in the art. See *e.g.*, Morrison, 1985, Science 229:1202; Oi et al., 1986, BioTechniques 4:214; Gillies et al., 1989, J. Immunol. Methods 125:191-202; and U.S. Patent Nos. 5,807,715, 4,816,567, 4,816,397, and 6,331,415, which are incorporated herein by reference in their entirety.

A humanized antibody is an antibody or its variant or fragment thereof which is capable of binding to a predetermined antigen and which comprises a framework region having substantially the amino acid sequence of a human immunoglobulin and a CDR having substantially the amino acid sequence of a non-human immuoglobulin. A humanized antibody comprises substantially all of at least one, and typically two, variable domains (Fab, Fab', $F(ab')_2$ Fabc, Fv) in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin (*i.e.*, donor antibody) and all or substantially all of the framework regions are those of a human immunoglobulin consensus sequence. Preferably, a humanized antibody also comprises at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin. Ordinarily, the antibody will contain both the light chain as well as at least the variable domain of a heavy chain. The antibody also may include the CH1, hinge, CH2, CH3, and CH4 regions of the heavy chain. The humanized antibody can be selected from any class of immunoglobulins, including IgM, IgG, IgD, IgA and IgE, and any isotype, including IgG_1, IgG_2, IgG_3 and IgG_4. Usually the constant domain is a complement fixing constant domain where it is desired that the humanized antibody exhibit cytotoxic activity, and the class is typically IgG_1. Where such

cytotoxic activity is not desirable, the constant domain may be of the IgG_2 class. The humanized antibody may comprise sequences from more than one class or isotype, and selecting particular constant domains to optimize desired effector functions is within the ordinary skill in the art. The framework and CDR regions of a humanized antibody need not correspond precisely to the parental sequences, *e.g.*, the donor CDR or the consensus framework may be mutagenized by substitution, insertion or deletion of at least one residue so that the CDR or framework residue at that site does not correspond to either the consensus or the import antibody. Such mutations, however, will not be extensive. Usually, at least 75% of the humanized antibody residues will correspond to those of the parental FR and CDR sequences, more often 90%, and most preferably greater than 95%. Humanized antibody can be produced using variety of techniques known in the art, including but not limited to, CDR-grafting (European Patent No. EP 239,400; International Publication No. WO 91/09967; and U.S. Patent Nos. 5,225,539, 5,530,101, and 5,585,089), veneering or resurfacing (European Patent Nos. EP 592,106 and EP 519,596; Padlan, 1991, Molecular Immunology 28(4/5):489-498; Studnicka et al., 1994, Protein Engineering 7(6):805-814; and Roguska et al. , 1994, PNAS 91:969-973), chain shuffling (U.S. Patent No. 5,565,332), and techniques disclosed in, *e.g.* ., U.S. Pat. No. 6,407,213, U.S. Pat. No. 5,766,886, WO 9317105, Tan et al., 2002, J. Immunol. 169:1119-25, Caldas et al., 2000, Protein Eng. 13(5):353 - 60, Morea et al., 2000, Methods 20(3):267-79, Baca et al., 1997, J. Biol. Chem. 272(16):10678-84, Roguska et al., 1996, Protein Eng. 9(10):895-904, Couto et al., 1995, Cancer Res. 55 (23 Supp):5973s - 5977s, Couto et al., 1995, Cancer Res. 55(8):1717-22, Sandhu JS, 1994, Gene 150(2):409-10, and Pedersen et al., 1994, J. Mol. Biol. 235(3):959-73. Often, framework residues in the framework regions will be substituted with the corresponding residue from the CDR donor antibody to alter, preferably improve, antigen binding. These framework substitutions are identified by methods well known in the art, *e.g.*, by modeling of the interactions of the CDR and framework residues to identify framework residues important for antigen binding and sequence comparison to identify unusual framework residues at particular positions. (See, *e.g.*, Queen et al., U.S. Patent No. 5,585,089; and Riechmann et al., 1988, Nature 332:323, which are incorporated herein by reference in their entireties.)

Single domain antibodies, for example, antibodies lacking the light chains, can be produced by methods well-known in the art. See Riechmann et al., 1999, J. Immuno. 231:25-38; Nuttall et al., 2000, Curr. Pharm. Biotechnol. 1(3):253-263; Muylderman, 2001, J. Biotechnol. 74(4):277302; U.S. Patent No. 6,005,079; and International Publication Nos. WO 94/04678, WO 94/25591, and WO 01/44301, each of which is incorporated herein by reference in its entirety.

Further, the antibodies that specifically bind to an antigen can, in turn, be utilised to generate anti-idiotype antibodies that "mimic" an antigen using techniques well known to those skilled in the art. (See, *e.g.*, Greenspan & Bona, 1989, FASEB J. 7(5):437-444; and Nissinoff, 1991, J. Immunol. 147(8):2429-2438). Such antibodies can be used, alone or in combination with other therapies, in the prevention, treatment, management or amelioration of bladder cancer or a symptom thereof.

The invention encompasses polynucleotides comprising a nucleotide sequence encoding an antibody or fragment thereof that specifically binds to an antigen. The invention also encompasses polynucleotides that hybridize under high stringency, intermediate or lower stringency hybridization conditions to polynucleotides that encode an antibody of the invention.

The polynucleotides may be obtained, and the nucleotide sequence of the polynucleotides determined, by any method known in the art. The nucleotide sequences encoding known antibodies can be determined using methods well known in the art, *i.e.*, nucleotide codons known to encode particular amino acids are assembled in such a way to generate a nucleic acid that encodes the antibody. Such a polynucleotide encoding the antibody may be assembled from chemically synthesised oligonucleotides (*e.g.*, as described in Kutmeier et al., 1994, BioTechniques 17:242), which, briefly, involves the synthesis of overlapping oligonucleotides containing portions of the sequence encoding the antibody, fragments, or variants thereof, annealing and ligating of those oligonucleotides, and then amplification of the ligated oligonucleotides by PCR.

Alternatively, a polynucleotide encoding an antibody may be generated from nucleic acid from a suitable source. If a clone containing a nucleic acid encoding a particular antibody is not available, but the sequence of the antibody molecule is known, a nucleic acid

encoding the immunoglobulin may be chemically synthesised or obtained from a suitable source (*e.g.*, an antibody cDNA library or a cDNA library generated from, or nucleic acid, preferably poly A+ RNA, isolated from, any tissue or cells expressing the antibody, such as hybridoma cells selected to express an antibody of the invention) by PCR amplification using synthetic primers hybridizable to the 3' and 5' ends of the sequence or by cloning using an oligonucleotide probe specific for the particular gene sequence to identify, *e.g.*, a cDNA clone from a cDNA library that encodes the antibody. Amplified nucleic acids generated by PCR may then be cloned into replicable cloning vectors using any method well known in the art.

Once the nucleotide sequence of the antibody is determined, the nucleotide sequence of the antibody may be manipulated using methods well known in the art for the manipulation of nucleotide sequences, *e.g.*, recombinant DNA techniques, site directed mutagenesis, PCR, etc. (see, for example, the techniques described in Sambrook et al., 1990, Molecular Cloning, A Laboratory Manual, 2d Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY and Ausubel et al., eds., 1998, Current Protocols in Molecular Biology, John Wiley & Sons, NY, which are both incorporated by reference herein in their entireties), to generate antibodies having a different amino acid sequence, for example to create amino acid substitutions, deletions, and/or insertions.

Once a polynucleotide encoding an antibody molecule, heavy or light chain of an antibody, or fragment thereof (preferably, but not necessarily, containing the heavy or light chain variable domain) of the invention has been obtained, the vector for the production of the antibody molecule may be produced by recombinant DNA technology using techniques well-known in the art.

In one preferred embodiment, monoclonal antibodies are produced in mammalian cells. Preferred mammalian host cells for expressing the clone antibodies or antigen-binding fragments thereof include Chinese Hamster Ovary (CHO cells) (including dhfr- CHO cells, described in Urlaub and Chasin (1980, Proc. Natl. Acad. Sci. USA 77:4216-4220), used with a DHFR selectable marker, *e.g.*, as described in Kaufman and Sharp (1982, Mol. Biol. 159:601-621), lymphocytic cell lines, *e.g.*, NS0 myeloma cells and SP2 cells, COS

cells, and a cell from a transgenic animal, *e.g.*, a transgenic mammal. For example, the cell is a mammary epithelial cell.

In addition to the nucleic acid sequence encoding the diversified immunoglobulin domain, the recombinant expression vectors may carry additional sequences, such as sequences that regulate replication of the vector in host cells (*e.g.*, origins of replication) and selectable marker genes. The selectable marker gene facilitates selection of host cells into which the vector has been introduced (see *e.g.*, U.S. Patents Nos. 4,399,216, 4,634,665 and 5,179,017). For example, typically the selectable marker gene confers resistance to drugs, such as G418, hygromycin or methotrexate, on a host cell into which the vector has been introduced. Preferred selectable marker genes include the dihydrofolate reductase (DHFR) gene (for use in *dhfr*$^-$ host cells with methotrexate selection/amplification) and the *neo* gene (for G418 selection).

In an exemplary system for recombinant expression of an antibody, or antigen-binding portion thereof, of the invention, a recombinant expression vector encoding both the antibody heavy chain and the antibody light chain is introduced into *dhfr*$^-$ CHO cells by calcium phosphate-mediated transfection. Within the recombinant expression vector, the antibody heavy and light chain genes are each operatively linked to enhancer/promoter regulatory elements (*e.g.*, derived from SV40, CMV, adenovirus and the like, such as a CMV enhancer/AdMLP promoter regulatory element or an SV40 enhancer/AdMLP promoter regulatory element) to drive high levels of transcription of the genes. The recombinant expression vector also carries a DHFR gene, which allows for selection of CHO cells that have been transfected with the vector using methotrexate selection/amplification. The selected transformant host cells are cultured to allow for expression of the antibody heavy and light chains and intact antibody is recovered from the culture medium. Standard molecular biology techniques are used to prepare the recombinant expression vector, transfect the host cells, select for transformants, culture the host cells and recover the antibody from the culture medium. For example, some antibodies can be isolated by affinity chromatography with a Protein A or Protein G.

For antibodies that include an Fc domain, the antibody production system preferably synthesizes antibodies in which the Fc region is glycosylated. For example, the Fc

domain of IgG molecules is glycosylated at asparagine 297 in the CH2 domain. This asparagine is the site for modification with biantennary-type oligosaccharides. It has been demonstrated that this glycosylation is required for effector functions mediated by Fc□ receptors and complement C1q (Burton and Woof, 1992, Adv. Immunol. 51 :1-84 ; Jefferis et al., 1998, Immunol. Rev. 163:59-76). In a preferred embodiment, the Fc domain is produced in a mammalian expression system that appropriately glycosylates the residue corresponding to asparagine 297. The Fc domain can also include other eukaryotic post-translational modifications.

Once an antibody molecule has been produced by recombinant expression, it may be purified by any method known in the art for purification of an immunoglobulin molecule, for example, by chromatography (*e.g.*, ion exchange, affinity, particularly by affinity for the specific antigen after Protein A, and sizing column chromatography), centrifugation, differential solubility, or by any other standard technique for the purification of proteins. Further, the antibodies or fragments thereof may be fused to heterologous polypeptide sequences known in the art to facilitate purification.

Gene Therapy Techniques

Gene therapy refers to therapy performed by the administration to a subject of an expressed or expressible nucleic acid. Any of the methods for gene therapy available in the art can be used according to the present invention. Exemplary methods are described below.

In specific embodiments, one or more antisense oligonucleotides for one or more biomarkers of the invention are administered to prevent, treat, manage or ameliorate bladder cancer or a symptom thereof, by way of gene therapy. In other embodiments, one or more nucleic acid molecules comprising nucleotides encoding one or more antibodies that specifically bind to one or more protein products of one or more biomarkers of the invention are administered to prevent, treat, manage or ameliorate bladder cancer or a symptom thereof, by way of gene therapy. In other embodiments, one or more nucleic acid molecules comprising nucleotides encoding protein products of one or more biomarkers of the invention or analogs, derivatives or fragments thereof, are administered to prevent, treat, manage or ameliorate bladder cancer or a symptom thereof, by way of gene therapy. In yet other

embodiments, one or more nucleic acid molecules comprising nucleotides encoding one or more dominant-negative polypeptides of one or more protein products of one or more biomarker of the invention are administered to prevent, treat, manage or ameliorate bladder cancer or a symptom thereof, by way of gene therapy.

For general reviews of the methods of gene therapy, see Goldspiel et al., 1993, Clinical Pharmacy 12:488-505; Wu and Wu, 1991, Biotherapy 3:87-95; Tolstoshev, 1993, Ann. Rev. Pharmacol. Toxicol. 32:573-596; Mulligan, 1993, Science 260:926-932; and Morgan and Anderson, 1993, Ann. Rev. Biochem. 62:191-217; May, 1993, TIBTECH 11(5):155-215). Methods commonly known in the art of recombinant DNA technology which can be used are described in Ausubel et al. (eds.), 1993, Current Protocols in Molecular Biology, John Wiley & Sons, NY; and Kriegler, 1990, Gene Transfer and Expression, A Laboratory Manual, Stockton Press, NY.

In one aspect, a composition of the invention comprises nucleic acid sequences encoding one or more antibodies that specifically bind to one or more protein products of one or more biomarkers of the invention, said nucleic acid sequences being part of expression vectors that express one or more antibodies in a suitable host. In particular, such nucleic acid sequences have promoters operably linked to the antibodies, said promoter being inducible or constitutive, and, optionally, tissue-specific.

In another aspect, a composition of the invention comprises nucleic acid sequences encoding dominant-negative polypeptides of one or protein products of one or more biomarkers of the invention, said nucleic acid sequences being part of expression vectors that express dominant-negative polypeptides in a suitable host. In particular, such nucleic acid sequences have promoters operably linked to the dominant-negative polypeptides, said promoter being inducible or constitutive, and, optionally, tissue-specific. In another particular embodiment, nucleic acid molecules are used in which the dominant-negative coding sequences and any other desired sequences are flanked by regions that promote homologous recombination at a desired site in the genome, thus providing for intrachromosomal expression of the dominant-negative nucleic acids (Koller and Smithies, 1989, Proc. Natl. Acad. Sci. USA 86:8932-8935; Zijlstra et al., 1989, Nature 342:435-438).

Delivery of the nucleic acids into a patient may be either direct, in which case the patient is directly exposed to the nucleic acid or nucleic acid-carrying vectors, or indirect, in which case, cells are first transformed with the nucleic acids *in vitro*, then transplanted into the patient. These two approaches are known, respectively, as *in vivo* or *ex vivo* gene therapy.

In a specific embodiment, the nucleic acid sequence is directly administered *in vivo*, where it is expressed to produce the encoded product. This can be accomplished by any of numerous methods known in the art, *e.g.*, by constructing it as part of an appropriate nucleic acid expression vector and administering it so that they become intracellular, *e.g.*, by infection using defective or attenuated retrovirals or other viral vectors (see U.S. Patent No. 4,980,286), or by direct injection of naked DNA, or by use of microparticle bombardment (*e.g.*, a gene gun; Biolistic, Dupont), or coating with lipids or cell-surface receptors or transfecting agents, encapsulation in liposomes, microparticles, or microcapsules, or by administering them in linkage to a peptide which is known to enter the nucleus, by administering it in linkage to a ligand subject to receptor-mediated endocytosis (see, *e.g.*, Wu and Wu, 1987, J. Biol. Chem. 262:4429-4432) (which can be used to target cell types specifically expressing the receptors), etc. In another embodiment, nucleic acid-ligand complexes can be formed in which the ligand comprises a fusogenic viral peptide to disrupt endosomes, allowing the nucleic acid to avoid lysosomal degradation. In yet another embodiment, the nucleic acid can be targeted *in vivo* for cell specific uptake and expression, by targeting a specific receptor (see, *e.g.*, International Publication Nos. WO 92/06180 dated April 16, 1992 (Wu et al.); WO 92/22635 dated December 23, 1992 (Wilson et al.); WO92/20316 dated November 26, 1992 (Findeis et al.); WO 93/14188 dated July 22, 1993 (Clarke et al.), WO 93/20221 dated October 14, 1993 (Young)). Alternatively, the nucleic acid can be introduced intracellularly and incorporated within host cell DNA for expression, by homologous recombination (Koller and Smithies, 1989, Proc. Natl. Acad. Sci. USA 86:8932-8935; Zijlstra et al., 1989, Nature 342:435-438).

For example, a retroviral vector can be used. These retroviral vectors have been modified to delete retroviral sequences that are not necessary for packaging of the viral genome and integration into host cell DNA. The nucleic acid sequences encoding the

antibodies of interest, or proteins of interest or fragments thereof to be used in gene therapy are cloned into one or more vectors, which facilitates delivery of the gene into a patient. More detail about retroviral vectors can be found in Boesen et al., 1994, Biotherapy 6:291-302, which describes the use of a retroviral vector to deliver the mdr1 gene to hematopoietic stem cells in order to make the stem cells more resistant to chemotherapy. Other references illustrating the use of retroviral vectors in gene therapy are: Clowes et al., 1994, J. Clin. Invest. 93:644-651; Kiem et al., 1994, Blood 83:1467-1473; Salmons and Gunzberg, 1993, Human Gene Therapy 4:129-141; and Grossman and Wilson, 1993, Curr. Opin. in Genetics and Devel. 3:110-114.

Adenoviruses are other viral vectors that can be used in gene therapy. Adenoviruses are especially attractive vehicles for delivering genes to respiratory epithelia. Adenoviruses naturally infect respiratory epithelia where they cause a mild disease. Other targets for adenovirus-based delivery systems are liver, the central nervous system, endothelial cells, and muscle. Adenoviruses have the advantage of being capable of infecting non-dividing cells. Kozarsky and Wilson, 1993, Current Opinion in Genetics and Development 3:499-503 present a review of adenovirus-based gene therapy. Bout et al., 1994, Human Gene Therapy 5:3-10 demonstrated the use of adenovirus vectors to transfer genes to the respiratory epithelia of rhesus monkeys. Other instances of the use of adenoviruses in gene therapy can be found in Rosenfeld et al., 1991, Science 252:431-434; Rosenfeld et al., 1992, Cell 68:143-155; Mastrangeli et al., 1993, J. Clin. Invest. 91:225-234; PCT Publication WO94/12649; and Wang, et al., 1995, Gene Therapy 2:775-783. In a preferred embodiment, adenovirus vectors are used.

Adeno-associated virus (AAV) has also been proposed for use in gene therapy (Walsh et al., 1993, Proc. Soc. Exp. Biol. Med. 204:289-300; U.S. Patent No. 5,436,146).

Another approach to gene therapy involves transferring a gene to cells in tissue culture by such methods as electroporation, lipofection, calcium phosphate mediated transfection, or viral infection. Usually, the method of transfer includes the transfer of a selectable marker to the cells. The cells are then placed under selection to isolate those cells that have taken up and are expressing the transferred gene. Those cells are then delivered to a patient.

In this embodiment, the nucleic acid is introduced into a cell prior to administration *in vivo* of the resulting recombinant cell. Such introduction can be carried out by any method known in the art, including but not limited to transfection, electroporation, microinjection, infection with a viral or bacteriophage vector containing the nucleic acid sequences, cell fusion, chromosome-mediated gene transfer, microcell-mediated gene transfer, spheroplast fusion, etc. Numerous techniques are known in the art for the introduction of foreign genes into cells (see, *e.g.*, Loeffler and Behr, 1993, Meth. Enzymol. 217:599-618; Cohen et al., 1993, Meth. Enzymol. 217:618-644; Cline, 1985, Pharmac. Ther. 29:69-92) and may be used in accordance with the present invention, provided that the necessary developmental and physiological functions of the recipient cells are not disrupted. The technique should provide for the stable transfer of the nucleic acid to the cell, so that the nucleic acid is expressible by the cell and preferably heritable and expressible by its cell progeny.

The resulting recombinant cells can be delivered to a patient by various methods known in the art. Recombinant blood cells (*e.g.*, hematopoietic stem or progenitor cells) and/or chondrocytes are preferably administered intravenously. The amount of cells envisioned for use depends on the desired effect, patient state, etc., and can be determined by one skilled in the art.

Cells into which a nucleic acid can be introduced for purposes of gene therapy encompass any desired, available cell type, and include but are not limited to epithelial cells, endothelial cells, keratinocytes, chondrocytes, fibroblasts, muscle cells, hepatocytes; blood cells such as T lymphocytes, B lymphocytes, monocytes, macrophages, neutrophils, eosinophils, megakaryocytes, granulocytes; various stem or progenitor cells, in particular hematopoietic stem or progenitor cells, *e.g.*, as obtained from bone marrow, umbilical cord blood, peripheral blood, fetal liver, etc.

In a preferred embodiment, the cell used for gene therapy is autologous to the patient.

In one embodiment in which recombinant cells are used in gene therapy, nucleic acid sequences encoding antibodies of interest, or proteins of interest or fragments thereof are introduced into the cells such that they are expressible by the cells or their

progeny, and the recombinant cells are then administered *in vivo* for therapeutic effect. In a specific embodiment, stem or progenitor cells are used. Any stem and/or progenitor cells which can be isolated and maintained *in vitro* can potentially be used in accordance with this embodiment of the present invention (see, *e.g.*, International Publication No. WO 94/08598, dated April 28, 1994; Stemple and Anderson, 1992, Cell 71:973-985; Rheinwald, 1980, Meth. Cell Bio. 21A:229; and Pittelkow and Scott, 1986, Mayo Clinic Proc. 61:771).

Promoters that may be used to control the expression of nucleic acid sequences encoding antibodies of interest, proteins of interest or fragments thereof may be constitutive, inducible or tissue- specific. Non-limiting examples include the SV40 early promoter region (Bernoist and Chambon, 1981, Nature 290:304-310), the promoter contained in the 3' long terminal repeat of Rous sarcoma virus (Yamamoto, et al., 1980, Cell 22:787-797), the herpes thymidine kinase promoter (Wagner et al., 1981, Proc. Natl. Acad. Sci. USA 78:1441-1445), the regulatory sequences of the metallothionein gene (Brinster et al., 1982, Nature 296:39-42); prokaryotic expression vectors such as the β-lactamase promoter (Villa-Kamaroff et al., 1978, Proc. Natl. Acad. Sci. USA 75:3727-3731), or the *tac* promoter (DeBoer et al., 1983, Proc. Natl. Acad. Sci. USA 80:21-25); see also "Useful proteins from recombinant bacteria" in Scientific American, 1980, 242:74-94; plant expression vectors comprising the nopaline synthetase promoter region (Herrera-Estrella et al., Nature 303:209-213) or the cauliflower mosaic virus 35S RNA promoter (Gardner et al., 1981, Nucl. Acids Res. 9:2871), and the promoter of the photosynthetic enzyme ribulose biphosphate carboxylase (Herrera-Estrella et al., 1984, Nature 310:115-120); promoter elements from yeast or other fungi such as the Gal 4 promoter, the ADC (alcohol dehydrogenase) promoter, PGK (phosphoglycerol kinase) promoter, alkaline phosphatase promoter, and the following animal transcriptional control regions, which exhibit tissue specificity and have been utilised in transgenic animals: elastase I gene control region which is active in pancreatic acinar cells (Swift et al., 1984, Cell 38:639-646; Ornitz et al., 1986, Cold Spring Harbor Symp. Quant. Biol. 50:399-409; MacDonald, 1987, Hepatology 7:425-515); insulin gene control region which is active in pancreatic beta cells (Hanahan, 1985, Nature 315:115-122), immunoglobulin gene control region which is active in lymphoid cells (Grosschedl et al., 1984, Cell 38:647-658; Adames et al., 1985, Nature 318:533-538; Alexander et al., 1987, Mol. Cell. Biol. 7:1436-1444), mouse mammary tumor virus control

region which is active in testicular, breast, lymphoid and mast cells (Leder et al., 1986, Cell 45:485-495), albumin gene control region which is active in liver (Pinkert et al., 1987, Genes and Devel. 1:268-276), alpha-fetoprotein gene control region which is active in liver (Krumlauf et al., 1985, Mol. Cell. Biol. 5:1639-1648; Hammer et al., 1987, Science 235:53-58; alpha 1-antitrypsin gene control region which is active in the liver (Kelsey et al., 1987, Genes and Devel. 1:161-171), beta-globin gene control region which is active in myeloid cells (Mogram et al., 1985, Nature 315:338-340; Kollias et al., 1986, Cell 46:89-94; myelin basic protein gene control region which is active in oligodendrocyte cells in the brain (Readhead et al., 1987, Cell 48:703-712); myosin light chain-2 gene control region which is active in skeletal muscle (Sani, 1985, Nature 314:283-286), and gonadotropic releasing hormone gene control region which is active in the hypothalamus (Mason et al., 1986, Science 234:1372-1378).

In a specific embodiment, the nucleic acid to be introduced for purposes of gene therapy comprises an inducible promoter operably linked to the coding region, such that expression of the nucleic acid is controllable by controlling the presence or absence of the appropriate inducer of transcription.

5.20 Pharmaceutical Compositions

Biologically active compounds identified using the methods of the invention or a pharmaceutically acceptable salt thereof can be administered to a patient, preferably a mammal, more preferably a human, suffering from bladder cancer. In a specific embodiment, a compound or pharmaceutically acceptable salt thereof is administered to a patient, preferably a mammal, more preferably a human, suffering from the following bladder cancer and/or early stage bladder cancer. In another embodiment, a compound or a pharmaceutically acceptable salt thereof is administered to a patient, preferably a mammal, more preferably a human, as a preventative measure against bladder cancer. In accordance with these embodiments, the patient may be a child, an adult or elderly, wherein a "child" is a subject between the ages of 24 months of age and 18 years of age, an "adult" is a subject 18 years of age or older, and "elderly" is a subject 65 years of age or older.

When administered to a patient, the compound or a pharmaceutically acceptable salt thereof is preferably administered as component of a composition that

optionally comprises a pharmaceutically acceptable vehicle. The composition can be administered orally, or by any other convenient route, for example, by infusion or bolus injection, by absorption through epithelial or mucocutaneous linings (*e.g.*, oral mucosa, rectal, and intestinal mucosa, *etc.*) and may be administered together with another biologically active agent. Administration can be systemic or local. Various delivery systems are known, *e.g.*, encapsulation in liposomes, microparticles, microcapsules, capsules, *etc.*, and can be used to administer the compound and pharmaceutically acceptable salts thereof.

Methods of administration include but are not limited to intradermal, intramuscular, intraperitoneal, intravenous, subcutaneous, intranasal, epidural, oral, sublingual, intranasal, intracerebral, intravaginal, transdermal, rectally, by inhalation, or topically, particularly to the ears, nose, eyes, or skin. The mode of administration is left to the discretion of the practitioner. In most instances, administration will result in the release of the compound or a pharmaceutically acceptable salt thereof into the bloodstream.

In specific embodiments, it may be desirable to administer the compound or a pharmaceutically acceptable salt thereof locally. This may be achieved, for example, and not by way of limitation, by local infusion during surgery, topical application, *e.g.*, in conjunction with a wound dressing after surgery, by injection, by means of a catheter, by means of a suppository, or by means of an implant, said implant being of a porous, non-porous, or gelatinous material, including membranes, such as sialastic membranes, or fibers.

In certain embodiments, it may be desirable to introduce the compound or a pharmaceutically acceptable salt thereof into the central nervous system by any suitable route, including intraventricular, intrathecal and epidural injection. Intraventricular injection may be facilitated by an intraventricular catheter, for example, attached to a reservoir, such as an Ommaya reservoir.

Pulmonary administration can also be employed, *e.g.*, by use of an inhaler or nebulizer, and formulation with an aerosolizing agent, or via perfusion in a fluorocarbon or synthetic pulmonary surfactant. In certain embodiments, the compound and pharmaceutically acceptable salts thereof can be formulated as a suppository, with traditional binders and vehicles such as triglycerides.

In another embodiment, the compound and pharmaceutically acceptable salts thereof can be delivered in a vesicle, in particular a liposome (see Langer, 1990, Science 249:1527-1533; Treat et al., in Liposomes in the Therapy of Infectious Disease and Cancer, Lopez-Berestein and Fidler (eds.), Liss, New York, pp. 353-365 (1989); Lopez-Berestein, *ibid.*, pp. 317-327; see generally *ibid.*).

In yet another embodiment, the compound and pharmaceutically acceptable salts thereof can be delivered in a controlled release system (see, *e.g.*, Goodson, in Medical Applications of Controlled Release, supra, vol. 2, pp. 115-138 (1984)). Other controlled-release systems discussed in the review by Langer, 1990, Science 249:1527-1533 may be used. In one embodiment, a pump may be used (see Langer, *supra*; Sefton, 1987, CRC Crit. Ref. Biomed. Eng. 14:201; Buchwald et al., 1980, Surgery 88:507; Saudek et al., 1989, N. Engl. J. Med. 321:574). In another embodiment, polymeric materials can be used (see Medical Applications of Controlled Release, Langer and Wise (eds.), CRC Pres., Boca Raton, Florida (1974); Controlled Drug Bioavailability, Drug Product Design and Performance, Smolen and Ball (eds.), Wiley, New York (1984); Ranger and Peppas, 1983, J. Macromol. Sci. Rev. Macromol. Chem. 23:61; see also Levy et al., 1985, Science 228:190; During et al., 1989, Ann. Neurol. 25:351; Howard et al., 1989, J. Neurosurg. 71:105). In yet another embodiment, a controlled-release system can be placed in proximity of a target RNA of the compound or a pharmaceutically acceptable salt thereof, thus requiring only a fraction of the systemic dose.

The compounds described herein can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the active compound and a pharmaceutically acceptable carrier. As used herein the language "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

The invention includes methods for preparing pharmaceutical compositions for modulating the expression or activity of a polypeptide or nucleic acid of interest. Such methods comprise formulating a pharmaceutically acceptable carrier with an agent that modulates expression or activity of a polypeptide or nucleic acid of interest. Such compositions can further include additional active agents. Thus, the invention further includes methods for preparing a pharmaceutical composition by formulating a pharmaceutically acceptable carrier with an agent that modulates expression or activity of a polypeptide or nucleic acid of interest and one or more additional active compounds.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*, inhalation), transdermal (topical), transmucosal, and rectal administration. Intravenous administration is preferred. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersions. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF; Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example,

glycerol, propylene glycol, and liquid polyetheylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as mannitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (*e.g.*, a polypeptide or antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed.

Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as

magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from a pressurized container or dispenser which contains a suitable propellant, *e.g.*, a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (*e.g.*, with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as

used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

For antibodies, the preferred dosage is 0.1 mg/kg to 100 mg/kg of body weight (more preferably, 0.1 to 20 mg/kg, 0.1-10 mg/kg, or 0.1 to 1.0 mg/kg). If the antibody is to act in the brain, a dosage of 50 mg/kg to 100 mg/kg is usually appropriate. Generally, partially human antibodies and fully human antibodies have a longer half-life within the human body than other antibodies. Accordingly, lower dosages and less frequent administration is often possible. Modifications such as lipidation can be used to stabilize antibodies and to enhance uptake and tissue penetration (*e.g.*, into the brain). A method for lipidation of antibodies is described by Cruikshank et al. (1997, J. Acquired Immune Deficiency Syndromes and Human Retrovirology 14:193).

In a specific embodiment, an effective amount of protein or polypeptide (*i.e.*, an effective dosage) ranges from about 0.001 to 30 mg/kg body weight, preferably about 0.01 to 25 mg/kg body weight, more preferably about 0.1 to 20 mg/kg body weight, and even more preferably about 0.1 to 1.0 mg/kg, 1 to 10 mg/kg, 2 to 9 mg/kg, 3 to 8 mg/kg, 4 to 7 mg/kg, or 5 to 6 mg/kg body weight.

The skilled artisan will appreciate that certain factors may influence the dosage required to effectively treat a subject, including but not limited to the severity of the disease or disorder, previous treatments, the general health and/or age of the subject, and other diseases present. Moreover, treatment of a subject with a therapeutically effective amount of a protein, polypeptide, or antibody can include a single treatment or, preferably, can include a series of treatments.

In addition to those compounds described above, the present invention encompasses the use of small molecules that modulate expression or activity of a nucleic acid or polypeptide of interest. Non-limiting examples of small molecules include peptides,

peptidomimetics, amino acids, amino acid analogs, polynucleotides, polynucleotide analogs, nucleotides, nucleotide analogs, organic or inorganic compounds (*i.e.*,. including heteroorganic and organometallic compounds) having a molecular weight less than about 10,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 5,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 1,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 500 grams per mole, and salts, esters, and other pharmaceutically acceptable forms of such compounds.

It is understood that appropriate doses of small molecule agents depends upon a number of factors within the ken of the ordinarily skilled physician, veterinarian, or researcher. The dose(s) of the small molecule will vary, for example, depending upon the identity, size, and condition of the subject or sample being treated, further depending upon the route by which the composition is to be administered, if applicable, and the effect which the practitioner desires the small molecule to have upon the nucleic acid or polypeptide of the invention. Exemplary doses include milligram or microgram amounts of the small molecule per kilogram of subject or sample weight (*e.g.*, about 1 microgram per kilogram to about 500 milligrams per kilogram, about 100 micrograms per kilogram to about 5 milligrams per kilogram, or about 1 microgram per kilogram to about 50 micrograms per kilogram). It is furthermore understood that appropriate doses of a small molecule depend upon the potency of the small molecule with respect to the expression or activity to be modulated. Such appropriate doses may be determined using the assays described herein. When one or more of these small molecules is to be administered to a subject (*e.g.*, a human) in order to modulate expression or activity of a polypeptide or nucleic acid of the invention, a physician, veterinarian, or researcher may, for example, prescribe a relatively low dose at first, subsequently increasing the dose until an appropriate response is obtained. In addition, it is understood that the specific dose level for any particular animal subject will depend upon a variety of factors including the activity of the specific compound employed, the age, body weight, general health, gender, and diet of the subject, the time of administration, the route of administration, the rate of excretion, any drug combination, and the degree of expression or activity to be modulated.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

5.21 Kits

The present invention provides kits for measuring the expression of the protein and RNA products of at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, all or any combination of the biomarkers of the invention. Such kits comprise materials and reagents required for measuring the expression of such protein and RNA products. In specific embodiments, the kits may further comprise one or more additional reagents employed in the various methods, such as: (1) reagents for stabilizing and/or purifying RNA from blood, or tissue (2) primers for generating test nucleic acids; (3) dNTPs and/or rNTPs (either premixed or separate), optionally with one or more uniquely labelled dNTPs and/or rNTPs (*e.g.*, biotinylated or Cy3 or Cy5 tagged dNTPs); (4) post synthesis labelling reagents, such as chemically active derivatives of fluorescent dyes; (5) enzymes, such as reverse transcriptases, DNA polymerases, and the like; (6) various buffer mediums, *e.g.*, reaction, hybridization and washing buffers; (7) labelled probe purification reagents and components, like spin columns, etc.; and (8) protein purification reagents; (9) signal generation and detection reagents, *e.g.*, streptavidin-alkaline phosphatase conjugate, chemifluorescent or chemiluminescent substrate, and the like. In particular embodiments, the kits comprise prelabeled quality controlled protein and or RNA isolated from a sample (*e.g.*, blood or chondrocytes or synovial fluid) for use as a control.

In some embodiments, the kits are RT-PCR or qRT-PCR kits. In other embodiments, the kits are nucleic acid arrays and protein arrays. Such kits according to the subject invention will at least comprise an array having associated protein or nucleic acid members of the invention and packaging means therefore. Alternatively the protein or nucleic acid members of the invention may be prepackaged onto an array.

In some embodiments, the kits are Quantitative RT-PCR kits. In one embodiment, the quantitative RT-PCR kit includes the following: (a) primers used to amplify each of a combination of biomarkers of the invention; (b) buffers and enzymes including an reverse transcripate; (c) one or more thermos table polymerases; and (d) Sybr® Green. In

another embodiment, the kit of the invention also includes (a) a reference control RNA and (b) a spiked control RNA.

The invention provides kits that are useful for (a) diagnosing individuals as having bladder cancer and/or early stage bladder cancer. For example, in a particular embodiment of the invention a kit is comprised a forward and reverse primer wherein the forward and reverse primer are designed to quantitate expression of all of the species of mRNA corresponding to each of the biomarkers as identified in accordance with the invention useful in determining whether an individual has bladder cancer and/or early stage bladder cancer or not. In certain embodiments, at least one of the primers is designed to span an exon junction.

The invention provides kits that are useful for detecting, diagnosing, monitoring and prognosing bladder cancer based upon the expression of protein or RNA products of at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, all or any combination of the biomarkers of the invention in a sample. In certain embodiments, such kits do not include the materials and reagents for measuring the expression of a protein or RNA product of a biomarker of the invention that has been suggested by the prior art to be associated with bladder cancer. In other embodiments, such kits include the materials and reagents for measuring the expression of a protein or RNA product of a biomarker of the invention that has been suggested by the prior art to be associated with bladder cancer and at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45 or more genes other than the biomarkers of the invention.

The invention provides kits useful for monitoring the efficacy of one or more therapies that a subject is undergoing based upon the expression of a protein or RNA product of any number of up to at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, all or any combination of the biomarkers of the invention in a sample. In certain embodiments, such kits do not include the materials and reagents for measuring the expression of a protein or RNA product of a biomarker of the invention that

has been suggested by the prior art to be associated with bladder cancer. In other embodiments, such kits include the materials and reagents for measuring the expression of a protein or RNA product of a biomarker of the invention that has been suggested by the prior art to be associated with bladder cancer and any number of up to at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45 or more genes other than the biomarkers of the invention.

The invention provides kits using for determining whether a subject will be responsive to a therapy based upon the expression of a protein or RNA product of any number of up to at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, all or any combination of the biomarkers of the invention in a sample. In certain embodiments, such kits do not include the materials and reagents for measuring the expression of a protein or RNA product of a biomarker of the invention that has been suggested by the prior art to be associated with bladder cancer. In other embodiments, such kits include the materials and reagents for measuring the expression of a protein or RNA product of a biomarker of the invention that has been suggested by the prior art to be associated with bladder cancer and any number of up to at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45 or more genes other than the biomarkers of the invention.

The invention provides kits for measuring the expression of a RNA product of any number of up to at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, all or any combination of the biomarkers of the invention in a sample. In a specific embodiment, such kits comprise materials and reagents that are necessary for measuring the expression of a RNA product of a biomarker of the invention. For example, a microarray or RT-PCR kit may be produced for bladder cancer and contain only those reagents and materials necessary for measuring the levels of RNA products of any number of up to at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at

least 8, at least 9, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, all or any combination of the biomarkers of the invention. Alternatively, in some embodiments, the kits can comprise materials and reagents that are not limited to those required to measure the levels of RNA products of any number of up to1,2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, all or any combination of the biomarkers of the invention. For example, a microarray kit may contain reagents and materials necessary for measuring the levels of RNA products of not necessarily associated with or indicative of bladder cancer, in addition to reagents and materials necessary for measuring the levels of the RNA products of any number of up to at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, all or any combination of the biomarkers of the invention. In a specific embodiment, a microarray or RT-PCR kit contains reagents and materials necessary for measuring the levels of RNA products of any number of up to at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, all or any combination of the biomarkers of the invention, and any number of up to 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 125, 150, 175, 200, 225, 250, 300, 350, 400, 450, or more genes other than the biomarkers of the invention, or 1-10, 1-100, 1-150, 1-200, 1-300, 1-400, 1-500, 1-1000, 25-100, 25-200, 25-300, 25-400, 25-500, 25-1000, 100-150, 100-200, 100-300, 100-400, 100-500, 100-1000, 500-1000 other genes than the biomarkers of the invention.

For nucleic acid micoarray kits, the kits generally comprise probes attached to a solid support surface. The probes may be labelled with a detectable label. In a specific embodiment, the probes are specific for an exon(s), an intron(s), an exon junction(s), or an exon-intron junction(s)), of RNA products of any number of up to 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, all or any combination of the biomarkers of the invention. The microarray kits may comprise instructions for performing the assay and methods for interpreting and analyzing the data resulting from the performance of the assay. In a specific embodiment, the kits comprise instructions for diagnosing bladder cancer. The kits may also comprise hybridization reagents and/or reagents necessary for detecting a signal produced when a probe hybridizes to a target nucleic acid sequence. Generally, the materials and

reagents for the microarray kits are in one or more containers. Each component of the kit is generally in its own a suitable container.

For RT-PCR kits, the kits generally comprise pre-selected primers specific for particular RNA products (*e.g.*, an exon(s), an intron(s), an exon junction(s), and an exon-intron junction(s)) of any number of up to 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, all or any combination of the biomarkers of the invention. The RT-PCR kits may also comprise enzymes suitable for reverse transcribing and/or amplifying nucleic acids (*e.g.*, polymerases such as Taq), and deoxynucleotides and buffers needed for the reaction mixture for reverse transcription and amplification. The RT-PCR kits may also comprise probes specific for RNA products of any number of up to 1,2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, all or any combination of the biomarkers of the invention. The probes may or may not be labelled with a detectable label (*e.g.*, a fluorescent label). Each component of the RT-PCR kit is generally in its own suitable container. Thus, these kits generally comprise distinct containers suitable for each individual reagent, enzyme, primer and probe. Further, the RT-PCR kits may comprise instructions for performing the assay and methods for interpreting and analyzing the data resulting from the performance of the assay. In a specific embodiment, the kits contain instructions for diagnosing bladder cancer.

In a specific embodiment, the kit is a real-time RT-PCR kit. Such a kit may comprise a 96 well plate and reagents and materials necessary for SYBR Green detection. The kit may comprise reagents and materials so that beta-actin can be used to normalize the results. The kit may also comprise controls such as water, phosphate buffered saline, and phage MS2 RNA. Further, the kit may comprise instructions for performing the assay and methods for interpreting and analyzing the date resulting from the performance of the assay. In a specific embodiment, the instructions state that the level of a RNA product of any number of up to 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, all or any combination of the biomarkers of the invention should be examined at two concentrations that differ by, *e.g.*, 5 fold to 10-fold.

For antibody based kits, the kit can comprise, for example: (1) a first antibody (which may or may not be attached to a solid support) which binds to protein of interest (*e.g.*, a protein product of any number of up to 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45,

50, all or any combination of the biomarkers of the invention); and, optionally, (2) a second, different antibody which binds to either the protein, or the first antibody and is conjugated to a detectable label (*e.g.*, a fluorescent label, radioactive isotope or enzyme). The antibody-based kits may also comprise beads for conducting an immunoprecipitation. Each component of the antibody-based kits is generally in its own suitable container. Thus, these kits generally comprise distinct containers suitable for each antibody. Further, the antibody-based kits may comprise instructions for performing the assay and methods for interpreting and analyzing the data resulting from the performance of the assay. In a specific embodiment, the kits contain instructions for diagnosing bladder cancer.

6. EXAMPLES

The examples below are non-limiting and are merely representative of various aspects and features of the present invention

EXAMPLE 1

MICROARRAY CONSTRUCTION

An array according to one aspect of the invention was constructed as follows.

PCR products (~40 ul) of cDNA clones from OA cartilage cDNA libraries, in the same 96-well tubes used for amplification, are precipitated with 4 ul (1/10 volume) of 3M sodium acetate (pH 5.2) and 100 ul (2.5 volumes) of ethanol and stored overnight at –20°C. They are then centrifuged at 3,300 rpm at 4°C for 1 hour. The obtained pellets were washed with 50 ul ice-cold 70% ethanol and centrifuged again for 30 minutes. The pellets are then air-dried and resuspended well in 50% dimethylsulfoxide (DMSO) or 20 ul 3X SSC overnight. The samples are then deposited either singly or in duplicate onto Gamma Amino Propyl Silane (Corning CMT-GAPS or CMT-GAP2, Catalog No. 40003, 40004) or polylysine-coated slides (Sigma Cat. No. P0425) using a robotic GMS 417 or 427 arrayer (Affymetrix, CA). The boundaries of the DNA spots on the microarray are marked with a diamond scriber. The invention provides for arrays where 10-20,000 PCR products are spotted onto a solid support to prepare an array.

The arrays are rehydrated by suspending the slides over a dish of warm particle free ddH_20 for approximately one minute (the spots will swell slightly but not run

into each other) and snap-dried on a 70-80°C inverted heating block for 3 seconds. DNA is then UV crosslinked to the slide (Stratagene, Stratalinker, 65 mJ – set display to "650" which is 650 x 100 uJ) or baked at 80C for two to four hours. The arrays are placed in a slide rack. An empty slide chamber is prepared and filled with the following solution: 3.0 grams of succinic anhydride (Aldrich) is dissolved in 189 ml of 1-methyl-2-pyrrolidinone (rapid addition of reagent is crucial); immediately after the last flake of succinic anhydride dissolved, 21.0 ml of 0.2 M sodium borate is mixed in and the solution is poured into the slide chamber. The slide rack is plunged rapidly and evenly in the slide chamber and vigorously shaken up and down for a few seconds, making sure the slides never leave the solution, and then mixed on an orbital shaker for 15-20 minutes. The slide rack is then gently plunged in 95°C dd$H_2$0 for 2 minutes, followed by plunging five times in 95% ethanol. The slides are then air dried by allowing excess ethanol to drip onto paper towels. The arrays are then stored in the slide box at room temperature until use.

EXAMPLE 2

RNA ISOLATION FROM UNFRACTIONATED WHOLE BLOOD

(a) LYSED BLOOD

Ten ml of peripheral whole blood was collected in EDTA Vacutainer tubes (Becton Dickinson, Franklin Lakes, N.J.) and stored on ice until processing (within 6 hours). Upon centrifugation, blood samples separated into plasma, buffy coat and red blood cell layers. The plasma was removed and a hypotonic buffer (1.6 mM EDTA, 10 mM $KHCO_3$, 153 mM NH_4Cl, pH 7.4) was added to lyse the red blood cells at a 3:1 volume ratio. The mixture was centrifuged to yield a cell pellet, which was dissolved and homogenized into 1.0 ml of TRIzol® Reagent (Invitrogen Corp., Carlsbad, CA) and 0.2 ml of chloroform according to the manufacture's instructions. After centrifugation, isopropanol was added to the aqueous phase at a 1:1 ratio and allowed to precipitate at –20°C. Subsequent centrifugation yielded an RNA pellet that was resuspended in water for experimental use. RNA quality was assessed on Agilent 2100 Bioanalyzer RNA 6000 Nano Chips as specified by the manufacturer, and RNA quantity was determined by absorbance at 260 nm in a Beckman-Coulter DU640 Spectrophotometer.

(b) CENTRIFUGED LYSED BLOOD

10 ml whole blood is obtained in a Vacutainer and spun at 2,000 rpm (800g) for 5 min at 4°C and the plasma layer removed. Lysis Buffer is added to blood sample in a ratio of 3 parts Lysis Buffer to 1 part blood (Lysis Buffer (1L) 0.6g EDTA; 1.0g KHCO2, 8.2g NH4Cl adjusted to pH 7.4 (using NaOH)). Sample is mixed and placed on ice for 5-10 minutes until transparent. Lysed sample is centrifuged at 1000 rpm for 10 minutes at 4°C, and supernatant is aspirated. Pellet is resuspended in 5 ml Lysis Buffer, and centrifuged again at 1000 rpm for 10 minutes at 4°C. Pelleted cells are homogenized using TRIzol (GIBCO/BRL) in a ratio of approximately 6 ml of TRIzol for every 10 ml of the original blood sample and vortexed well. Samples are left for 5 minutes at room temperature. RNA is extracted using 1.2 ml of chloroform per 1 ml of TRIzol. Sample is centrifuged at 12,000 x g for 5 minutes at 4°C and upper layer is collected. To upper layer, isopropanol is added in ratio of 0.5 ml per 1 ml of TRIzol. Sample is left overnight at -20°C or for one hour at -20°C. RNA is pelleted in accordance with known methods, RNA pellet air dried, and pellet resuspended in DEPC treated ddH2O. RNA samples can also be stored in 75% ethanol where the samples are stable at room temperature for transportation.

(c)FROM SERUM FREE WHOLE BLOOD

10 ml whole blood is obtained in a Vacutainer and spun at 2,000 rpm (800g) for 5 min at 4°C and the plasma layer removed. Pelleted cells are homogenized using TRIzol (GIBCO/BRL) in a ratio of approximately 6 ml of TRIzol for every 10 ml of the original blood sample and vortexed well. Samples are left for 5 minutes at room temperature. RNA is extracted using 1.2 ml of chloroform per 1 ml of TRIzol. Sample is centrifuged at 12,000 x g for 5 minutes at 4°C and upper layer is collected. To upper layer, isopropanol is added in ratio of 0.5 ml per 1 ml of TRIzol. Sample is left overnight at -20°C or for one hour at -20°C. RNA is pelleted in accordance with known methods, RNA pellet air dried, and pellet resuspended in DEPC treated ddH2O. RNA samples can also be stored in 75% ethanol where the samples are stable at room temperature for transportation.

(d)FROM WHOLE BLOOD USING PAXGENE™

2.5 ml whole blood is collected into PAXgene Blood RNA Tubes and processed in accordance with the instructions of the PAXgene™ Blood RNA Kit protocol. In brief after storing the blood in the PAXgene™ tube for at least 2 hours, the blood sample is centriguted and the supernatant discarded. To the remaining sample, 360 ul of the supplied Buffer BR1 is added and the sample is pipetted into the spin column and centrifuged an then washed with numerous wash steps and finally eluted and stored.

(e) FROM WHOLE BLOOD USING PAXGENE™ AND SUBSEQUENT GLOBIN REDUCTION

RNA isolated from PAXGene™ as noted in (d) above is subsequently treated to selectively remove globin mRNA as is described in Affymetrix® technical note entitled "Globin Reduction Protocol". Oligonucleotides specific for the alpha 1, alpha 2 and beta globin species are incubated with an oligonucleotide hybridization buffer and RNAse H used to specifically target degradation of the globin mRNA and the cRNA clean up column from Affymetrix used to remove the globin mRNA.

EXAMPLE 3

TARGET NUCLEIC ACID PREPARATION AND HYBRIDIZATION

Preparation of Fluorescent DNA Probe from mRNA

Fluorescently labelled target nucleic acid samples of RNA were prepared for analysis with an array of the invention.

1 μg Oligo-dT primers were annealed to 10 μg of total RNA isolated from unfractionated whole blood from patient diagnosed with bladder cancer in a total volume of 10 μl, by heating to 70°C for 10 min, and cooled on ice. Patients were diagnosed as positive or negative for bladder cancer by a registered physician. A positive colon pathology diagnosis was followed by histological examination of the excised tissue(s). The mRNA was reverse transcribed by incubating the sample at 42°C for 40 min in a 25μl volume containing a final concentration of 50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM $MgCl_2$, 25 mM DTT, 25 mM unlabeled dNTPs, 400 units of Superscript II (200 U/μL, Gibco BRL), and 15 mM of Cy3 or

Cy5 (Amersham). The reaction was stopped by the addition of 2.5µl of 500mM EDTA and5µl of 1M NaOH, and incubation at 65°C for 10 min. The reaction mixture is neutralized by addition of 12.5µl of 1M TrisHCl (pH7.6).

The labeled target nucleic acid sample was purified by centrifugation in a Centricon-30 micro-concentrator (Amicon).

Labeled nucleic acid was denatured by heating for 2 min at 100°C, and incubated at 37°C for 20-30 min before being placed on a nucleic acid array under a 22mm x 22mm glass cover slip. Hybridization was carried out at 65°C for 14 to 18 hours in a custom slide chamber with humidity maintained by a small reservoir of 3XSSC. The array was washed by submersion and agitation for 2-5 min in 2X SSC with 0.1%SDS, followed by 1X SSC, and 0.1X SSC. Finally, the array was dried by centrifugation for 2 min in a slide rack in a Beckman GS-6 tabletop centrifuge in Microplus carriers at 650 RPM for 2 min.

Total RNA (5 µg) was labeled and hybridized onto Affymetrix U133Plus 2.0 GeneChips (Affymetrix; Santa Clara, CA) along with other similarly prepared samples from individuals having or not having bladder cancer and hybridized according to the manufacturer's instructions. Briefly, hybridization signals were scaled in the Affymetrix GCOS software (version 1.1.1), using a scaling factor determined by adjusting the global trimmed mean signal intensity value to 500 for each array, and imported into GeneSpring version 7.2 (Silicon Genetics; Redwood City, CA). Signal intensities were then centered to the 50th percentile of each chip, and for each individual gene, to the median intensity of the whole sample set. Only genes called present or marginal by the GCOS software in at least 80% of each group of samples were used for further analysis. Differentially expressed genes were identified using 1) the non-parametric Wilcoxon-Mann-Whitney non-parametric test (P<0.05), 2) parametric t test (P <0.05), and 3) unsupervised analysis method (14). In the un-supervised analysis, the signal-intensity filtered genes were used to select genes with at least 2-fold change (up or down) in expression level, away from the mean, in at least 15% of the samples. Hierarchical cluster analysis was performed on each comparison to assess correlation analysis using Spearman correlation among samples for each identified gene set as the similarity measure with average centroid linkage in GeneSpring v6.0. Results from

experiments comparing individuals having bladder cancer with individuals not having bladder cancer are provided in **Table 1.** Results from experiments comparing individuals having early stage bladder cancer as compared with individuals not having bladder cancer are provided in **Table 4.** Results from experiments comparing individuals having bladder cancer as compared with individuals having either testicular cancer or renal cell carcinoma are shown in **Table 11.** A selection of those genes identified in **Table 1** as biomarkers of bladder cancer are shown in **Table 10** below. These genes are particularly useful because they demonstrate a significant fold change with a significant p value.

TABLE 10

							Supervised non-parametric:WMW					
Affy ID	Accession	Gene Name	Description	Ctrl mean	Ctrl SD	SD/mean Ctrl	BC mean	BC SD	SD/mean BC	BC/ctrl	direction	P-value
208478_s_at	NM_004324	BAX	BCL2-associated X protein /// BCL2-associated X protein	0.83	0.25	0.30	1.28	0.81	0.64	1.55	↑	0.007554
201163_s_at	NM_001553	IGFBP7	insulin-like growth factor binding protein 7 /// insulin-like growth factor binding protein 7	0.83	0.28	0.34	1.29	0.87	0.68	1.56	↑	0.030171
242903_at	NM_00416	IFNGR1	interferon gamma receptor 1	0.86	0.58	0.67	1.31	0.47	0.36	1.53	↑	0.004288
205992_s_at	NM_000585	IL15	interleukin 15 /// interleukin 15	0.82	0.21	0.26	1.19	0.63	0.53	1.45	↑	0.007554

203085_s_at	NM_000660	TGFB1	transforming growth factor, beta 1 (Camurati-Engelmann disease) /// transforming growth factor, beta 1 (Camurati-Engelmann disease)	0.85	0.22	0.26	1.16	0.48	0.41	1.37	↑	0.014664
214525_x_at	NM_014381	MLH3	mutL homolog 3 (E. coli)	1.48	0.54	0.37	0.91	0.26	0.29	0.61	↓	0.001177
235372_at	NM_145141	FREB	Fc receptor homolog expressed in B cells	1.71	0.85	0.49	1.14	0.50	0.44	0.66	↓	0.005494
205539_at	NM_006576	AVIL	advillin /// advillin	1.42	0.51	0.36	0.95	0.28	0.29	0.67	↓	9.62E-04
206608_s_at	NM_020366	RPGRIP1	retinitis pigmentosa GTPase regulator interacting protein 1 /// retinitis pigmentosa GTPase regulator interacting protein 1	1.50	0.55	0.37	1.01	0.21	0.21	0.67	↓	0.011846
1554014_at	NM_001271	CHD2	chromodomain helicase DNA binding protein 2	1.06	0.17	0.16	0.98	0.10	0.11	0.92	↓	0.040701

EXAMPLE 4

QUANTITATIVE REAL TIME RT PCR

A selection of genes from **Table 1** identified using Affymetrix U133Plus 2.0 GeneChips and noted in **Table 13** below were quantified using quantitative real time RT PCR (QRT-PCR). **Table 13** provides a list of the selected biomarkers. **Table 14** provides a list of the products of the biomarkers noted in **Table 13**. A summary of the results of the quantitative real time RT-PCR using primers as noted in **Table 15** is shown in **Table 16**. QRT-PCR was done using either the SYBR® Green Kit from Qiagen (Product Number 204143) and/or using Applied Biosystems PCR reagent kits (Cat 4334973). Amplicons were detected in real time using a Prism 7500 instrument (Applied Biosystems).

Reverse transcription was first performed using the High-Capacity cDNA Archive Kit from Applied Biosystems (Product number 4322171) and following the protocol utilized therein.

More specifically, purified RNA as described previously herein was incubated with Reverse Transcriptase buffer, dNTPs, Random primers and Reverse transcriptase and incubated for 25°C for 10 minutes and subsequently for 37°C for two hours and the resulting mixture utilized as the starting product for quantitative PCR.

cDNA resulting from reverse transcription was incubated with the QuantiTect SYBR® Green PCR Master Mix as provided and no adjustments were made for magnesium concentration. Uracil-N-Glycosylase was not added. 5μM of both forward primer and reverse primer specific to the genes of the invention were added and the reaction was incubated and monitored in accordance with the standard protocol utilizing the ABI PRISM 7700/ABI GeneAmp 5700/iCycler/DNA Engine Opticon. Forward and reverse primers for the candidate biomarkers were designed using "PrimerQuest". A tool found at http://biotools.idtdna.com/primerquest (Integrated DNA Technologies, Coralville, IA). ΔCt measurements for each of the genes relative to a housekeeping gene (beta-actin) were determined The melting temperature [Tm] in thermal dissociation, and examination on agarose gels provided confirmation of specific PCR amplification and the lack of primer-dimer formation in each

reaction well. For individual target gene analysis, changes in Ct value between each gene and the Beta-actin house-keeping was calculated as ΔCt= Ct (target gene) - Ct (house-keeping gene).

QRT PCR can also be performed on the RNA products of the remaining biomarkers disclosed in **Table 1** to the products noted in **Table 3** using, for example, the SYBR® Green Kit from Qiagen (Product Number 204143).

Either a one step (reverse transcription and PCR combined) or a two step (reverse transcription first and then subsequent PCR) can be used. In the case of the two step protocol, reverse transcription was first performed using the High-Capacity cDNA Archive Kit from Applied Biosystems (Product number 4322171) and following the protocol utilized therein.

More specifically purified RNA as described previously herein was incubated with Reverse Transcriptase buffer, dNTPs, Random primers and Reverse transcriptase and incubated for 25°C for 10 minutes and subsequently for 37°C for two hours and the resulting mixture utilized as the starting product for quantitative PCR.

cDNA resulting from reverse transcription can be incubated with the QuantiTect SYBR® Green PCR Master Mix as provided and no adjustments made for magnesium concentration. Uracil-N-Glycosylase is optional. 5μM of both forward primer and reverse primer specific to the genes of the invention are added and the reaction incubated and monitored in accordance with the standard protocol utilizing the ABI PRISM 7700/ABI GeneAmp 5700/iCycler/DNA Engine Opticon.

Table 13

Selection of biomarkers of Table 1. Table demonstrates the Hugo Gene Name (GeneSymbol) the Entrez Gene ID (formerly locus link) and the Gene description.

GeneSymbol	*GeneID*	*Annotation*
CHD2	1106	Homo sapiens chromodomain helicase DNA binding protein 2 (CHD2), mRNA
CSPG6	9126	Homo sapiens chondroitin sulfate proteoglycan 6 (bamacan) (CSPG6), mRNA
CTSD	1509	Homo sapiens cathepsin D (lysosomal aspartyl protease) (CTSD), mRNA
FREB	84824	Homo sapiens Fc receptor homolog expressed in B cells (FREB), mRNA

GAS7	8522	Homo sapiens growth arrest-specific 7 (GAS7), transcript variant b, mRNA
GAS7	8522	Homo sapiens growth arrest-specific 7 (GAS7), transcript variant c, mRNA
HIST1H1C	3006	Homo sapiens histone 1, H1c (HIST1H1C), mRNA
IGFBP7	3490	Homo sapiens insulin-like growth factor binding protein 7 (IGFBP7), mRNA
IRAK3	11213	Homo sapiens interleukin-1 receptor-associated kinase 3 (IRAK3), mRNA
IREB2	3658	Homo sapiens iron-responsive element binding protein 2 (IREB2), mRNA
MLH3	27030	Homo sapiens mutL homolog 3 (E. coli) (MLH3), mRNA
MTHFS	10588	Homo sapiens 5,10-methenyltetrahydrofolate synthetase (5-formyltetrahydrofolate cyclo-ligase) (MTHFS), mRNA
NELL2	4753	Homo sapiens NEL-like 2 (chicken) (NELL2), mRNA
PLDN	26258	Homo sapiens pallidin homolog (mouse) (PLDN), mRNA
RPS24	6229	Homo sapiens ribosomal protein S24 (RPS24), transcript variant 1, mRNA
RPS24	6229	Homo sapiens ribosomal protein S24 (RPS24), transcript variant 2, mRNA
SNTB2	6645	Homo sapiens syntrophin, beta 2 (dystrophin-associated protein A1, 59kDa, basic component 2) (SNTB2), transcript variant 1, mRNA
SNTB2	6645	Homo sapiens syntrophin, beta 2 (dystrophin-associated protein A1, 59kDa, basic component 2) (SNTB2), transcript variant 2, mRNA
SNX16	64089	Homo sapiens sorting nexin 16 (SNX16), transcript variant 1, mRNA
SNX16	64089	Homo sapiens sorting nexin 16 (SNX16), transcript variant 2, mRNA
SNX16	64089	Homo sapiens sorting nexin 16 (SNX16), transcript variant 3, mRNA
TNFRSF7	939	Homo sapiens tumor necrosis factor receptor superfamily, member 7 (TNFRSF7), mRNA
ZAK	51776	Homo sapiens sterile alpha motif and leucine zipper containing kinase AZK (ZAK), mRNA

Table 14

Products of the biomarkers noted in **Table 13.** Listed is the Entrez Gene ID along with the human RNA accession number and human protein accession numbers for reference sequence from Genbank.

GeneSymbol	*GeneID*	*RNAAccession*	*ProteinAccession*
CHD2	1106	NM_001271	NP_001262
CSPG6	9126	NM_005445	NP_005436
CTSD	1509	NM_001909	NP_001900
FREB	84824	NM_032738	NP_116127
GAS7	8522	NM_003644	NP_003635
GAS7	8522	NM_201432	NP_958836
GAS7	8522	NM_201433	NP_958839
HIST1H1C	3006	NM_005319	NP_005310
IGFBP7	3490	NM_001553	NP_001544
IRAK3	11213	NM_007199	NP_009130
IREB2	3658	NM_004136	NP_004127
MLH3	27030	NM_014381	NP_055196
MTHFS	10588	NM_006441	NP_006432
NELL2	4753	NM_006159	NP_006150
PLDN	26258	NM_012388	NP_036520
RPS24	6229	NM_001026	NP_001017
RPS24	6229	NM_033022	NP_148982
SNTB2	6645	NM_006750	NP_006741
SNTB2	6645	NM_130845	NP_570896
SNX16	64089	NM_022133	NP_071416
SNX16	64089	NM_152836	NP_690049
SNX16	64089	NM_152837	NP_690050
TNFRSF7	939	NM_001242	NP_001233
ZAK	51776	NM_016653	NP_057737
ZAK	51776	NM_133646	NP_598407

Table 15 Primers utilized for quantitative real-time RT-PCR of biomarkers noted in **Table 13**. The reference sequences to which the primers amplify are shown along with the 5' and 3' primer and the amplicon size.

Primer Pair ID	*RefSeq*	*Sense Primer Seq*	*Sense Primer Start*	*Antisense Primer Seq*	*Antisense Primer Start*	*Amplicon Size*
SEPT6	NM_015129, NM_145799, NM_145800, NM_145802	AGGACAGCTACAAGCCTATCGT	600	ACAAGCAGACATGGATTCGGGA	722	123
CHD2	NM_001271	TGACAAGAAGCCAAAGCGCA	4037	TATGCACTCCAGCCGTTCAAGA	4175	139
CSPG6	NM_005445	ACTCGTGCCAAACTTGATGAGC	794	ACTTGGCGTTCGATATCCTCCA	912	119
CTSD	NM_001909	TGTGGAGGACCTGATTGCCAAA	259	ACTGGGCGTCCATGTAGTTCTT	368	110
FREB_NS	NM_032738	AGCCACTGAGGACAACCAAGT	1019	AGGAGCTGGATTCAATGTGGGA	1121	103
GAS7	NM_003644 /// NM_005890 /// NM_201432 /// NM_201433	AGTTGCTGGAGAAGCCTGGAAT	334	TTCCCAGGTGGTCTCATTGGT	479	146
HIST1H1C	NM_005319	TCTGGTGCAAACGAAAGGCA	276	TGGCTTCTTAGGTTTGGTTCCG	393	118
IGFBP7	NM_001553	TGCGAGCAAGGTCCTTCCATA	488	ATGAGGACAGGTGTCGGGATTC	597	110
IRAK3	NM_007199	ATTGCAGTGTGTAGGTGACACG	833	AGCATGGTTGAACGTTGTGCAG	939	107

IREB2	NM_004136	ACCAGTGCCTGAACCTGAAACA	672	AGGAGGGATCACTGCCACATT	795	124
MLH3	NM_005784 /// NM_014381	ACACGGAGGATGACTTGACTGA	3493	AAACTGGCTTGGCTGGACCTT	3621	129
MTHFS	NM_006441	AGCCTGGTGAGGGTGATGTT	351	CCAGTCGGTTGCCATGTTTGT	452	102
NELL2	NM_006159	CCAGCTGTGAAACGGACATTGA	1742	TCATGGTAGCCATCTCTGCACT	1859	118
PLDN	NM_012388	AGCCGACGCCTGGTTTAAGTGA	296	AGCAATCCTTCTGCCAGTTGCT	395	100
RPS24	NM_001026 /// NM_033022	GCAACGAAAGGAACGCAAGAAC	352	ACTCCTTCGGCTGTGATCCAAT	459	108

Table 16 Results of quantitative real time RT-PCR of biomarkers noted in **Table 13** comparing 20 bladder cancer patients with 14 individuals not having bladder cancer are summarized.

Gene Symbol	Description	RT-PCR Result								
		Ctrl			Cancer			BC/ctrl	direction	P-value
		mean	SD	SD/mean	mean	SD	SD/mean			
SEPT6	septin 6	7.09	0.56	0.08	7.86	0.85	0.11	0.59	↓	0.003
CHD2	chromodomain helicase DNA binding protein 2	7.44	0.36	0.05	8.14	0.75	0.09	0.61	↓	0.001
CSPG6	chondroitin sulfate proteoglycan 6 (bamacan)	8.51	0.48	0.06	9.05	0.63	0.07	0.69	↓	0.008
CTSD	cathepsin D (lysosomal aspartyl protease) /// cathepsin D (lysosomal aspartyl protease)	5.09	0.35	0.07	4.83	0.36	0.07	1.20	↑	0.044
FREB	Fc receptor homolog expressed in B cells	10.61	0.54	0.05	11.85	2.26	0.19	0.42	↓	0.027
GAS7	growth arrest-specific 7	9.85	0.60	0.06	9.67	0.57	0.06	1.13	↑	0.399
HIST1H1C; H1.2; H1F2; MGC:3992	histone 1, H1c /// histone 1, H1c	4.81	0.64	0.13	4.82	0.77	0.16	0.99	---	0.938
IGFBP7	insulin-like growth factor binding protein 7 /// insulin-like growth factor binding protein 7	11.23	0.52	0.05	10.29	0.61	0.06	1.92	↑	0.000
IRAK3	interleukin-1 receptor-associated kinase 3	7.42	0.49	0.07	7.21	0.53	0.07	1.16	↑	0.242
IREB2	iron-responsive element binding protein 2	8.46	0.47	0.06	8.80	0.66	0.08	0.79	↓	0.088
MLH3	mutL homolog 3 (E. coli)	12.53	0.51	0.04	13.77	2.34	0.17	0.42	↓	0.031
MTHFS	5,10-methenyltetrahydrofolate synthetase (5-formyltetrahydrofolate cyclo-ligase) /// 5,10-methenyltetrahydrofolate synthetase (5-formyltetrahydrofolate cyclo-ligase)	7.61	0.59	0.08	8.09	0.45	0.06	0.72	↓	0.018
NELL2	NEL-like 2 (chicken) /// NEL-like 2 (chicken)	9.83	0.83	0.08	10.73	1.55	0.14	0.54	↓	0.037
PLDN	pallidin homolog (mouse)	9.21	0.30	0.03	9.23	0.49	0.05	0.98	---	0.865
RPS24	ribosomal protein S24	5.51	0.71	0.13	5.91	0.84	0.14	0.76	↓	0.145
SNTB2	syntrophin, beta 2 (dystrophin-associated protein A1, 59kDa, basic component 2) /// syntrophin, beta 2 (dystrophin-associated protein A1, 59kDa, basic component 2)	9.83	0.37	0.04	9.95	0.45	0.05	0.92	↓	0.396
SNX16	sorting nexin 16 /// sorting nexin 16	10.89	0.52	0.05	11.27	0.62	0.05	0.77	↓	0.063
TNFRSF7	tumor necrosis factor receptor superfamily, member 7 /// tumor necrosis factor receptor superfamily, member 7	7.81	0.66	0.08	8.59	1.40	0.16	0.58	↓	0.040
ZAK	sterile alpha motif and leucine zipper containing kinase AZK	10.74	0.33	0.03	10.56	0.63	0.06	1.13	↑	0.288

Table 17 Commercially available antibodies to the Protein Products of the Biomarkers of the Invention Noted in **Table 13.**

Gene Symbol	Description	Commercially Available Antibody Reference	Scientific Reference
CHD2			
CSPG6 (SMC3)	Ab9236, Ab1277 and Ab1297 Antihuman Rabbit polyclonal	Abcam Ab9236; Ab1277; Ab1297	See http://www.abcam.com/index.html?datasheet=9263
CTSD (CPSD)	Ab826; Ab 19555; Ab925 Antihuman Rabbit Polyclonal Ab6313; Ab7433 Antihuman mouse monoclonal	Abcam Ab826, Ab1955, Ab6313, Ab7433, Ab925	Henry JA *et al.* Prognostic significance of the estrogen-regulated protein, cathepsin D, in breast cancer. An immunohistochemical study. *Cancer* **65**:265-71 (1990) Garcia M *et al.* Immunohistochemical distribution of the 52-kDa protein in mammary tumors: a marker associated with cell proliferation rather than with hormone responsiveness. *J Steroid Biochem* **27**:439-45 (1987).
FREB			
GAS7	Antihuman chicken polyclonal	GenWay Biotech, Inc. A22099A; A22099B ProteinTechGroup Inc. 10072-1-AP	
HIST1H1C	Antihuman Rabbit polyclonal	Abcam Ab4086; Ab17677	See http://www.abcam.com/index.html?datasheet=4086
IGFBP7			
IRAK3	Antihuman rabbit polyclonal	Abcam Ab8116; ab26319	Wesche H *et al.* IRAK-M is a novel member of the Pelle/interleukin-1 receptor-associated kinase (IRAK) family. *J Biol Chem* **274**:19403-10 (1999).

			Muzio M *et al.* IRAK (Pelle) family member IRAK-2 and MyD88 as proximal mediators of IL-1 signaling. *Science* **278**:1612-5 (1997). Cao Z *et al.* IRAK: a kinase associated with the interleukin-1 receptor. *Science* **271**:1128-31 (1996).
IREB2			
MLH3	Antihuman goat polyclonal	Abcam ab4834	Lipkin SM *et al.* MLH3: a DNA mismatch repair gene associated with mammalian microsatellite instability. *Nat Genet* **24**:27-35 (2000).
MTHFS			
NELL2 (NRP2)	Antihuman chicken polyclonal	Abcam Ab15874	Gluzman-Poltorak Z *et al.* Neuropilin-2 is a receptor for the vascular endothelial growth factor (VEGF) forms VEGF-145 and VEGF-165 [corrected]. *J Biol Chem* **275**:18040-5 (2000).
PLDN	Antihuman goat polyclonal antibody	Abcam Ab10145	Huang L *et al.* The pallid gene encodes a novel, syntaxin 13-interacting protein involved in platelet storage pool deficiency. *Nat Genet* **23**:329-32 (1999).
RPS24			
SNTB2			
SNX16	Antihuman goat polyclonal	Abcam Ab4151;	Teasdale RD *et al.* A large family of endosome-localized proteins related to sorting nexin 1. *Biochem J* **358**:7-16 (2001).
TNFRSF7 (CD27)		Abcam Ab30366; Ab30367; Ab10456; Ab10952; Ab30365	Sembries S *et al.* Reduced expression of adhesion molecules and cell signaling receptors by chronic lymphocytic leukemia cells with 11q deletion. *Blood* **93**:624-31 (1999). Yoshino N *et al.* Upgrading of flow cytometric analysis for absolute counts, cytokines and other antigenic molecules of cynomolgus monkeys (Macaca fascicularis) by using anti-human cross-reactive antibodies. *Exp*

			Anim **49**:97-110 (2000).
ZAK	AP7823b Antihuman goat polyclonal 52-3667 Polyclonal Antibody	Abgent AP7823b Invitrogen 52-3667	

EXAMPLE 5

TaqMan® QRT PCR can also be performed using the QuantiTect™ Probe RT-PCR system from Qiagen (Product Number 204343) in conjunction with a TaqMan® dual labelled probe and primers corresponding to the gene of interest. The TaqMan® probe and primers can be ordered from Applied Biosystems Assays-On-Demand™.

A dual labelled probe contains both a fluorophore and a quencher molecule. The proximity of the fluorescent reporter with the quencher prevents the reporter from fluorescing, but during the PCR extension step, the 5' – 3' exonuclease activity of the Taq DNA polymerase releases the fluorophore which allows it to fluoresce. As such, the amount of fluorescence correlates with the amount of PCR product generated. Examples of TaqMan® probes which can be utilized in accordance with the invention and with the products of the selected biomarkers disclosed in **Table 13** are shown in **Table 18**.

Table 18

Accession	Transcript Length	5'	Start	Primer Length	Right	Start	Primer Length	Probe	Start	Probe Length
NM_001271	7764	AAGACAAAGAAGGGGACAAGG	4932	21	ACAGGTTCACTTCCTGCTGTAA	5067	22	GGTGATGCCAAATCTTCGAGTA	4994	22
NM_001271	7764	CAAAAGCCAAGAGGAGGACA	724	20	TGGATCACTTCCCTGCTCA	841	19	TTCGGGTTCAGACTCAGGCA	790	20
NM_001271	7764	GATGATGACAAGAAGCCAAAGC	4031	22	CGTTCAAGAGGGAGACCAA	4161	19	AAGGACCTCGTGGAGGGATTT	4079	21
NM_001271	7764	AGGAAAGACAAAGAAGGGGACA	4928	22	GACAGGTTCACTTCCTGCTGTAA	5068	23	GTGGTGATGCCAAATCTTCGAG	4992	22
NM_001271	7764	CAAAGCGATCTCAGGGTCCT	5019	20	AGCCCCTTGTCAGGTTTGT	5184	19	CGTGAAAAAGGCACTGAAACAGCT	5140	24
NM_001271	7764	AATCTTCTGAGAGTCAGTCGGA	879	22	CTTGTCTGCTTCGGTTTGAC	1058	20	GTTCCAAATCCCAGCCAGTCCT	921	22
NM_001271	7764	AGAAGATGAACAGGAACAAGGC	1174	22	TAACACGAGGTTTGGGCA	1280	18	AGTAAAAGCCAGAAGACCTGTCC	1228	23
NM_001271	7764	TCAGAAGCCTCATTTGCCT	5515	19	TTGTTGCCACCACCATAGTTG	5664	21	CCCCAACAAGAGACACTTCAGT	5584	22
NM_001271	7764	GGGGAATCGAGTGCTTATCTTC	3124	22	TTTCTCCCTTGATGGAACCA	3239	20	CTCAGATGGTGAGAATGTTGGA	3147	22
NM_001271	7764	CTGGGGATTTATGAACATGGC	4517	21	CTTCAACAAGTAATCCGCTCG	4663	21	TAAAAAGCCTCAGGGGAAGC	4612	20
NM_005445	4077	TCACAAAGCAGTGTCCCATC	3337	20	AGCAAGGGCTACCAAGGATT	3468	20	GAGAAATGCAACAGCTTTCAGG	3419	22
NM_005445	4077	CACATGCGTGGAAGTCACTG	1731	20	GATAGGCTGTATCCCTGACATCTA	1891	24	CCTGGAGAGGTTACTTTTCTGCC	1834	23
NM_005445	4077	AGGGGTTCTGGCTCACAAA	3325	19	TCAGAGCAAGGGCTACCAA	3472	19	AATGCAACAGCTTTCAGGTGG	3423	21
NM_005445	4077	TGCTAGACCACTTCCGTCGA	1634	20	TCTTCGTGCTGACTTCATCTGA	1804	22	ACATGCGTGGAAGTCACTGCT	1732	21
NM_005445	4077	CCGTCGAAAAGGAATAAACCAG	1647	22	TCTTCGTGCTGACTTCATCTG	1804	21	ATGCGTGGAAGTCACTGCTG	1734	20
NM_005445	4077	AGCATGGAAGTTTCAACCCA	1993	20	TCCTTGTGTCATAATAACCCCC	2107	22	GCCCGTGCTTTCACTATGGA	2017	20
NM_005445	4077	GAGAGAGACAGAAGGGGGTACTGT	2640	24	CTGAAGCTCCTTAATTCCAGCT	2784	22	AGACACTATGGCACGATCAGAAG	2712	23
NM_005445	4077	TGAGATTCGTCAACTTCAGCAG	2496	22	TACCCCCTTCTGTCTCTCTCAG	2659	22	ATCTGAGAAAACGCTTGGACC	2594	21
NM_005445	4077	ACAAAGCAGTGTCCCATCAG	3339	20	GGGCTACCAAGGATTTCTGTC	3463	21	GAGAAATGCAACAGCTTTCAGG	3419	22
NM_005445	4077	GTTAGATGTCAGGGATACAGCCTATC	1866	26	ATAGTGAAAGCACGGGCCA	2033	19	TGTCGTAGCATGGAAGTTTCAACC	1987	24
NM_001909	2120	GGTGGCACAGACTCCAAGTATT	874	22	ATGAGGGAAGTGCCTGTGTC	1034	20	CAAGGGTTCTCTGTCCTACCTGA	897	23
NM_001909	2120	TGCTTCACAGTCGTCTTCG	403	19	GCTGGACTTGTCGCTGTTGT	519	20	CCATCCACTGCAAACTGCTGGA	452	22
NM_001909	2120	GGAGGACCTGATTGCCAAAG	261	20	TGTCGAAGACGACTGTGAAGC	424	21	TACTACGGGGAGATTGGCATCG	367	22

NM_001909	2120	GGTGGCACAGACTCCAAGTAT	874	21	CATGAGGGAAGTGCCTGTGT	1035	20	ACAAGGGTTCTCTGTCCTACCTGA	896	24
NM_001909	2120	GCTTCACAGTCGTCTTCGACA	404	21	TAGGTGCTGGACTTGTCGCT	524	20	TCCACTGCAAACTGCTGGACATC	455	23
NM_001909	2120	CGAGGTGCTCAAGAACTACATG	336	22	GATGTCCAGCAGTTTGCAGT	477	20	TCACAGTCGTCTTCGACACG	407	20
NM_001909	2120	AAAGGCCCCGTCTCAAAGTA	277	20	CGTGTCGAAGACGACTGTGA	426	20	TACTACGGGGAGATTGGCATC	367	21
NM_001909	2120	GTGCTTCACAGTCGTCTTCG	402	20	GCCATAGTGGATGTCAAACGA	561	21	TCCCCTCCATCCACTGCAAA	446	20
NM_001909	2120	ATTCCCGAGGTGCTCAAGA	331	19	TCCAGCAGTTTGCAGTGGA	473	19	TCACAGTCGTCTTCGACACG	407	20
NM_001909	2120	AAGGCCCCGTCTCAAAGTA	278	19	CGAAGACGACTGTGAAGCACT	421	21	TTCCCGAGGTGCTCAAGAACT	332	21
NM_032738	2345	ACTTGACTGATGCAAGGGAA	444	20	AAAACTGGCTTGGCTGGAC	561	19	TCAGTGAACCCTTCCACCTGAT	492	22
NM_032738	2345	CACAGTCACCATGAAGCTGG	283	20	CATCAGTCAAGTCATCCTCCG	455	21	GCCAGTTTTGAGACGCTGCA	374	20
NM_032738	2345	CCTCTACCTTTCCCTTGGTGT	325	21	GCATCAGTCAAGTCATCCTCC	456	21	GCCAGTTTTGAGACGCTGCA	374	20
NM_032738	2345	TCACTCCGGGTCATACTGGT	991	20	TCTGAGGAGCTGGATTCAATGT	1124	22	TTGGAAACAGAGCCCCCAGCTA	1039	22
NM_032738	2345	CACACGGAGGATGACTTGACT	431	21	GTCCCCTTCAAAAACTGGCTT	571	21	TCAGTGAACCCTTCCACCTG	492	20
NM_032738	2345	CTCTACCTTTCCCTTGGTGTG	326	21	CTTCCCTTGCATCAGTCAAGT	464	21	GCCAGTTTTGAGACGCTGCA	374	20
NM_032738	2345	ACTTGACTGATGCAAGGGAAG	444	21	AGCAGGTCCCCTTCAAAAACT	576	21	TCAGTGAACCCTTCCACCTG	492	20
NM_032738	2345	TTGTGGCTATCACAGTCCAAG	771	21	TTGTAGAAGGAGAAGAGGAGGC	924	22	ACTGTTTCCAGCGCCAATTCTC	793	22
NM_032738	2345	GGGGTTGCAGGAGACCTAAA	263	20	TGCAGCGTCTCAAAACTGG	393	19	TCACCATGAAGCTGGGCTGTGT	288	22
NM_032738	2345	GAGGATGACTTGACTGATGCA	437	21	GTCCCCTTCAAAAACTGGC	571	19	CAGTGAACCCTTCCACCTGA	493	20
NM_003644	7773	GCTGAGCAACAAGACAGAGGAG	802	22	ATTTGGACTGGGCCTGGTT	911	19	ACCTCATGCGCTGTGTGGAT	867	20
NM_003644	7773	TCGCCAAGCAAAAAGCAG	209	18	CCTTGGGGGTCCTTCTTATC	360	20	AGCTGCTGAAACCAACCGAGT	294	21
NM_003644	7773	TGATAAGAAGGACCCCCAAG	340	20	AAGCCAAGGAGTTCTGAGAGAG	515	22	TGGCTGGGTTTGAACTACTGC	372	21
NM_003644	7773	GGAGAACAGCTTTGACGATGT	154	21	CACTCGGTTGGTTTCAGCAG	315	20	GAGCAAGGAAAACACCATCACA	226	22
NM_003644	7773	AGAACTCCTTGGCTTCACAGG	501	21	GGAAGTTCATCAGGGGCTT	644	19	TCACCTCAAGTTCTCTGCCAA	586	21
NM_003644	7773	TCATGCGCTGTGTGGATCT	870	19	ATCATCTCTACCCTCTCCACCT	978	22	GCCCAGTCCAAATGGTTTGA	899	20
NM_003644	7773	GAGCAAGGAAAACACCATCAC	226	21	CAGTAGTTCAAACCCAGCCA	391	20	AGCTGCTGAAACCAACCGAGT	294	21
NM_003644	7773	AAGAAGTGCGACCACCACATT	668	21	TGTCCTCCTCTGTCTTGTTGC	827	21	CTGGAGATGAAGACCCAGCA	770	20
NM_003644	7773	GCTGAGCAACAAGACAGAGGAG	802	22	CTCTTCAAACCATTTGGACTGG	922	22	TCATGCGCTGTGTGGATCTC	870	20
NM_003644	7773	CTGAAACCAACCGAGTGGA	299	19	TCCTTCTGCATTTGTTTGCC	429	20	ACCGTGGCTGGGTTTGAACTAC	368	22
NM_201432	8181	TCACCTCAAGTTCTCTGCCA	994	20	ATGTGGTGGTCGCACTTCT	1095	19	TTCACAGCGAGGTGGAGAAG	1017	20
NM_201432	8181	TCAAGCTGAGCAACAAGACAG	1206	21	TCAAACCATTTGGACTGGG	1326	19	ACCTCATGCGCTGTGTGGAT	1275	20
NM_201432	8181	TGAGTGTCCGAAAATCCACC	570	20	TCCACTCGGTTGGTTTCAG	725	19	TCGCCAAGCAAAAAGCAGAG	617	20
NM_201432	8181	TGCGCTGTGTGGATCTCTA	1281	19	GGATCATCTCTACCCTCTCCA	1388	21	AGGCCCAGTCCAAATGGTTTGAAGA	1305	25
NM_201432	8181	CTGAAACCAACCGAGTGGAG	707	20	CATTTGTTTGCCCTTCAGCT	829	20	GGGTTTGAACTACTGCTCCAGA	785	22
NM_201432	8181	CAGAGCAAGGAAAACACCATC	632	21	CTGGAGCAGTAGTTCAAACCC	805	21	AGCTGCTGAAACCAACCGAGT	702	21
NM_201432	8181	AAGTGCGACCACCACATTG	1079	19	GTCCTCCTCTGTCTTGTTGCT	1234	21	GACCTGGAGATGAAGACCCA	1175	20

NM_201432	8181	CAGAGCAAGGAAAACACCATC	632	21	CCTTGGGGGTCCTTCTTATC	768	20	AGCTGCTGAAACCAACCGAGT	702	21
NM_201432	8181	GCAGAGCAAGGAAAACACCA	631	20	GGAGCAGTAGTTCAAACCCAG	803	21	AGCTGCTGAAACCAACCGAGT	702	21
NM_201432	8181	AACAAGACAGAGGAGGACATCA	1217	22	TGCTGCCGGATCATCTCTA	1395	19	TCTACAACCAGGCCCAGTCCAAA	1296	23
NM_201433	8211	GCAAGGAAAACACCATCACA	666	20	GGAGCAGTAGTTCAAACCCAG	833	21	GCTGCTGAAACCAACCGAGT	733	20
NM_201433	8211	CTGAAACCAACCGAGTGGA	737	19	CATTTGTTTGCCCTTCAGCT	859	20	TGGCTGGGTTTGAACTACTGCTCC	810	24
NM_201433	8211	TCAAGCTGAGCAACAAGACAG	1236	21	CTCTTCAAACCATTTGGACTGG	1360	22	TCATGCGCTGTGTGGATCTC	1308	20
NM_201433	8211	GAGCAAGGAAAACACCATCA	664	20	GTTTCTGGAGCAGTAGTTCAAACC	839	24	GCTGCTGAAACCAACCGAGT	733	20
NM_201433	8211	AACAAGACAGAGGAGGACATCA	1247	22	AACCATTTGGACTGGGCCT	1353	19	ACCTCATGCGCTGTGTGGAT	1305	20
NM_201433	8211	CACCTCAAGTTCTCTGCCAA	1025	20	ATGTGGTGGTCGCACTTCTT	1125	20	TTCACAGCGAGGTGGAGAAG	1047	20
NM_201433	8211	ACAAGACAGAGGAGGACATCAAG	1248	23	TGCTGCCGGATCATCTCTAC	1425	20	ACCTCATGCGCTGTGTGGATCT	1305	22
NM_201433	8211	TGAGTGTCCGAAAATCCACC	600	20	AGAAGTAGTCGCAGTAGCTCCACT	773	24	CAGAGCAAGGAAAACACCATCA	662	22
NM_201433	8211	ACCTGGAGATGAAGACCCAG	1206	20	CAAACCATTTGGACTGGGC	1355	19	ACCTCATGCGCTGTGTGGAT	1305	20
NM_201433	8211	ATGAAGAAGTGCGACCACCA	1103	20	ATGTCCTCCTCTGTCTTGTTGCT	1266	23	ATTGCCGACCTTCGCAAGCA	1124	20
NM_001553	1108	GAGCTGTGAGGTCATCGGAAT	558	21	GCACCCAGCCAGTTACTTCA	721	20	GGGGTCACTATGGAGTTCAAAGG	614	23
NM_001553	1108	CAGGTGTACTTGAGCTGTGAGGT	547	23	GCACCCAGCCAGTTACTTCA	721	20	GGGGTCACTATGGAGTTCAAAGG	614	23
NM_001553	1108	ACAGAACTCCTGCCTGGTGA	637	20	CCTTGGGAATTGGATGCA	785	18	AAGTAACTGGCTGGGTGCTG	704	20
NM_001553	1108	TACTTGAGCTGTGAGGTCATCG	553	22	CCAGCCAGTTACTTCATGCTT	717	21	GGGGTCACTATGGAGTTCAAAGG	614	23
NM_001553	1108	GCATGAAGTAACTGGCTGGGT	699	21	CATGTAAGGCATCAACCACTG	832	21	TGCATCCAATTCCCAAGGACAGG	768	23
NM_001553	1108	AGCATGAAGTAACTGGCTGGG	698	21	TGATGCTGAAGCCTGTCCT	801	19	CCTCTAAGTAAGGAAGATGCTGGA	730	24
NM_001553	1108	GAAGTAACTGGCTGGGTGCT	703	20	CATGTAAGGCATCAACCACTG	832	21	GCATCCAATTCCCAAGGACAGG	769	22
NM_001553	1108	GCATGAAGTAACTGGCTGGG	699	20	CTGATGCTGAAGCCTGTCCT	802	20	CCTCTAAGTAAGGAAGATGCTGGA	730	24
NM_001553	1108	AAAGCATGAAGTAACTGGCTGG	696	22	TGCTGATGCTGAAGCCTGT	804	19	CCTCTAAGTAAGGAAGATGCTGGA	730	24
NM_001553	1108	GGCCCAGAAAAGCATGAAGT	688	20	CTGATGCTGAAGCCTGTCCT	802	20	CCTCTAAGTAAGGAAGATGCTGGA	730	24
NM_007199	2288	CTGCCAAGCTCTTCTGTTTG	1306	20	GAGAAGGACACCTGAAGGACTTT	1464	23	CGGGCAAAGTTAAGACCATCAA	1347	22
NM_007199	2288	TCGGTCATCTGTGGCAGTATA	924	21	TGCTGCTGCTGGTCATATTT	1065	20	ACTTCCGGTCCCACCTAGAA	1009	20
NM_007199	2288	TTCAACCATGCTCGGTCA	913	18	TTTACTGCTGCTGCTGGTCA	1070	20	ACTTCCGGTCCCACCTAGAA	1009	20
NM_007199	2288	AGCCATTCACTACCTGCACAA	890	21	ACAACTCTGATGTTCTAGGTGGGA	1040	24	GCTCGGTCATCTGTGGCAGTAT	922	22
NM_007199	2288	TTCAACCATGCTCGGTCA	913	18	TGTTTACTGCTGCTGCTGG	1072	19	ACTTCCGGTCCCACCTAGAA	1009	20
NM_007199	2288	GCGGGCAAAGTTAAGACCA	1346	19	CCAGGAATAGAGGAGAAGGACA	1476	22	GCCAGCTTGTATTTTGCTGAAG	1404	22
NM_007199	2288	ACCATCGGTGACCTTTTACAG	306	21	TTATCCACGGTGACATTGGC	469	20	GATGGGACATCGTCGAGCTA	338	20
NM_007199	2288	GCCAAGCTCTTCTGTTTGG	1308	19	TGGTACATTCTCCAGGAATAGAGG	1487	24	GCCAGCTTGTATTTTGCTGAAGATCC	1404	26

NM_007199	2288	GCCATTCACTACCTGCACAA	891	20	CAACTCTGATGTTCTAGGTGGGA	1039	23	CTCGGTCATCTGTGGCAGTAT	923	21
NM_007199	2288	AACCATGCTCGGTCATCTG	916	19	ATGTTTACTGCTGCTGCTGGT	1073	21	ACTTCCGGTCCCACCTAGAA	1009	20
NM_004136	3928	ACCTGCATGATATTTGGCCT	2032	20	TGAATCCGGTGCTTCTAAGG	2171	20	TGGGGAATAAACGGTGGAAT	2131	20
NM_004136	3928	TGGAATTGGCATAGCTCCAC	2705	20	AACGAAGCAATCACGCTGAA	2872	20	CCTGAAGAACTGTCTCCTGGAA	2796	22
NM_004136	3928	TGCAATCCATCTGTCATGC	1599	19	CCCACTGCCTGGAGATAAACT	1706	21	GTTGAAGCTGGTCTGCGTGTT	1647	21
NM_004136	3928	GCAAAGCCAAACTCGAATC	1255	19	TCTATCCTGAGGTCTTTTTGGACC	1415	24	CAGGAGAACCTGAATACTCCCAG	1327	23
NM_004136	3928	TGGGGAATAAACGGTGGAA	2131	19	TTCAATAGCCTGGAGTGCAA	2267	20	AAGCACCGGATTCAGTTTTG	2158	20
NM_004136	3928	AGGAAACTCCAGAGACTGGG	2606	20	ACTGAAGTGGAGCTATGCCAAT	2730	22	GTGTGAAAGCTGTTTTGGCC	2650	20
NM_004136	3928	TGTTGAAGCTGGTCTGCGT	1646	19	CATCCATAGCCAACGATTTC	1786	20	TTATCTCCAGGCAGTGGGATGGTT	1689	24
NM_004136	3928	CACTTCAGTTCCTTCCAGGAGA	2722	22	CGAAGCAATCACGCTGAAT	2870	19	CCTGAAGAACTGTCTCCTGGAAT	2796	23
NM_004136	3928	GTTATGCTTGGTCTGCCAGTTT	963	22	ACTTTCCAGCCACTCCTACTTG	1107	22	TCTTACTTTACCAGAGGTGGTTGG	986	24
NM_004136	3928	CTGCCCGTGTTCTTCTTCA	298	19	TTTCTCAGGATCACCTCCAAGA	398	22	TTTGCTGCTATGAGGGAGGCAG	348	22
NM_014381	4889	CCCCAACTGAGGACATTCAG	3319	20	TCTCGGAAGGAAAGGAAGAA	3452	20	AAAGACCTGACAACTGTGGCTGTG	3351	24
NM_014381	4889	TTGCCACTGACTGTCCAGA	4086	19	CTTCAATAAGGCGGCAACTT	4188	20	TAATGATGGCCTGAGCTTACAGG	4145	23
NM_014381	4889	ACCTTCTATGCTGCCGTTAGCT	4232	22	CTGCCTTGTATCACACTCTGCT	4364	22	ACTTCGCAAAATGGCCCAGG	4304	20
NM_014381	4889	GTCCTTGTGGGAAAAGTACCAC	3936	22	TTCTGGACAGTCAGTGGCAAT	4105	21	GCCAATGAACTTCGGAGAGG	3978	20
NM_014381	4889	CCTTCTATGCTGCCGTTAGC	4233	20	GCCTTGTATCACACTCTGCTTTTC	4362	24	ACTTCGCAAAATGGCCCAGG	4304	20
NM_014381	4889	CTCAGAATGGGACAATCCAGT	3539	21	CCTTTGGTGAAACGATAGGGAT	3664	22	GCTGTTGATGTAAGCAGTGGC	3582	21
NM_014381	4889	AAGCGACCTTGTTCTTCCTT	3422	20	ACTGGATTGTCCCATTCTGAG	3559	21	TCGAGCAGAGAGGACTGTGATGA	3455	23
NM_014381	4889	CCTTGTTCTTCCTTTCCTTCC	3428	21	GGCAAATACTGGATTGTCCCA	3566	21	TCGAGCAGAGAGGACTGTGATGA	3455	23
NM_014381	4889	TTGTTCTTCCTTTCCTTCCGAG	3430	22	GGCCACTGCTTACATCAACAG	3603	21	TCGAGCAGAGAGGACTGTGATGA	3455	23
NM_014381	4889	ATTGCCACTGACTGTCCAGAAG	4085	22	CTTCAATAAGGCGGCAACTT	4188	20	TGATGGCCTGAGCTTACAGG	4148	20
NM_006441	857	GTTCCAGAGCAATCACATGG	264	20	CCTGGCATGAAGATGAGATC	419	20	CCCAAAACATCCTGGAATATCC	325	22
NM_006441	857	GAATATCCCTCAGCCTGGTG	339	20	TCAGATAGGCATCATAGTAGCCCT	484	24	TGACAAACATGGCAACCGAC	429	20
NM_006441	857	GAATATCCCTCAGCCTGGTG	339	20	CAGTCGGTTGCCATGTTTGT	450	20	CTTGATCTCATCTTCATGCCAGG	397	23
NM_006441	857	GAATATCCCTCAGCCTGGTG	339	20	AGGGCTTCACTTCCTGATGC	517	20	GACAAACATGGCAACCGACTGG	430	22
NM_006441	857	CGGTTCCAGAGCAATCACA	262	19	CCTGGCATGAAGATGAGATCA	419	21	CCCAAAACATCCTGGAATATCC	325	22
NM_006441	857	TTCCAGAGCAATCACATGG	265	19	CCTGGCATGAAGATGAGATC	419	20	CCCAAAACATCCTGGAATATCC	325	22
NM_006441	857	AGCCTGGTGAGGGTGATGTT	350	20	AGTCGGTTGCCATGTTTGT	449	19	CTTGATCTCATCTTCATGCCAGGCC	397	25
NM_006441	857	AGCCTGGTGAGGGTGATGTT	350	20	AGATAGGCATCATAGTAGCCCTTG	482	24	CTTGATCTCATCTTCATGCCAGGCC	397	25
NM_006441	857	AGCCTGGTGAGGGTGATGTT	350	20	CATCATAGTAGCCCTTGCCC	475	20	ACAAACATGGCAACCGACTGGG	431	22

NM_006441	857	AGCCTGGTGAGGGTGATGTT	350	20	CAGTCGGTTGCCATGTTTG	450	19	TTGATCTCATCTTCATGCCAGGCC	398	24
NM_006159	3196	TACAGGGAATGGAACGACATG	1634	21	GCACGACTGTCACATTGAACA	1801	21	GCCTGTATTGCCGCTAATGTGTGT	1692	24
NM_006159	3196	TGACCCTAAAACAGACCCACTT	364	22	CCACTTGTCATCAGCCAAA	542	19	GGCCATCGGAATGAAGTCAGA	450	21
NM_006159	3196	TAGCCAAAACATCAGCCAAG	835	20	TATCCAGGACTCAAATTCTCGG	956	22	CACCATGAAGGGAACCACCT	914	20
NM_006159	3196	GTACCACTGTGAGTGCAGAGATG	1826	23	TTCTTTCCATGAGGACATCG	1996	20	ACCAAGTGGAGAATCGTGTGAA	1874	22
NM_006159	3196	CACCAAGTGGAGAATCGTGT	1873	20	AGTCCCCTGTGCAATTCTTT	2010	20	TATTGATGAGTGTGGGACCGGGA	1898	23
NM_006159	3196	TGTGCAGAAAATCATGGAGC	803	20	GCAGTTCTTACAGCCGTCTATCC	974	23	CACCATGAAGGGAACCACCT	914	20
NM_006159	3196	ATGACAGGTGCTCTGTGTGC	2062	20	GCACTGACTACTAAGCCTTGGGT	2192	23	GGATGGTCTGTGACTGTGAGAATC	2113	24
NM_006159	3196	ATGACAAGTGGCACAAGCTC	532	20	CCTAGAGGCAAGTCTGTGGA	649	20	AGCCATCAGTGCTTCCCATTT	557	21
NM_006159	3196	TGGTCAGATTTGGGTGTTGG	2039	20	ACTACTAAGCCTTGGGTCACATTC	2186	24	ATGGTCTGTGACTGTGAGAATCC	2115	23
NM_006159	3196	GAACCACCTACCGAGAATTTG	925	21	TAAGTGGGCAGTCAGGATTTG	1038	21	GGCTGTAAGAACTGCACATGCC	960	22
NM_012388	3941	GGCTAAACACTATCATGCCAAG	545	22	CAAACTCCTTCTCTCGTTGCTG	699	22	TGCAGCAGAAGAGGCAAAAAG	640	21
NM_012388	3941	GGCTAAACACTATCATGCCAAG	545	22	TTGCTGCTCCCTTTCCAA	683	18	TGCAGCAGAAGAGGCAAAAAGAAG	640	24
NM_012388	3941	GGCTAAACACTATCATGCCAAG	545	22	CTGCTCCCTTTCCAACTCTT	680	20	TGCAGCAGAAGAGGCAAAAA	640	20
NM_012388	3941	GGCTAAACACTATCATGCCAAG	545	22	TTGCTGCTCCCTTTCCAAC	683	19	TGCAGCAGAAGAGGCAAAAAGAAG	640	24
NM_012388	3941	GGCTAAACACTATCATGCCAAG	545	22	ACTCCTTCTCTCGTTGCTGC	696	20	TGCAGCAGAAGAGGCAAAAAG	640	21
NM_012388	3941	GGCTAAACACTATCATGCCAAG	545	22	TCCTTCTCTCGTTGCTGCT	694	19	TGCAGCAGAAGAGGCAAAAAG	640	21
NM_012388	3941	GGCTAAACACTATCATGCCAAG	545	22	TTCAAACTCCTTCTCTCGTTGC	701	22	GCAGCAGAAGAGGCAAAAAGAAG	641	23
NM_012388	3941	GGCTAAACACTATCATGCCAAG	545	22	TGCTGCTCCCTTTCCAACT	682	19	TGCAGCAGAAGAGGCAAAAAGA	640	22
NM_012388	3941	GGCTAAACACTATCATGCCAAG	545	22	GTTGCTGCTCCCTTTCCAA	684	19	CAGCAGAAGAGGCAAAAAGAAG	642	22
NM_012388	3941	GGCTAAACACTATCATGCCAAG	545	22	TCAAACTCCTTCTCTCGTTGC	700	21	CAGCAGAAGAGGCAAAAAGAAG	642	22
NM_001026	515	GACCTCAAGAAAGCAACGAA	339	20	CATTGCAGCACCTTTACTCCTT	451	22	AGAAAGTCAGGGGGACTGCAAAG	380	23
NM_001026	515	GACCTCAAGAAAGCAACGAA	339	20	TGCAGCACCTTTACTCCTTC	448	20	GAAAGTCAGGGGGACTGCAAAG	381	22
NM_001026	515	GACCTCAAGAAAGCAACGAA	339	20	CATTGCAGCACCTTTACTCCT	451	21	AAAGTCAGGGGGACTGCAAA	382	20
NM_001026	515	GACCTCAAGAAAGCAACGAA	339	20	CATTGCAGCACCTTTACTCC	451	20	CAATGTTGGTGCTGGCAAAA	405	20
NM_001026	515	AGAAAGCAACGAAAGGAACG	396	20	GCCACAGCTAACATCATTGC	515	20	AATGAAGAAAGTCAGGGGGACT	425	22
NM_001026	515	GCGACAGTGCCTAAGACAGA	183	20	ATCATGCCAAAGCCAGTTGT	310	20	CAGAACTCATTTTGGTGGTGG	266	21
NM_001026	515	AACGAAAGGAACGCAAGAAC	403	20	GCCACAGCTAACATCATTGC	515	20	AATGAAGAAAGTCAGGGGGACT	425	22
NM_001026	515	AGAAAGCAACGAAAGGAACG	396	20	GGCCACAGCTAACATCATTG	516	20	AATGAAGAAAGTCAGGGGGACT	425	22
NM_001026	515	AACGAAAGGAACGCAAGAAC	403	20	GGCCACAGCTAACATCATTG	516	20	AATGAAGAAAGTCAGGGGGACT	425	22
NM_001026	515	ACGAAAGGAACGCAAGAACA	404	20	GCCACAGCTAACATCATTGC	515	20	AATGAAGAAAGTCAGGGGGACT	425	22

NM_033022	537	CTCATTTTGGTGGTGGCAAG	221	20	TGTTCTTGCGTTCCTTTCG	373	19	CAACTGGCTTTGGCATGATT	242	20
NM_033022	537	CAACTGGCTTTGGCATGATT	242	20	CTTTGCAGTCCCCCTGACTT	402	20	ACGAAAGGAACGCAAGAACAG	354	21
NM_033022	537	CATTTTGGTGGTGGCAAGAC	223	20	CTTTGCAGTCCCCCTGACTT	402	20	CGAAAGGAACGCAAGAACAG	355	20
NM_033022	537	AGACAACTGGCTTTGGCAT	239	19	AGTCCCCCTGACTTTCTTCATT	396	22	ACGAAAGGAACGCAAGAACA	354	20
NM_033022	537	CAGAACTCATTTTGGTGGTGG	216	21	GTTCTTGCGTTCCTTTCGTT	372	20	GACAACTGGCTTTGGCATGATT	240	22
NM_033022	537	GAATGAAGAAAGTCAGGGGG	374	20	CATTGCAGCACCTTTACTCCTT	473	22	CAATGTTGGTGCTGGCAAAA	405	20
NM_033022	537	CTCATTTTGGTGGTGGCA	221	18	GTCCCCCTGACTTTCTTCATT	395	21	ACAACTGGCTTTGGCATGAT	241	20
NM_033022	537	AGAACTCATTTTGGTGGTGGC	217	21	TCTGTTCTTGCGTTCCTTTC	375	20	GACAACTGGCTTTGGCATGATT	240	22
NM_033022	537	ACAACTGGCTTTGGCATGA	241	19	GTCCCCCTGACTTTCTTCATT	395	21	ACGAAAGGAACGCAAGAACA	354	20
NM_033022	537	TTTGGTGGTGGCAAGACAA	226	19	TTTGCAGTCCCCCTGACTTT	401	20	CGAAAGGAACGCAAGAACAG	355	20
NM_006750	1700	ATGCCGTGGACAAGAGATG	1106	19	GCCTGTCCTGGTAGCAAATG	1240	20	ATGCCACAGCTACCCACTTGTTGC	1138	24
NM_006750	1700	TGTATGCCGTGGACAAGAGAT	1103	21	CCTGTCCTGGTAGCAAATGTAAG	1239	23	TGCCACAGCTACCCACTTGTTG	1139	22
NM_006750	1700	CTGAACTCAACGCCATGCTT	942	20	TCACAGCCATGAGGACAGGT	1076	20	GTGAAGCATATTGCCTGGCT	995	20
NM_006750	1700	GACTGAGAAGGATTTGCTGCT	1075	21	ATGTAAGGTCAGATCCAAGGGA	1223	22	CATGCCACAGCTACCCACTTGT	1137	22
NM_006750	1700	ACTGTATGCCGTGGACAAGA	1101	20	AAGGTCAGATCCAAGGGAGG	1219	20	CATGCCACAGCTACCCACTT	1137	20
NM_006750	1700	GAGGTGAAGCATATTGCCTG	992	20	CTCTTGTCCACGGCATACAGT	1121	21	GATGGTGGAAGACAGCAATGGAG	1034	23
NM_006750	1700	CAGCACCAAGGACAGGAAGAT	733	21	ATCTTTGCAGCGTAGGATCA	871	20	CCTCTCAAAATGTGCTTTGCTG	758	22
NM_006750	1700	GACTGTATGCCGTGGACAAGA	1100	21	GAGCCTGTCCTGGTAGCAAAT	1242	21	TGCCACAGCTACCCACTTGT	1139	20
NM_006750	1700	GTGAAGCATATTGCCTGGCT	995	20	TCTCTTGTCCACGGCATACAG	1122	21	ATGGTGGAAGACAGCAATGGA	1035	21
NM_006750	1700	CCAAAACACCAGAACAGCA	719	19	GCAGCGTAGGATCAACGTGT	865	20	CCCTCTCAAAATGTGCTTTGCTGC	757	24
NM_130845	1489	ACAGCTACCCACTTGTTGCC	932	20	GGATGACAGATCCCGATGT	1089	19	ATTTGCTACCAGGACAGGCTC	1011	21
NM_130845	1489	TAGTGGCAGTGAGGACTCTGGT	694	22	CATTGCTGTCTTCCACCATC	842	20	TCCCTCTCAAAATGTGCTTTGC	756	22
NM_130845	1489	ATGTGCTTTGCTGCTAGAAACC	767	22	CATCTCTTGTCCACGGCATA	913	20	TGACTGAGAAGGATTTGCTGC	863	21
NM_130845	1489	ACTGTATGCCGTGGACAAGAG	890	21	AGAGCCTGTCCTGGTAGCAAA	1032	21	TGCCACAGCTACCCACTTGT	928	20
NM_130845	1489	CAGGATACTTGTTCAGGGTTGC	1095	22	TTCCCTTGAGATGGTGAACC	1224	20	CCAAGAGGTGAGGCTTACTATTCA	1170	24
NM_130845	1489	GGATACTTGTTCAGGGTTGCC	1097	21	GCCTCCATTTTCCCTTGA	1233	18	CCAAGAGGTGAGGCTTACTATTCAC	1170	25
NM_130845	1489	TCTCTTCAGGGTGGAGACACA	1053	21	CCTTGAGATGGTGAACCCATT	1221	21	CAGGATACTTGTTCAGGGTTGCC	1095	23
NM_130845	1489	GGATACTTGTTCAGGGTTGCC	1097	21	AGCCTCCATTTTCCCTTGAG	1234	20	CCAAGAGGTGAGGCTTACTATTCA	1170	24
NM_130845	1489	TCTCTTCAGGGTGGAGACACA	1053	21	TGAATAGTAAGCCTCACCTCTTGG	1193	24	AGGATACTTGTTCAGGGTTGCCA	1096	23
NM_130845	1489	TGACTGTATGCCGTGGACAA	888	20	CTGTCCTGGTAGCAAATGTAAGG	1027	23	TGCCACAGCTACCCACTTGTTG	928	22
NM_022133	3289	CTGGGTTATGAAGTGATGGAAG	678	22	AACCAGCGTTTTGGAGGA	856	18	GTTTCCAGGTTTTCGACTAGCA	815	22
NM_022133	3289	CTGGGTTATGAAGTGATGGAAG	678	22	AAACCAGCGTTTTGGAGG	857	18	GTTTCCAGGTTTTCGACTAGCA	815	22
NM_022133	3289	CTGGGTTATGAAGTGATGGAAG	678	22	ACCAGCGTTTTGGAGGAAG	855	19	TGTTTCCAGGTTTTCGACTAGC	814	22

NM_022133	3289	CTGGGTTATGAAGTGATGGAAG	678	22	AAACCAGCGTTTTGGAGGAA	857	20	TGTTTCCAGGTTTTCGACTAGC	814	22
NM_022133	3289	CTGGGTTATGAAGTGATGGAAG	678	22	AACCAGCGTTTTGGAGGAAG	856	20	TGTTTCCAGGTTTTCGACTAGC	814	22
NM_022133	3289	CTGGGTTATGAAGTGATGGAAG	678	22	AAACCAGCGTTTTGGAGGA	857	19	GTTTCCAGGTTTTCGACTAGCA	815	22
NM_022133	3289	CTGGGTTATGAAGTGATGGAAG	678	22	AACCAGCGTTTTGGAGGAA	856	19	GTTTCCAGGTTTTCGACTAGCA	815	22
NM_022133	3289	CTGGGTTATGAAGTGATGGAAG	678	22	ACCAGCGTTTTGGAGGAA	855	18	GTTTCCAGGTTTTCGACTAGCA	815	22
NM_022133	3289	AGCCAAAAGAAGATGGCAAC	164	20	TTGAGCTTGTTGAGACACTGC	287	21	CCATAGGAAACTCTGCTTCCA GT	207	23
NM_022133	3289	TTGGCAGTGTCTCAACAAGC	431	20	ATGAGGGGACTGCTACAGACA	560	21	AGGGCCAGTTAGAAGACTCAA A	464	22
NM_152836	3122	CTGGGTTATGAAGTGATGGAAG	511	22	AAACCAGCGTTTTGGAGG	690	18	TGTTTCCAGGTTTTCGACTAGC	647	22
NM_152836	3122	CTGGGTTATGAAGTGATGGAAG	511	22	AACCAGCGTTTTGGAGGAA	689	19	TGTTTCCAGGTTTTCGACTAGC	647	22
NM_152836	3122	CTGGGTTATGAAGTGATGGAAG	511	22	ACCAGCGTTTTGGAGGAA	688	18	GTTTCCAGGTTTTCGACTAGCA	648	22
NM_152836	3122	CTGGGTTATGAAGTGATGGAAG	511	22	AAACCAGCGTTTTGGAGGAA	690	20	TGTTTCCAGGTTTTCGACTAGC	647	22
NM_152836	3122	CTGGGTTATGAAGTGATGGAAG	511	22	AAACCAGCGTTTTGGAGGA	690	19	TGTTTCCAGGTTTTCGACTAGC	647	22
NM_152836	3122	CTGGGTTATGAAGTGATGGAAG	511	22	AACCAGCGTTTTGGAGGAAG	689	20	TGTTTCCAGGTTTTCGACTAGC	647	22
NM_152836	3122	CTGGGTTATGAAGTGATGGAAG	511	22	AACCAGCGTTTTGGAGGA	689	18	GTTTCCAGGTTTTCGACTAGCA	648	22
NM_152836	3122	CTGGGTTATGAAGTGATGGAAG	511	22	ACCAGCGTTTTGGAGGAAG	688	19	TGTTTCCAGGTTTTCGACTAGC	647	22
NM_152836	3122	TTGGCAGTGTCTCAACAAGC	264	20	ATGAGGGGACTGCTACAGACA	393	21	AGGGCCAGTTAGAAGACTCAA A	297	22
NM_152836	3122	CCCAGAAGAAAGCTGGGTAG	577	20	GCGTTTTGGAGGAAGTGCTA	685	20	GAGATGTTTCCAGGTTTTCGAC	644	22
NM_152837	3035	CGGAAACAGTGAATTGGGAAG	467	21	TTAAACCAGCGTTTTGGAGG	605	20	TGTTTCCAGGTTTTCGACTAGC	560	22
NM_152837	3035	GACACTGAAGAACAAAATCCGG	448	22	TTAAACCAGCGTTTTGGAGG	605	20	GTTTCCAGGTTTTCGACTAGCA	561	22
NM_152837	3035	CGGAAACAGTGAATTGGGAAG	467	21	TAAACCAGCGTTTTGGAGGA	604	20	GTTTCCAGGTTTTCGACTAGCA	561	22
NM_152837	3035	CGGAAACAGTGAATTGGGAAG	467	21	TAAACCAGCGTTTTGGAGG	604	19	TGTTTCCAGGTTTTCGACTAGC	560	22
NM_152837	3035	CGGAAACAGTGAATTGGGAAG	467	21	AACCAGCGTTTTGGAGGAAG	602	20	TGTTTCCAGGTTTTCGACTAGC	560	22
NM_152837	3035	CGGAAACAGTGAATTGGGAAG	467	21	AAACCAGCGTTTTGGAGGAA	603	20	GTTTCCAGGTTTTCGACTAGCA	561	22
NM_152837	3035	CACTGAAGAACAAAATCCGG	450	20	AAACCAGCGTTTTGGAGG	603	18	GTTTCCAGGTTTTCGACTAGCA	561	22
NM_152837	3035	GACACTGAAGAACAAAATCCGG	448	22	AAACCAGCGTTTTGGAGG	603	18	TGTTTCCAGGTTTTCGACTAGC	560	22
NM_152837	3035	CGGAAACAGTGAATTGGGAAG	467	21	ACCAGCGTTTTGGAGGAAG	601	19	TGTTTCCAGGTTTTCGACTAGC	560	22
NM_152837	3035	GACACTGAAGAACAAAATCCGG	448	22	TAAACCAGCGTTTTGGAGGA	604	20	GTTTCCAGGTTTTCGACTAGCA	561	22
NM_001242	1305	TGTGATCCTTGCATACCGG	379	19	ATTGGCAGTGATGGTGCAG	498	19	AGCTGTCGGCACTGTAACTCT G	442	22
NM_001242	1305	AGAAAGGCTGCTCAGTGTGA	364	20	ATTGGCAGTGATGGTGCAG	498	19	AGAGCTGTCGGCACTGTAACT C	440	22
NM_001242	1305	GTTCCTTGTTTTCACCCTGG	792	20	CCGGTTTTCGGTAATCCTC	958	19	CCTGTTCCTCCATCAACGAAG	819	21
NM_001242	1305	AGCTGTCGGCACTGTAACTC	442	20	TCAGCGAAGGGTTTGGAA	580	18	CAACTGCACCATCACTGCCA	477	20
NM_001242	1305	CTGCTCAGTGTGATCCTTGC	371	20	TGATGGTGCAGTTGCGAA	490	18	AGCTGTCGGCACTGTAACTCT G	442	22
NM_001242	1305	AGAGCTGTCGGCACTGTAACT	440	21	CAGCGAAGGGTTTGGAAGA	579	19	AACTGCACCATCACTGCCAAT GC	478	23
NM_001242	1305	AGAAAGGCTGCTCAGTGTGAT	364	21	ATTGGCAGTGATGGTGCA	498	18	AGAGCTGTCGGCACTGTAACT CT	440	23

NM_001242	1305	AGCCCACCCACTTACCTTATG	617	21	AGAAGATCACAAGGATGCGA	784	20	AAAGATCCCTGTGCAGCTCCGAT	737	23
NM_001242	1305	CAGCATAGAAAGGCTGCTCAG	358	21	ATTGGCAGTGATGGTGCA	498	18	AGAGCTGTCGGCACTGTAACTCT	440	23
NM_001242	1305	GCTCAGTGTGATCCTTGCA	373	19	TGGCAGTGATGGTGCAGTT	496	19	AGAGCTGTCGGCACTGTAACTCTG	440	24
NM_016653	3699	CCTCTCGGTTCCATAACCATA	494	21	TCCCTTGTTAGCATCTCCCAG	641	21	GAAGTTATCCAGAGTCTCCCTGTGTCAGA	559	29
NM_016653	3699	CCAGAAGTTATCCAGAGTCTCCC	556	23	GCAACTGCTTGGAATGGTT	726	19	TCTCTGGGAGATGCTAACAAGG	618	22
NM_016653	3699	AGTGAAAGCAGTCCAACTTGC	1659	21	GGAACGCTGTAAAGAAGTGTTG	1806	22	TCAGATGGCAACCCTGGAAG	1702	20
NM_016653	3699	CCTTTGAGATTGGTGCATGG	1031	20	ATGCCCATGTCTTTCAGGTC	1205	20	ACAACATTACAGGGAAGCGGC	1145	21
NM_016653	3699	TCAGCAGCTCGTCAGAAAA	1077	19	ATGCCCATGTCTTTCAGGTC	1205	20	ACAACATTACAGGGAAGCGGC	1145	21
NM_016653	3699	TGGGAGATGCTAACAAGGGA	622	20	TTGCTTGAATGATGGCCG	798	18	CTGTTACATCAGTGTTGGGAAGC	745	23
NM_016653	3699	CAGTAACAGAAGTGAGGAGATGG	318	23	CGTTTCTTGACTTGAGGTCTCTGT	448	24	TATGACCTGGGCCACTGATG	354	20
NM_016653	3699	GCAGTCCAACTTGCCATTC	1666	19	TTGCTCTGGGAACGCTGTA	1814	19	TTCAGATGGCAACCCTGGAAG	1701	21
NM_016653	3699	CCAGAGTCTCCCTGTGTCAGAA	567	22	GCAACTGCTTGGAATGGTT	726	19	TCTCTGGGAGATGCTAACAAGGGA	618	24
NM_016653	3699	CTGGGAGATGCTAACAAGGG	621	20	TGCTTGAATGATGGCCGT	797	18	TTGCCCCAGAAGTTTTGCTGAAC	723	23
NM_133646	7179	TGCCCCAGAAGTTTTGCT	885	18	AGGAAGGCTCGTGTCATTTG	1004	20	ACGGCCATCATTCAAGCAAA	941	20
NM_133646	7179	TCTCAGCTTTAAGGAGCAGGA	1094	21	ACATCTCTGCACTGTTTGACTCC	1263	23	TGTGGGAGCAAAAGCTGACA	1144	20
NM_133646	7179	ATTCCTACACAACAAGGCGGA	1019	21	TGCTCTGTCAGCTTTTGCTC	1168	20	CGTGATCTCAGCTTTAAGGAGCAGGAG	1089	27
NM_133646	7179	ATGACACGAGCCTTCCTGA	988	19	CTGCTCCTTAAAGCTGAGATCAC	1112	23	ACTCATTCCTACACAACAAGGCGG	1015	24
NM_133646	7179	GAGATGCTAACAAGGGAGGTCC	786	22	AGCTTCCCAACACTGATGTAACAG	929	24	TTGCCCCAGAAGTTTTGCTG	884	20
NM_133646	7179	TGTTACATCAGTGTTGGGAAGC	907	22	GCCTTGTTGTGTAGGAATGAGT	1036	22	CGGCCATCATTCAAGCAAAT	942	20
NM_133646	7179	CATGGATGGCTCCAGAAGTTAT	706	22	TTCCACTACAAGCCAAGCTACT	854	22	TCTCTGGGAGATGCTAACAAGGG	779	23
NM_133646	7179	TGATCTCAGCTTTAAGGAGCAG	1091	22	CTGCACTGTTTGACTCCTCTGT	1257	22	TGTGGGAGCAAAAGCTGACAGAG	1144	23
NM_133646	7179	AAAGCTGACAGAGCAGTCCAAC	1154	22	CAGACTTGGGTTCATGCCA	1319	19	ATCACAGCAACAAGTAACGGGG	1272	22
NM_133646	7179	GTTTTGGGAGTGTTTATCGAGC	271	22	AATGCCATAGTTGGGAGGTT	434	20	CAGGACAAGGAGGTGGCTGTAAA	306	23
NM_005319	732	AACACCGAAGAAAGCGAAGA	502	20	AGCCTTAGCAGCACTTTTGG	619	20	GTAACCAAGAAAGTGGCTAAGAGC	539	24
NM_005319	732	CACCGAAGAAAGCGAAGAAG	504	20	AGCCTTAGCAGCACTTTTGG	619	20	GTAACCAAGAAAGTGGCTAAGAGC	539	24
NM_005319	732	AGCGTAGCGGAGTTTCTCTG	201	20	CTTTCGTTTGCACCAGAGTG	335	20	GGAGAAAAACAACAGCCGTATC	262	22
NM_005319	732	AGCGTAGCGGAGTTTCTCTG	201	20	CCAGGCTCTTGAGACCAAGT	305	20	GGAGAAAAACAACAGCCGTATC	262	22
NM_005319	732	ACTCTGGTGCAAACGAAAGG	317	20	CCCCAACTGGCTTCTTAGGT	443	20	AAGCCCAAGGTTAAAAAGGC	392	20
NM_005319	732	AAACACCGAAGAAAGCGAAG	501	20	AGCCTTAGCAGCACTTTTGG	619	20	GTAACCAAGAAAGTGGCTAAGAGC	539	24
NM_005319	732	AGCCAAGCCCAAGGTTAAAA	388	20	TTCTTCGCTTTCTTCGGTGT	522	20	AAACCTAAGAAGCCAGTTGGG	422	21

							422	21
							392	20
							422	21
						AACCTAAGAAGCCAGTTGGG		
						AAGCCCAAGGTTAAAAAGGC		
						AAACCTAAGAAGCCAGTTGGG		
					521	20		
					443	20		
					522	20		
				TCTTCGCTTTCTTCGGTGTT				
				CCCCAACTGGCTTCTTAGGT				
				TTCTTCGCTTTCTTCGGTGT				
			388	20				
			316	20				
			389	19				
		AGCCAAGCCCAAGGTTAAAA						
		CACTCTGGTGCAAACGAAAG						
		GCCAAGCCCAAGGTTAAAA						
NM_005319	732							
NM_005319	732							
NM_005319	732							

EXAMPLE 6

Statistical Analysis of Real-Time RT-PCR Results

Real Time PCR analysis on blood samples isolated from individuals categorized as having bladder cancer (or a stage thereof) or not having bladder cancer were statistically analyzed using known methods in order to obtain data corresponding the level of abundance of the biomarkers of the invention in a training population.

Populations of individuals having bladder cancer and not having bladder cancer are chosen. Similar distributions of individuals, e.g. for phenotypes such as age, sex, body mass index (BMI) were selected. Selection of samples for which comparisons can be made on the basis of age and BMI are determined using KW One Way Analysis of Variance on Ranks as would be understood by a person skilled in the art.

EXAMPLE 7

ANALYSIS OF GENE EXPRESSION PROFILES OF BLOOD SAMPLES FROM INDIVIDUALS HAVING BLADDER CANCER AS COMPARED WITH GENE EXPRESSION PROFILES FROM NORMAL INDIVIDUALS USING THE PRODUCTS OF THE BIOMARKERS DESCRIBED IN FIGURE 1

This example demonstrates the use of the claimed invention to identify individual biomarkers of the invention by detecting differential gene expression in blood samples taken from patients with bladder cancer as compared to blood samples taken from healthy patients.

Blood samples are taken from patients who are clinically diagnosed with bladder cancer as defined herein; patients who do not having bladder cancer as defined herein and one or more test patients. Gene expression profiles of combinations of biomarkers of the invention are then analyzed and the test individuals profile compared with the two control profiles.

Total mRNA from lysed whole blood is taken from each patient is first isolated using TRIzol® reagent (GIBCO) and fluorescently labelled probes for each blood sample are then generated, denatured and hybridized to the Affymetrix® U133Version 2.0 microarray

containingoligonucleotides corresponding to each of the genes as described in Table 1. Detection of specific hybridization to the array is then measured by scanning with a GMS Scanner 418 and processing of the experimental data with Scanalyzer software (Michael Eisen, Stanford University), followed by GeneSpring software (Silicon Genetics, CA) analysis. Differential expression of the RNA products in blood samples corresponding to the biomarkers as between the two control populations and the test individual is determined by statistical analysis using the Wilcox Mann Whitney rank sum test (Glantz SA. Primer of Biostatistics. 5th ed. New York, USA: McGraw-Hill Medical Publishing Division, 2002).

EXAMPLE 8

Application of Logistic Regression to a Combination of the Invention to Identify Classifiers and Combinations Useful in Differentiating Bladder Cancer or a stage of Bladder Cancer from Non Bladder Cancer.

Data corresponding to the level of RNA products for the biomarkers in **Table 10**, and/or **Table 13** can be determined using Quantitative real-time RT-PCR (QRT-PCR) using Qiagen's QuantiTect™ SYBR® Green PCR kit on a reference (training population). In one case the reference population consists of individuals having bladder cancer (irrespective of stage of bladder cancer) and individuals not having bladder cancer. In another instance, the reference population consists of individuals having a stage of bladder cancer (e.g. early stage bladder cancer) and individuals not having bladder cancer (so as to identify combinations useful for screening/diagnosis of early stage bladder cancer). A reference dataset consisting of ΔCt values (i.e. comparing the Ct of the RNA products of a biomarkers of Table 10 as compared with the Ct from a housekeeping gene) from each of the individuals in the reference population for each gene can be created using the data generated from the QRT-PCR across a population of individuals, some having bladder cancer and some not having bladder cancer. Note that the housekeeping gene, as the name implies is one which does not vary with a statistical significance as between individuals having bladder cancer and not having bladder cancer. We note that neither GAPDH or Beta-actin is useful as a housekeeping gene, therefore another housekeeping gene should be utilized, for example RPLP0, B2M, RPS18 and the like. A logistic regression model can be applied to each possible combination of the 10 genes using the ΔCt values generated. Resulting classifiers in the form of Equation 1 below are thus generated.

$$X = \text{Logit}(P) = \ln(P/(1-P)) = b_0 + b_1 \Delta Ct_1 + b_2 \Delta Ct_2 + \ldots + b_n \Delta Ct_n \qquad \text{(Eq 1)}$$

Where P = probability that a patient sample is diagnosed as "diseased" (e.g. having bladder cancer) As such, X = Logit (p) can be defined as follows:

If $X \geq$ threshold then Y=1 (diagnosis = "diseased" e.g. having bladder cancer), and if $X <$ threshold then Y=0 (diagnosis = "control"). As would be understood, the threshold is chosen so as to set the required specificity or sensitivity desired. The resulting ROC area and the remaining sensitivity or specificity are determined thus providing a measure of the likelihood that the diagnosis generated is a true positive or a false positive result.

Cross Validation

Having identified useful combinations of biomarkers and classifiers specific to these combinations using methods as noted above, one can cross validate the results using the training set using one or more methods as would be known. For example, one can cross validate the resulting classifiers using WEKA (http://www.cs.waikato.ac.nz/~ml/weka/index.html) (22). In one embodiment, two different cross-validation schemes can be used. For example, the WEKA MetaAnalysis function can be used to construct 100 bootstrap replicates of the training data set and each new dataset analyzed using a simple logistic model. Each of the resultant logistic equations (classifiers) can then be analyzed by 10-fold cross-validation and the esults for all equations averaged ("bootstrap aggregating").

Prospective (Blind) Test

To further confirm and select one or more resulting classifiers, a blind set of clinical samples can be tested. A test set consisted of controls and subjects with bladder cancer is used. Preferably, none of the blind test set individuals were used in creating the reference data set.

The measured values for each blind sample can be evaluated using one or more of the resulting classifiers defined by the training set and cross validated. In addition, voting schemes using two or more of the resulting classifiers can also be used. For example, and algorithm twhich consists of an initial calculation, a binary logic gate, and a "committee machine" vote can be used. *Initial Calculation:* The logit function X_i can be computed for each of the equations (I = 1 to N) that gave ROC Area >for example, 0.6, 0.7, 0.8, or 0.9 against the reference data set. *Logic Gate*: If $X_i <$ threshold of j (where j stands for an equation number), then the blind sample is given score $S_i = -1$ for Eqn #j. If $X_i \geq$ threshold of j, then it is given score $S_i = +1$ for Eqn #j. *Vote by Committee Machine (23):* A vote can then be taken over the scores from the 10 or more logit equations used for diagnosis. By definition, Vote = $\sum S_i$. If Vote ≤ 0 then the sample is called "no prostate cancer", while if Vote > 0 then the sample is called "diseased or having bladder cancer".

EXAMPLE 9

Application of Two Gene "Ratios" of Selected Biomarkers as a Mathematical Model resulting in Classifiers Useful in Differentiating Bladder Cancer from Non Bladder Cancer.

In order to avoid the necessity to identify a housekeeping gene, a strategy of directly comparing the differential expression of two expressed genes was implemented and surprising proved more beneficial than utilizing a housekeeping gene. The use of pair-wise gene ratios for disease group discrimination using c-DNA array data from tissues (rather than from blood) has previously been described *(16, 17)*. The technique reduces inter-individual variation caused by biological and technical factors and is also independent of the expression measuring platform for data acquisition. Other advantages of the pair-wise gene method include the fact that the method is relatively independent of the input of sample amount making it unnecessary to use a housekeeping gene as loading control, and the method requires only small amounts of RNA. Finally, the use of the ratio of an up-regulated gene to a down-regulated gene amplifies the signal thus making the assay more sensitive.

As shown in Table 10, five of the biomarkers were identified as being up-regulated as between bladder cancer patients and control patients and five of the biomarkers were identified as down-regulated. In this example, combinations of two biomarkers (one up regulated and one down regulated (wherein upregulated refers to genes which have a higher level of expression in

patients with bladder cancer as compared with patients not having bladder cancer and donwregulated refers to genes which have a lower level of expression in patients with bladder cancer as compared with patients without bladder cancer)) were selected for input into logistic regression to determine the diagnostic capabilities of different pair combinations from the biomarkers listed in Table 10. A number of combinations were evaluated in maximum-likelihood logistic regression to determine the discrimination ability (ROC Area > 0.5) of "bladder cancer" vs. "control".

Table 18: Real-time RT-PCR result demonstrating several 2 gene combinations with classification power to discriminate between bladder cancer and non-cancer samples

Ratio ΔCt	AUC	Sensitivity (95%CI,)	Specificity (95%CI)	p value
IGFBP7-CHD2	0.907±0.058	80 (56.3- 94.1)	92.9 (66.1- 98.8)	<0.001
IGFBP7-NELL2	0.879±0.066	90 (68.3- 98.5)	85.7 (57.2- 97.8)	<0.001
IGFBP7-CSPG6	0.896±0.061	85 (62.1- 96.6)	85.7 (57.2- 97.8)	<0.001
IGFBP7-TNFRSF7	0.864±0.069	90 (75.1- 99.2)	71.4 (41.9- 91.4)	<0.001
IGFBP7-SNX16	0.871±0.067	75 (50.9- 91.2)	85.7 (57.2- 97.8)	<0.001
CTSD-CHD2	0.825±0.077	90 (68.3- 98.5)	64.3 (35.2- 87.1)	0.002
CTSD-NELL2	0.737±0.090	60 (36.1- 80.8)	85.7 (57.2- 97.8)	0.021
CTSD-CSPG6	0.764±0.087	70 (45.7- 88.0)	85.7 (57.2- 97.8)	0.01
CTSD-TNFRSF7	0.668±0.097	45 (23.1- 68.4)	85.7 (57.2- 97.8)	0.104
CTSD-SNX16	0.723±0.092	45 (23.1- 68.4)	92.9 (66.1- 98.8)	0.03

Of the pairs tested using a training population consisting of the ratio ΔCt between IGFBP7 and CHD2 performed the best in discriminating bladder cancer patients (n=20) from normal patients (n=14) Scoring of the pair yields an AUC of 0.907, a sensitivity of 80.0% (95% CI 56.3-94.1%), and a specificity of 92.9% (95% CI 66.1-98.8%). The ratio ΔCt between IGFBP7 and NELL2 was ranked the second best with an AUC of 0.879, a sensitivity of 90.0% (95% CI 68.3-98.5%), and specificity of 85.7% (95% CI 57.2-97.8).

This ratio was subsequently tested on a second training population consisting of more individuals, namely resulting in a sensitivity of the assay in this cohort is 83%, as 34 of 40 bladder cancer cases were predicted correctly. Of note, 3 of the 6 bladder cancer patients predicted wrongly by the assay to have no bladder cancer were patients with resected superficial bladder cancer. This suggests that the assay might be useful in determining patient prognosis. The specificity of the assay is 93%, as only 2 of the 27 healthy controls were predicted wrongly to have bladder cancer.

EXAMPLE 9

Application of Two Gene "Ratios" of Selected Biomarkers as a Mathematical Model resulting in Classifiers Useful in Differentiating Bladder Cancer from Non Bladder Cancer.

A subset of markers from **Table 13** were chosen to more exhaustively identify useful combinations of ratios of genes for use of biomarkers to screen and/or diagnose patients with bladder cancer. A reference training population was selected consisting of 20 individuals having bladder cancer and 14 individuals not having bladder cancer. A reference data set corresponding to the ΔCt of ratio's of the RNA products of the selected biomarkers SNX16, CSPG6, IGFBP7, and CTSD and TNFRSF7, NELL2, and CHD2. Thus ΔCt ratios were determined for CHD2/CTSD, CHD2/SNX16, CHD2/CSPG6, CHD2/IGFBP7, CHD2/TNFRSF7, CHD2/NELL2, NELL2/SNX16, NELL2/CSPG6, NELL2/IGFBP7, NELL2/CTSD, NELL2/TNFRSF7, TNFRSF7/SNX16, TNFRSF7/CSPG6, TNFRSF7/IGFBP7, TNFRSF7/CSTD, CTSD/SNX16, CTSD/CSPG6, CTSD/IGFBP7, IGFBP7/SNX16 IGFBP7/CSPG6 and CSPG6/SNX16 using Quantitative real-time RT-PCR (QRT-PCR) using Qiagen's QuantiTect™ SYBR® Green PCR kit where the Ct of each pair was measured within a single 96 well plate so as to allow better comparison of Ct values. (Note similar results could be obtained using multiplexing and TaqMan probes with different coloured flourophores as would be understood by a person skilled in the art) Identifying pairs in this manner allows one to input a value term to represent the level of expression of the pair of genes as a single term into logistic regression. Therefore one can create combinations of one ratio, two ratios, three ratios, four ratios, five ratios, 6 ratios, or all ratios such that the resulting logistic regression equations (classifiers) in the general formula

$$X = \text{Logit}(P) = \ln(P/(1-P)) = b_0 + b_1 \Delta Ct_1 + b_2 \Delta Ct_2 + \ldots + b_n \Delta Ct_n \qquad \text{(Eq 1)}$$

Where P = probability that a patient sample is diagnosed as "diseased" (e.g. having bladder cancer), takes on the form for combinations of single ratios as follows:

$X = \text{Logit}(P) = \ln(P/(1-P)) = b_0 + b_1 \Delta Ct_1$ (Eq 2) Where ΔCt_1 is data corresponding to any one of the ratios noted above.

Or takes on the form for combinations of two ratios as follows:

$X = \text{Logit}(P) = \ln(P/(1-P)) = b_0 + b_1 \Delta Ct_1 + \Delta Ct_2$ (Eq 3) where ΔCt_1 is data corresponding to any one of the ratios noted above and ΔCt_2 is also data corresponding to any of the ratios noted above.

Similarly, all three ratio combinations, four ratio combinations, and five ratio combinations were tested.

Each resulting classifier with an ROC Area > 0.5, greater than 0.55, greater than 0.6, greater than 0.65, greater than 0.7, greater than 0.75, greater than 0.8, greater than 0.85, greater than 0.9 or greater than 0.95 can be used to diagnose or screen a test individuals having having or not having bladder cancer. The graphical results of classifiers resulting from combinations of single ratios, two ratios, three ratios and four ratios is shown in **Figures 1-4**.

Representative classifiers are of four ratios using an expanded training population of 40 individuals having bladder cancer and 27 control individuals are shown in **Table 19,** representative classifiers of three ratios are shown in **Table 20,** representative classifiers of two ratios are shown in **Table 21** and representative classifiers of single ratios are shown in **Table 22** below

Table 19

Combination #	# of Ratios	ROCarea	Sens @ 90% spec	Spec @ 90% sens	Constant	CHD2 / IGFBP7	CHD2 / CSPG6	IGFBP7 / CSPG6	CHD2 / CTSD	IGFBP7 / CTSD	CSPG6 / CTSD	CHD2 / NELL2	IGFBP7 / NELL2	CSPG6 / NELL2	CTSD / NELL2	CHD2 / SNX16	IGFBP7 / SNX16	CSPG6 / SNX16	CTSD / SNX16	NELL2 / SNX16	CHD2 / TNFRSF7	IGFBP7 / TNFRSF7	CSPG6 / TNFRSF7	CTSD / TNFRSF7	NELL2 / TNFRSF7	SNX16 / TNFRSF7	Threshold Acc	Threshold Sens	Threshold Spec
8	1	0.8917	80.00	70.37	-11.498				3.183																		0.691	-0.423	0.646
1	1	0.8537	55.00	59.26	4.392	2.292																					-0.698	-0.422	1.350
32	1	0.8278	50.00	55.56	-6.228						2.085																-0.368	-0.326	1.119
4	1	0.8093	40.00	48.15	5.305			-2.1																			-0.101	-0.395	1.378
1024	1	0.7602	42.50	18.52	5.432											2.068											0.428	-0.729	1.121
262144	1	0.7482	40.00	25.93	-2.544																			-1.06			0.252	-0.468	1.112
65536	1	0.7435	52.50	7.41	3.474																	-1.12					0.891	-0.742	0.811
128	1	0.7407	45.00	25.93	1.053								-0.9														-0.160	-0.457	0.000
512	1	0.7259	42.50	18.52	-3.413										-0.79												-0.204	-0.473	0.853
8192	1	0.7157	22.50	29.63	-5.785														-1								0.251	-0.087	0.942
2	1	0.6889	32.50	11.11	-0.654		1.817																				0.473	-0.345	0.774
1048576	1	0.6750	30.00	14.81	3.084																					-0.8	0.200	-0.307	1.010
2048	1	0.6630	15.00	29.63	-0.019												-0.62										-0.031	0.012	0.983
16384	1	0.6537	37.50	11.11	1.247															0.639							0.198	-0.269	0.780
4096	1	0.6306	22.50	7.41	3.653													1.083									0.068	-0.246	0.000
256	1	0.5954	37.50	7.41	-0.492									-0.53													0.207	-0.106	0.646
131072	1	0.5824	27.50	7.41	0.572																		-0.52				0.374	-0.079	0.639
524288	1	0.5769	5.00	14.81	-0.685																				0.535		0.138	0.154	0.824
64	1	0.5232	30.00	0.00	0.248							-0.13															0.118	0.224	0.471
32768	1	0.5213	27.50	0.00	0.430																-0.04						0.324	0.347	0.417
16	1	0.5037	10.00	22.22	0.531					-0.02																	0.375	0.379	0.413

Table 20

Combination #	# of Ratios	ROCarea	Sens @ 90% spec	Spec @ 90% sens	Constant	CHD2 / CSPG6	CHD2 / CTSD	CSPG6 / CTSD	CHD2 / IGFBP7	CSPG6 / IGFBP7	CTSD / IGFBP7	CHD2 / NELL2	CSPG6 / NELL2	CTSD / NELL2	IGFBP7 / NELL2	CHD2 / SNX16	CSPG6 / SNX16	CTSD / SNX16	IGFBP7 / SNX16	NELL2 / SNX16	CHD2 / TNFRSF7	CSPG6 / TNFRSF7	CTSD / TNFRSF7	IGFBP7 / TNFRSF7	NELL2 / TNFRSF7	SNX16 / TNFRSF7	Threshold Acc	Threshold Sens	Threshold Spec
131584	2	0.9250	90.00	64.29	7.455										-2.866								1.870				0.576	-0.567	0.500
16400	2	0.9214	85.00	71.43	6.958					2.402										0.532							0.045	-0.444	0.471
144	2	0.9143	85.00	78.57	5.255					2.372			-0.480														-0.371	-0.328	0.561
528	2	0.9143	85.00	78.57	5.255					1.892					-0.480												-0.371	-0.328	0.561
640	2	0.9143	85.00	78.57	5.255								1.892		-2.372												-0.371	-0.328	0.561
72	2	0.9143	85.00	71.43	5.846				2.255			-0.527															0.069	-0.111	0.552
520	2	0.9143	85.00	71.43	5.846				1.728						-0.527												0.069	-0.111	0.552
576	2	0.9143	85.00	71.43	5.846							1.728			-2.255												0.069	-0.111	0.552
80	2	0.9143	80.00	71.43	5.346					2.360		-0.299															0.052	-0.276	0.808
20	2	0.9143	70.00	64.29	10.182			-0.890		3.001																	0.669	-0.253	1.360
36	2	0.9143	70.00	64.29	10.182			2.111			3.001																0.669	-0.253	1.360
48	2	0.9143	70.00	64.29	10.182					2.111	0.890																0.669	-0.253	1.360
1040	2	0.9107	90.00	64.29	10.940					2.422						1.281											0.243	-0.170	0.234
33280	2	0.9107	90.00	64.29	2.451										-2.031						2.007						0.223	-0.750	0.096
524296	2	0.9107	90.00	64.29	2.519				2.038																2.010		0.223	-0.734	0.103
66048	2	0.9107	85.00	85.71	1.352										-2.173							2.280					-0.391	-0.322	0.731
524304	2	0.9107	85.00	85.71	1.932					2.214															1.935		-0.370	-0.333	0.690
4112	2	0.9107	80.00	78.57	18.265					3.462								1.563									-0.199	-0.042	0.900
16392	2	0.9107	80.00	71.43	7.759				2.222											0.539							-0.006	-0.436	0.954
1048584	2	0.9107	80.00	71.43	8.360				2.275																	-0.373	0.521	-0.033	0.998
1048592	2	0.9107	80.00	64.29	7.277					2.400																-0.337	0.164	-0.464	0.735
18	2	0.9107	65.00	64.29	9.540		-0.860			2.962																	0.342	-0.245	1.768
10	2	0.9071	85.00	71.43	8.323		-0.302		2.457																		1.055	0.085	1.007
34	2	0.9071	85.00	71.43	8.323		2.155				2.457																1.055	0.085	1.007
40	2	0.9071	85.00	71.43	8.323				2.155		0.302																1.055	0.085	1.007
32776	2	0.9071	85.00	71.43	7.002				2.286												-0.299						0.707	0.020	0.680
262152	2	0.9071	85.00	71.43	7.002				1.987															-0.299			0.707	0.020	0.680
294912	2	0.9071	85.00	71.43	7.002																1.987		-2.286				0.707	0.020	0.680
9	2	0.9071	85.00	64.29	6.613	-1.396			2.393																		1.029	-0.177	0.989
17	2	0.9071	85.00	64.29	6.613	0.998				2.393																	1.029	-0.177	0.989
24	2	0.9071	85.00	64.29	6.613				0.998	1.396																	1.029	-0.177	0.989
65544	2	0.9071	85.00	64.29	7.038				2.266													-0.036					0.835	-0.469	0.795
131080	2	0.9071	85.00	64.29	6.832				2.242														-0.046				0.794	-0.449	0.749
136	2	0.9071	80.00	71.43	6.276				2.214				-0.366														0.141	0.045	0.884
1032	2	0.9071	80.00	71.43	9.282				2.245							0.614											0.627	-0.065	1.069
8200	2	0.9071	80.00	71.43	9.282				2.859										0.614								0.627	-0.065	1.069
9216	2	0.9071	80.00	71.43	9.282											2.859			-2.245								0.627	-0.065	1.069
2056	2	0.9071	80.00	64.29	11.765				2.368								1.385										0.706	-0.443	0.839
288	2	0.9071	75.00	57.14	11.204						3.154			-1.307													0.004	-0.757	1.029
544	2	0.9071	75.00	57.14	11.204						1.848				-1.307												0.004	-0.757	1.029

Table 21

Combinatio n #	# of Ratios	ROCarea	Sens @ 90% spec	Spec @ 90% sens	Constant	IGFBP7 / CHD2	CSPG6 / CHD2	IGFBP7 / CSPG6	CTSD / CHD2	IGFBP7 / CTSD	CSPG6 / CTSD	CHD2 / NELL2	IGFBP7 / NELL2	CSPG6 / NELL2	CTSD / NELL2	CHD2 / SNX16	IGFBP7 / SNX16	CSPG6 / SNX16	CTSD / SNX16	NELL2 / SNX16	CHD2 / TNFRSF7	IGFBP7 / TNFRSF7	CSPG6 / TNFRSF7	CTSD / TNFRSF7	NELL2 / TNFRSF7	SNX16 / TNFRSF7	Threshold Acc	Threshold Sens	Threshold Spec
2056	2	0.9019	82.50	70.37	-11.578				3.147								-0.315										0.734	-0.586	0.680
12	2	0.8991	82.50	59.26	-8.477			-0.612	2.763																		0.832	-0.655	0.662
4104	2	0.8963	80.00	70.37	-12.537				3.302									-0.199									0.741	-0.463	0.741
10	2	0.8963	75.00	74.07	-12.205		1.417		3.164																		-0.296	-0.244	0.731
34	2	0.8963	75.00	74.07	-12.205		4.581				3.164																-0.296	-0.244	0.731
40	2	0.8963	75.00	74.07	-12.205				4.581		-1.417																-0.296	-0.244	0.731
136	2	0.8954	82.50	62.96	-10.403				2.941				-0.240														0.839	-0.811	0.764
9	2	0.8944	82.50	59.26	-6.662	1.077			2.390																		0.490	-0.436	0.431
17	2	0.8944	82.50	59.26	-6.662	3.468					2.390																0.490	-0.436	0.431
24	2	0.8944	82.50	59.26	-6.662				3.468		-1.077																0.490	-0.436	0.431
65544	2	0.8944	82.50	48.15	-8.715				2.826													-0.519					0.373	-0.969	0.793
16392	2	0.8944	80.00	70.37	-11.618				3.205																		0.669	-0.392	0.652
32776	2	0.8917	80.00	70.37	-11.483				3.183							-0.024											0.713	-0.440	0.653
262152	2	0.8917	80.00	70.37	-11.483				3.167							-0.016											0.713	-0.440	0.653
294912	2	0.8917	80.00	70.37	-11.483																-0.016						0.713	-0.440	0.653
72	2	0.8917	77.50	70.37	-11.534				3.228			0.135				3.167					-3.183						0.498	-0.208	0.729
520	2	0.8917	77.50	70.37	-11.534				3.363						0.135												0.498	-0.208	0.729
576	2	0.8917	77.50	70.37	-11.534							3.363			-3.228												0.498	-0.208	0.729
264	2	0.8907	80.00	70.37	-11.531				3.151					-0.100													0.848	-0.518	0.672
1048584	2	0.8869	80.00	70.37	-10.424				3.072																		0.475	-0.403	0.621
1032	2	0.8880	80.00	70.37	-10.039				2.996								0.311									-0.191	0.497	-0.319	0.481
8200	2	0.8880	80.00	70.37	-10.039				3.308									0.311									0.497	-0.319	0.481
9216	2	0.8880	80.00	70.37	-10.039												3.308		-2.996								0.497	-0.319	0.481
524296	2	0.8880	77.50	74.07	-10.992				3.391																	-0.648	-0.289	-0.090	0.892
33	2	0.8861	82.50	51.85	-0.601	1.888					1.353																0.704	-0.661	0.760
131080	2	0.8852	80.00	59.26	-11.274				3.183																		0.629	-0.847	0.768
262145	2	0.8833	67.50	66.67	-1.728	2.093																		-0.488			0.307	-0.145	0.977
513	2	0.8787	77.50	62.96	1.140	2.074									-0.615										-0.870		0.665	-0.075	0.717
16385	2	0.8769	72.50	48.15	5.206	2.216														0.637							0.325	-0.835	0.844
1048577	2	0.8759	67.50	59.26	7.511	2.292																				-0.894	0.479	-0.593	0.953
1025	2	0.8704	65.00	62.96	6.781	1.941										1.220											-0.515	-0.479	1.072
2049	2	0.8704	65.00	62.96	6.781	3.162											1.220										-0.515	-0.479	1.072
3072	2	0.8704	65.00	62.96	6.781											3.162	-1.941										-0.515	-0.479	1.072
4097	2	0.8657	65.00	59.26	7.580	2.216												1.096									-0.515	-0.479	1.072
514	2	0.8648	75.00	29.63	-9.481		4.645								-1.519												0.209	-0.523	1.009
262146	2	0.8620	65.00	40.74	-6.947		4.199																				0.954	-1.327	0.649
257	2	0.8602	62.50	62.96	3.641	2.249								-0.440										-1.795			0.749	-0.895	0.948
66560	2	0.8574	62.50	29.63	9.102											2.067											0.225	-0.471	1.211
65	2	0.8574	52.50	59.26	4.184	2.327					-0.279																0.603	-1.113	0.977
129	2	0.8574	52.50	59.26	4.184	2.048																-1.305					0.349	-0.413	1.220
													-0.279														0.349	-0.413	1.220

Table 22

Combination #	# of Ratios	ROCarea	Sens @ 90% spec	Spec @ 90% sens	Constant	CHD2 / CSPG6	CHD2 / CTSD	CSPG6 / CTSD	CHD2 / IGFBP7	CSPG6 / IGFBP7	CTSD / IGFBP7	CHD2 / NELL2	CSPG6 / NELL2	CTSD / NELL2	IGFBP7 / NELL2	CHD2 / SNX16	CSPG6 / SNX16	CTSD / SNX16	IGFBP7 / SNX16	NELL2 / SNX16	CHD2 / TNFRSF7	CSPG6 / TNFRSF7	CTSD / TNFRSF7	IGFBP7 / TNFRSF7	NELL2 / TNFRSF7	SNX16 / TNFRSF7	Threshold Acc	Threshold Sens	Threshold Spec
2336	3	0.9536	75.00	85.71	21.762						3.661			-1.138			2.202									-0.690	-0.292	1.628
2592	3	0.9536	75.00	85.71	21.762						2.523				-1.138		2.202									-0.690	-0.292	1.628
2816	3	0.9536	75.00	85.71	21.762									2.523	-3.661		2.202									-0.690	-0.292	1.628
16402	3	0.9429	85.00	78.57	15.074		-1.896			3.616										0.910						-0.883	-0.053	0.238
4176	3	0.9393	80.00	85.71	20.359					3.671		-0.716						2.071								-0.723	-0.138	0.776
4240	3	0.9393	75.00	85.71	19.816					3.582			-0.770					1.969								-0.531	-0.181	1.331
4624	3	0.9393	75.00	85.71	19.816					2.812					-0.770			1.969								-0.531	-0.181	1.331
4736	3	0.9393	75.00	85.71	19.816								2.812		-3.582			1.969								-0.531	-0.181	1.331
67072	3	0.9321	100.00	92.86	5.855										-2.148	1.208						2.334				-0.538	-0.075	-0.601
525328	3	0.9321	100.00	92.86	6.729					2.223						1.200								1.798		-0.546	-0.217	-0.546
131585	3	0.9321	90.00	64.29	8.500	-1.500									-3.344								2.399			0.442	-0.299	0.281
16404	3	0.9321	75.00	85.71	15.166			-1.727		3.437										0.926						-0.219	-0.127	1.138
16420	3	0.9321	75.00	85.71	15.166			1.710			3.437									0.926						-0.219	-0.127	1.138
16432	3	0.9321	75.00	85.71	15.166					1.710	1.727									0.926						-0.219	-0.127	1.138
65554	3	0.9321	70.00	71.43	10.542		-1.091			3.096												-0.405				0.242	-0.207	1.350
262162	3	0.9321	70.00	71.43	10.542		-1.091			2.692														-0.405		0.242	-0.207	1.350
327682	3	0.9321	70.00	71.43	10.542		-1.091															2.692		-3.096		0.242	-0.207	1.350
2120	3	0.9286	85.00	78.57	10.526				2.358			-0.442					1.300									-0.758	-0.257	0.646
2568	3	0.9286	85.00	78.57	10.526				1.916						-0.442		1.300									-0.758	-0.257	0.646
2624	3	0.9286	85.00	78.57	10.526							1.916			-2.358		1.300									-0.758	-0.257	0.646
37376	3	0.9286	85.00	78.57	9.401										-2.481			1.082			2.023					-0.985	-0.260	0.805
131216	3	0.9286	85.00	78.57	8.989					3.222			-1.862										1.467			-0.601	-0.015	0.630
131600	3	0.9286	85.00	78.57	8.989					1.360					-1.862								1.467			-0.601	-0.015	0.630
131712	3	0.9286	85.00	78.57	8.989								1.360		-3.222								1.467			-0.601	-0.015	0.630
4368	3	0.9286	80.00	85.71	19.756					3.410				-0.566				2.247								-0.237	0.033	0.971
16656	3	0.9286	80.00	85.71	19.756					3.410				1.681						2.247						-0.237	0.033	0.971
20496	3	0.9286	80.00	85.71	19.756					3.410								1.681		0.566						-0.237	0.033	0.971
36880	3	0.9286	75.00	78.57	19.604					3.542								1.763			-0.356					-0.772	-0.191	1.010
1048596	3	0.9286	75.00	71.43	14.819			-1.387		3.301																-0.700	-1.455	-0.219	0.969
1048612	3	0.9286	75.00	71.43	14.819			1.914			3.301															-0.700	-1.455	-0.219	0.969
1048624	3	0.9286	75.00	71.43	14.819					1.914	1.387															-0.700	-1.455	-0.219	0.969
1312	3	0.9286	70.00	85.71	15.817						3.088			-1.233		1.240										-0.724	0.228	1.707
1568	3	0.9286	70.00	85.71	15.817						1.855				-1.233	1.240										-0.724	0.228	1.707
1792	3	0.9286	70.00	85.71	15.817									1.855	-3.088	1.240										-0.724	0.228	1.707
133152	3	0.9286	60.00	78.57	23.774						3.588						2.262						-1.133			-0.526	-0.464	2.005
264224	3	0.9286	60.00	78.57	23.774						2.455						2.262							-1.133		-0.526	-0.464	2.005
395264	3	0.9286	60.00	78.57	23.774												2.262						2.455	-3.588		-0.526	-0.464	2.005
49168	3	0.9250	95.00	92.86	8.565					2.262										1.497	1.690					-0.105	-0.011	-0.155
526344	3	0.9250	85.00	85.71	7.166				2.142								1.260								1.848	-0.495	-0.310	0.718
65824	3	0.9250	85.00	78.57	7.845						3.099			-1.885								1.669				-0.527	-0.115	0.706

References

Zaleske DJ. Cartilage and Bone Development. Instr Course Lect 1998;47:461-

Buckwalter JA, Mankin HJ. Articular Cartilage: Tissue Design and Chondrocyte-Matrix Interactions. Instr Course Lect 1998;47:477-86.

Westacott CI, Sharif M. Cytokines in Osteoarthritis: Mediators or Markers of Joint Destruction? Semin Arthritis Rheum 1996;25:254-72

Adams MD, Kerlavage AR, Fleischmann RD, Fuldner RA, Bult CJ, Lee NH, et al. Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. Nature 1995;377 Suppl:3-174.

Hwang DM, Dempsey AA, Wang RX, Rezvani M, Barrans JD, Dai KS, et al. A Genome-Based Resource for Molecular Cardiovascular Medicine: Toward a Compendium of Cardiovascular Genes. Circulation 1997;96:4146-203.

Mao M, Fu G, Wu JS, Zhang QH, Zhou J, Kan LX, et al. Identification of genes expressed in human $CD34^+$ hematopoietic stem/progenitor cells by expressed sequence tags and efficient full-length cDNA cloning. Proc Natl Acad Sci 1998;95:8175-80.

Hillier LD, Lennon G, Becker M, Bonaldo MF, Chiapelli B, Chissoe S, et al. Generation and analysis of 280,000 human expressed sequence tags. Genome Res. 1996;6:807-28.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990;215:403-10.

Mundlos S, Zabel B. Developmental Expression of Human Cartilage Matrix Protein. Dev Dyn 1994;199:241-52.

Nakamura S, Kamihagi K, Satakeda H, Katayama M, Pan H, Okamoto H, et al. Enhancement of SPARC (osteonectin) synthesis in arthritic cartilage. Increased levels in synovial fluids from patients with rheumatoid arthritis and regulation by growth factors and cytokines in chondrocyte cultures. Arthritis Rheum 1996;39:539-51.

Eyre DR, The Collagens of Articular Cartilage. Semin Arthritis Rheum 1991;21 (3 Suppl 2):2-11.

Okihana H, Yamada K. Preparation of a cDNA Library and Preliminary Assessment of 1400 Genes from Mouse Growth Cartilage. J Bone Miner Res 1999;14:304-10.

Morrison EH, Ferguson MWJ, Bayliss MT, Archer CW. The developmental of articular cartilage: I. The spatial and temporal patterns of collagen types. J Anat 1996;189:9-22.

Treilleux I, Mallein-Gerin F, le Guellec D, Herbage D. Localization of the Expression of Type I, II, III Collagens, and Aggrecan Core Protein Genes in Developing Human Articular Cartilage. Matrix 1992;12:221-32.

Eyre DR, Wu JJ, Niyibizi C. The collagens of bone and cartilage: Molecular diversity and supramolecular assembly. In Cohn DV, Glorieux FH, Martin TJ, editors. Calcium Regulation and Bone Metabolism. Amsterdam. The Netherlands: Elsevier; 1990. p. 188-94.

Birnbacher R. Amann G, Breitschopf H, Lassmann H, Suchanek G, Heinz-Erian P. Cellular localization of insulin-like growth factor II mRNA in the human fetus and the placenta: detection with a digoxigenin-labeled cRNA probe and immunocytochemistry. Pediatr Res 1998;43:614-20.

Wang E, Wang J, Chin E, Zhou J, Bondy CA. Cellular patterns of insulin-like growth factor system gene expression in murine chondrogenesis and osteogenesis. Endocrinology 1995; 136:2741-51.

van Kleffens M, Groffen C, Rosato RR, van den Eijnde SM, van Neck JW, Lindenbergh-Kortleve DJ, et al. mRNA expression patterns of the IGF system during mouse limb bud development, determined by whole mount in situ hybridization. Mol Cell Endocrinol 1998;138:151-61.

Braulke T, Gotz W, Claussen M. Immunohistochemical localization of insulin-like growth factor binding protein-1, -3, and –4 in human fetal tissues and their analysis in media from fetal tissue explants. Growth Regul 1996;6:55-65.

Kessler E, Takahara K, Biniaminov L, Brusel M, Greenspan DS. Bone Morphogenetic Protein-1: The Type I Procollagen C-Proteinase. Science 1996;271:360-2.

Ausubel et al., John Weley & Sons, Inc., 1997, Current Protocols in Molecular Biology

Marshall, K. et al., 2000, 46th Annual Meeting, ORS, paper No. 919.

Kumar, S., et al., 2000, 46th Annual Meeting, ORS, paper No. 1031.

Marshall K., et al., 2002, 48th Annual meeting, ORS (submitted).

Migita K., et al., Biochem Biophys Res Commun 1997, 239:621-625.

Migita K., et al., Kidney Int 1999, 55:572-578.

Herr H, Shipley WU, Bajorin DF. Cancer of the Bladder. In: Rosenberg SA, ed. Cancer: Principles and Practice of Oncology. Philadelphia: Lippincott, Williams, and Wilkins, 2001:1396-1418

The contents of all references, patents and patent applications (including, published patent applications) cited throughout this application are hereby incorporated by reference.

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

What is claimed is:

1. A method of screening for bladder cancer, or a stage of bladder cancer, in a test subject, comprising:

(a) determining, in an isolated blood sample of said subject, a level of RNA product of a first biomarker selected from the group of biomarkers set out in Table 13; and a level of RNA product of a second biomarker selected from said group of biomarkers; and

(b) for each of said first and second biomarkers, comparing said level of RNA product to a control level of RNA encoded by said biomarker and detected in an isolated blood sample from one or more control subjects, wherein for each of said biomarkers, determination of a statistically significant difference between said level of RNA of step (a) and said control level is indicative of bladder cancer, or said stage of bladder cancer, in said test subject.

2. A method of screening for bladder cancer, or a stage of bladder cancer, in a test subject, comprising:

(a) determining, in an isolated blood sample of said subject, a level of RNA encoded by a biomarker selected from the group of biomarkers set out in Table 13; and

(b) comparing said level to a control level, said control level being a level of RNA products encoded by said biomarker and found in an isolated blood sample of one or more control subjects, wherein determination of a difference between said level of step (a) and said control level is indicative of bladder cancer, or said stage of bladder cancer, in said test subject.

3. The method according to any of claims 1 or 2, wherein said group of biomarkers are limited to SNX16, CSPG6, IGFBP7, CTSD and TNFRSF7, NELL2, and CHD2

4. The method according to any of claims 1 or 2 wherein said isolated blood sample consists of whole blood.

5. The method according to any of claims 1 or 2 wherein said isolated sample consists of a drop of blood.

6. The method according to claim 1 wherein said determination of said level of RNA of step (a) for each said biomarker is effected by first isolating RNA from said blood sample of said test subject.

7. The method according to claim 1, wherein said determination of said level of RNA of step (a) for each said biomarker is effected using quantitative real-time RT-PCR (QRT-PCR).

8. The method according to claim 1, wherein said determination of said level of RNA of step (a) for each said biomarker is effected by hybridizing said RNA products encoded by each said biomarker to an array comprising at least one probe complementary to said RNA products of each of said biomarkers.

9. The method according to claim 1, wherein said determination of said level of RNA of step (a) for each said biomarker is effected by hybridizing cDNA, PCR products or ESTs derived from said RNA products to an array comprising at least one probe complementary to said cDNA, PCR products or ESTs each of said biomarkers.

10. The method according to claim 1, wherein the control level is derived from said one or more individuals not having bladder cancer.

11. A method of screening for bladder cancer, or a stage of bladder cancer, in a test subject, comprising:

(a) determining, in an isolated blood sample of said subject, a level of protein product of a first biomarker selected from the group of biomarkers set out in Table 13; and a level of protein product of a second biomarker selected from said group of biomarkers; and

(b) for each of said first and second biomarkers, comparing said level of protein product to a control level of protein product encoded by said biomarker and detected in an isolated blood sample from one or more control subjects, wherein for each of said biomarkers, determination of a statistically significant difference between said level of protein product of step (a) and said control level is indicative of bladder cancer, or said stage of bladder cancer, in said test subject.

12. A method of screening for bladder cancer, or a stage of bladder cancer, in a test subject, comprising:

(a) determining, in an isolated blood sample of said subject, a level of protein product encoded by a biomarker selected from the group of biomarkers set out in Table 13; and

(b) comparing said level to a control level, said control level being a level of protein product encoded by said biomarker and found in an isolated blood sample of one or more control subjects, wherein determination of a difference between said level of step (a) and said control level is indicative of bladder cancer, or said stage of bladder cancer, in said test subject.

13. A composition comprising a collection of two or more isolated proteins which bind selectively to the protein products of at least two unique biomarkers, wherein each said unique biomarker is selected from the group of biomarkers listed Table 13.

14. A composition comprising a collection of two or more sets of biomarker specific primers wherein each set selectively amplifies double stranded DNA complementary to a unique biomarker, wherein each said unique biomarker is selected from the group of biomarkers listed in Table 13.

15. A composition comprising a collection of three or more sets of biomarker specific primers wherein each set selectively amplifies double stranded DNA complementary to a unique biomarker, and each said unique biomarker is selected from the group of biomarkers listed in Table 13.

16. A composition comprising a collection of four or more sets of biomarker specific primers wherein each set selectively amplifies double stranded DNA complementary to a unique biomarker, and each said unique biomarker is selected from the group of biomarkers listed in Table 13.

17. A composition comprising a collection of five or more sets of biomarker specific primers wherein each set selectively amplifies double stranded DNA complementary to a unique biomarker, and each said unique biomarker is selected from the group of biomarkers listed in Table 13.

18. A composition comprising a collection of two or more sets of biomarker specific primers wherein each set selectively amplifies double stranded DNA complementary to a unique biomarker, and each said unique biomarker is selected from the group of biomarkers consisting of SNX16, CSPG6, IGFBP7, and CTSD and TNFRSF7, NELL2 and CHD2.

19. A composition comprising a collection of three or more sets of biomarker specific primers wherein each set can selectively amplify double stranded DNA complementary to a unique biomarker, and each said unique biomarker is selected from the group of biomarkers consisting of SNX16, CSPG6, IGFBP7, and CTSD and TNFRSF7, NELL2 and CHD2.

20. A composition comprising a collection of two or more isolated proteins which bind selectively to the protein products of at least two unique biomarkers, wherein each unique biomarker is selected from the group of biomarkers consisting of SNX16, CSPG6, IGFBP7, and CTSD and TNFRSF7, NELL2 and CHD2.

21. The composition of claim 20 wherein said isolated proteins are antibodies.

22. Use of a composition according to claim 13 for selectively binding, in an isolated biological sample, a presence of at least two distinct proteins each of which is encoded by a unique biomarker, each said unique biomarker selected from the group consisting of the biomarkers set forth in Table 13.

23. Use of a composition according to claim 14 for selectively amplifying, in an isolated biological sample, double stranded DNA complementary to at least two unique biomarkers, and each said unique biomarker is selected from the group of biomarkers listed in Table 13.

24. Use of a composition according to claim 15 for selectively amplifying, in an isolated biological sample, double stranded DNA complementary to at least three unique biomarkers, and each said unique biomarker is selected from the group of biomarkers listed in Table 13.

25. Use of a composition according to claim 16 for selectively amplifying, in an isolated biological sample, double stranded DNA complementary to at least four unique biomarkers, and each said unique biomarker is selected from the group of biomarkers listed in Table 13.

26. Use of a composition according to claim 17 for selectively amplifying, in an isolated biological sample, double stranded DNA complementary to at least five unique

biomarkers, and each said unique biomarker is selected from the group of biomarkers listed in Table 13.

27. Use of a composition according to claim 18 for selectively amplifying, in an isolated biological sample, double stranded DNA complementary to at least two unique biomarkers, and each said unique biomarker is selected from the group of biomarkers consisting of SNX16, CSPG6, IGFBP7, and CTSD and TNFRSF7, NELL2 and CHD2.

28. Use of a composition according to claim 19 for selectively amplifying, in an isolated biological sample, double stranded DNA complementary to at least three unique biomarkers, and each said unique biomarker is selected from the group of biomarkers consisting of SNX16, CSPG6, IGFBP7, and CTSD and TNFRSF7, NELL2 and CHD2.

29. Use of the composition of any one of claims 13, 14, 15, 16, 17, 18 or 19 for screening for bladder cancer, or a stage of bladder cancer, in a blood sample isolated from a test subject.

30. A kit for conducting a method according to any claim 1 comprising at least two sets of biomarker specific primers wherein each set of biomarker specific primers produces double stranded DNA complementary to a unique biomarker, each unique biomarker selected from said group of biomarkers;

(b) an enzyme with reverse transcriptase activity;

(c) an enzyme with thermostable DNA polymerase activity, and

(d) a labeling means;

wherein each of said primer sets is used to detect the quantitative expression levels of said unique biomarker in a test subject.

31. A kit for conducting a method according to claim 1 comprising at least two sets of biomarker specific primers wherein each set of biomarker specific primers produces double stranded DNA complementary to a unique biomarker, each unique biomarker selected from the group of biomarkers consisting of SNX16, CSPG6, IGFBP7, CTSD, TNFRSF7, NELL2, CHD2.

(b) an enzyme with reverse transcriptase activity;

(c) an enzyme with thermostable DNA polymerase activity, and

(d) a labeling means;

wherein each of said primer sets is used to detect the quantitative expression levels of said unique biomarker in a test subject.

32. A kit comprising;

(a) the composition of any one of claims 14, 15, 16, 17, 18 or 19;

(b) an enzyme with reverse transcriptase activity;

(c) an enzyme with thermostable DNA polymerase activity, and

(d) a labeling means;

wherein each of said sets of biomarker specific primers is used to detect the quantitative expression levels of said unique biomarker in a test subject.