

URINE GENE EXPRESSION RATIOS FOR DETECTION OF CANCER

TECHNICAL FIELD

This invention relates to detection of cancer. Specifically, the invention relates to the use of markers for the detection of bladder cancer. More specifically, this invention relates to use of an under-expressed marker in combination with at least one other marker for the detection of bladder cancer.

BACKGROUND

Survival of cancer patients is greatly enhanced when the cancer is treated early. In the case of bladder cancer, patients diagnosed with early stage disease have 5-year survival rates of >90%, compared to approximately 15-30% for patients diagnosed with advanced disease. Therefore, developments that lead to early diagnosis of bladder cancer can lead to an improved prognosis for the patients. The established method for detecting bladder cancer using urine samples is cytology. However, cytology is known to be only about 75% sensitive for detecting invasive bladder cancer and only about 25% sensitive for detecting superficial bladder cancer (Lotan and Roehrborn, Urology 61, 109-118 (2003)).

Identification of specific markers for cancer in urine can provide a valuable approach for the early diagnosis of cancer, leading to early treatment and improved prognosis. Specific cancer markers also provide a means for monitoring disease progression, enabling the efficacy of surgical, radiotherapeutic and chemotherapeutic treatments to be monitored.

At present, the most reliable method for detecting bladder cancer is cystoscopy accompanied by histology of biopsied lesions. However, this technique is time consuming, invasive and its sensitivity is only approximately 90%, meaning that about 10 percent of cancers are not detected using these methods. Of the non-invasive methodologies, urine cytology, which detects exfoliated malignant cells microscopically, is the current preferred method. Although cytology has a specificity of about 95%, it has poor sensitivity (9-25%) for low-grade lesions, is extremely dependent on sample quality and suffers from high inter-observer variability.

Several urine protein markers are known. Tests for these markers offer better sensitivity than cytology, but tend to suffer from sub-optimal specificity because elevated levels of these markers are also commonly observed in patients with non-malignant diseases including inflammation, urolithiasis and benign prostatic hyperplasia. For example, NMP22, which

detects a specific nuclear matrix protein, has a sensitivity of 47-87% and a specificity of 58-91%.

One drawback associated with urine testing is that individual marker levels can vary significantly with: (i) different urine collection methods (catheterised, voided, urine pellets); (ii) the diurnal timing of urine sampling; (iii) the point of sampling during voiding (e.g. midstream vs end sample); and (iv) urine concentration associated with varying fluid intake, kidney function or diseases that affect plasma volume. These variations have the potential to lead to false positive and false negative tests. Although some of this variation can be reduced using strict standard operating procedures, patient compliance with these procedures can be unreliable. The effect of varying urine concentration can, in some instances, be accounted for by assessing marker levels relative to urinary creatinine, however, this increases the cost and complexity of testing, particularly when sample preparation or storage methods differ for marker detection and creatinine measurement.

There is a need for simple tools for the early detection and diagnosis of cancer. This invention provides further methods, devices and kits based on markers, specifically ratios, regression or classification analysis of bladder cancer markers, to aid in the detection and diagnosis of bladder cancer.

SUMMARY OF THE INVENTION

The present invention provides for a method for determining the presence of a cancer in a subject, comprising:

- (a) providing a sample from the subject;
- (b) detecting the expression level of at least two tumour marker (TM) family members in said sample, wherein at least one TM is an under-expressed TM;
- (c) establishing whether the patient has cancer according to a predetermined threshold.

Step (c) can be preformed by determining the ratio of expression of said TMs, or by performing regression or classification analysis on the TM expression levels.

The TM can be a BTM. The cancer to be detected can be bladder cancer, and in certain embodiments at least one of the TMs is an over-expressed BTM. The over-expressed BTM may be selected from the group outlined in Figure 11 or Figure 12.

- 3 -

In certain embodiments at least one under-expressed TM is a BTM selected from the group outlined in Figure 3 or Figure 4.

5 In other embodiments of the present invention the step of detecting is carried out by detecting over expression of BTM mRNA, a BTM protein, or a BTM peptide.

The sample can be any one of biopsy, blood, serum, peritoneal washes, cerebrospinal fluid, urine and stool samples

10 The present invention also provides for a device for detecting a TM, comprising:
a substrate having a TM capture reagent thereon; and
a detector associated with said substrate, said detector capable of detecting a TM associated with said capture reagent, wherein the TM is an under-expressed TM.

15 The TM can be a BTM.

The TM capture reagent can be an oligonucleotide or an antibody.

20 In certain embodiments the TM can be a BTM selected from the group outlined in Figure 3 or Figure 4.

The present invention also provides for a kit for determining the presence of a cancer in a subject, comprising:
a substrate;
25 at least two TM capture reagents, wherein at least one TM is an under-expressed TM;
and
instructions for use.

30 The TM can be a BTM.

The TM capture reagent may be a TM -specific oligonucleotide or a TM -specific antibody.

35 The TM detected by the kit may be a BTM selected from the group outlined in Figure 3 or Figure 4.

At least one of the TMs detected by the kit may be an over-expressed TM or an over-expressed BTM. The over-expressed BTM may be selected from the group outlined in Figure 11 or Figure 12.

5 BRIEF DESCRIPTION OF THE FIGURES

This invention is described with reference to specific embodiments thereof and with reference to the Figures, in which:

- 10 Figure 1 depicts a table showing the characteristics of urine samples used in the qPCR analyses.
- Figure 2 depicts a table of primers and oligonucleotide probes of markers for qPCR analysis of bladder cancer according to the present invention.
- 15 Figure 3 depicts a table of under-expressing bladder tumour markers identified using microarray methods on samples of bladder cancer.
- Figure 4 depicts a table of under-expressing bladder tumour markers identified using microarray methods on samples of bladder cancer that have insignificant expression in whole blood, but high expression in normal bladder tissue.
- 20 Figure 5 depicts box and whisker plots showing the ratios of three bladder transitional cell carcinoma (TCC) markers (HoxA13, IGFBP5, and MDK) with the under expressing marker LTB4DH for urine samples from patients with either non-malignant urological disease or TCC. The boxes define the 25th, 50th and 75th percentiles and the horizontal bars mark the adjacent values. Outliers are shown by circles. The unfilled boxes represent samples from non-malignant disease controls and the shaded boxes represent samples from patients with TCC.
- 25 Figure 6 shows examples of the sensitivities and specificities of TCC detection for tests that include LTB4DH. (a). single tests; (b). combination tests using LTB4DH and two of the three markers HoxA13, IGFBP5, and MDK.
- 30

Figure 7a-c shows ROC curves for the sensitivity and specificity of detection of TCC in urine samples using ratios that include LTB4DH. 7a. IGFBP5/LTB4DH; 7b. MDK/LTB4DH; 7c. HoxA13/LTB4DH.

Figure 8a-f shows scatter plots for combination tests, a-c using LTB4DH and two of the three markers HoxA13, IGFBP5, and MDK, and d-f repeated using BAG1 for LTB4DH. 8a. MDK/LTB4DH and IGFBP5/LTB4DH; 8b. MDK/LTB4DH and HoxA13/LTB4DH; 8c. IGFBP5/LTB4DH and HoxA13/LTB4DH; 8d. MDK/BAG1 and IGFBP5/BAG1; 8e. MDK/BAG1 and HoxA13/BAG1; 8f. IGFBP5/BAG1 and HoxA13/BAG1.

Figure 9a-b shows scatter plots showing the correlation between ΔCt for IGFBP5 and ΔCt ratios for IGFBP5/LTB4DH and urine creatinine concentration. 9a. Urine samples from patients with TCC 9b. Urine samples from patients with non-malignant disease

Figure 10a-f depicts self-self scatter plots showing the distribution of voided and catheterised urine samples from TCC patients analysed using the bladder tumour markers MDK, IGFBP5 and HoxA13 alone or in ratios with LTB4DH.

Figure 11 shows known over-expressed markers from invasive bladder tumours.

Figure 12 shows known over-expressed markers from superficial bladder tumours.

Figure 13 shows the clinical characteristics of low grade TCC samples and controls used in ROC curve analysis.

Figure 14 shows the results of a ROC Curve analysis. Illustration of the increased test accuracy obtained when LTB4DH is used in ratios with HoxA13 and IGFBP5.

Figure 15 shows the results of a Linear Discriminate Analysis of BTMs, with and without LTB4DH, for the detection of TCC.

DETAILED DESCRIPTION

Definitions

The term "marker" means a molecule that is associated quantitatively or qualitatively with the presence of a biological phenomenon. Examples of "markers" include a polynucleotide, such

as a gene, gene fragment, RNA, or RNA fragment; or a gene product, including a polypeptide such as a peptide, oligopeptide protein or protein fragment; or related metabolites, by products or other identifying molecules, such as antibodies or antibody fragments whether related directly or indirectly to a mechanism underlying the phenomenon.

5 The markers of the invention include the nucleotide sequences (e.g. GenBank sequences) as disclosed herein, in particular the full length sequences, any coding sequences, non-coding sequences and fragments, or any compliments thereof, and any measurable marker thereof as defined above.

10 The term "sensitivity" means the proportion of individuals with the disease who test (by the model) positive. Thus, increased sensitivity means fewer false negative test results.

The term "specificity" means the proportion of individuals without the disease who test (by the model) negative. Thus, increased specificity means fewer false positive test results.

15

The term "expression" includes production of polynucleotides and polypeptides, in particular, the production of RNA (e.g., mRNA) from a gene or portion of a gene, and includes the production of a polypeptide encoded by an RNA or gene or portion of a gene, and includes appearance of a detectable material associated with expression. For example, the formation of a complex, for example, from a polypeptide-polypeptide interaction, polypeptide-nucleotide interaction, or the like, is included within the scope of the term "expression". Another example, the binding of a binding ligand, such as a hybridization probe or antibody, to a gene or other polynucleotide, a polypeptide or a protein fragment and the visualization of the binding ligand. Thus, the density of a spot on a microarray, on a hybridization blot such as a Northern blot, or on an immunoblot, such as a Western blot, or on a bead array, or by PCR analysis, is included within the term "expression" of the underlying biological molecule.

25

The term "over expression" is used where the expression of a marker in one cell, or cell type, is greater than that of another equivalent cell, or cell type.

30

The term "under expression" is used where the expression of a marker in one cell, or cell type, is less than that of another equivalent cell, or cell type.

35

The term "TM" or "tumour marker" or "TM family member" means a marker that is associated with a particular cancer. The term TM also includes combinations of individual markers, whose combination improves the sensitivity and specificity of detecting cancer. It is to be

- 7 -

understood that the term TM does not require that the marker be specific only for a particular tumour. Rather, expression of TM can be altered in other types of cells, diseased cells, tumours, including malignant tumours.

- 5 A TM can be identified by extracting RNA from a tissue sample from a patient suspected of having bladder cancer, applying the RNA or cDNA copy to a microarray having a number of oligonucleotides thereon, permitting the sample RNA to hybridize to the oligonucleotides on the array, and then quantifying the level of measured RNA bound to the each array spot. A marker is considered to be a under expressing TM if its presence is below a threshold of at least about 1.2 times that found in normal, non-malignant tissue using microarray methods. Alternatively, the threshold can be below about 2 times normal, about 3 times less than normal, 4 times or even about 5 times less than normal. By "normal" we mean less than the 90th percentile of the normal population. In other cases, normal can mean a level of presence of the 95th percentile (i.e., about 2 Standard Deviations (SD) from the mean), and in 15 other cases, less than about 97.5th percentile (i.e., about 3 SD) or the 99th percentile.

The term "under expressing TM" means a marker that shows lower expression in bladder tumours than in non-malignant bladder tissue.

- 20 The term "over expressing TM" means a marker that shows higher expression in bladder tumours than in non-malignant tissue.

- The term "BTM" or "bladder tumour marker" or "BTM family member" means a TM that is associated with bladder cancer. The term BTM also includes combinations of individual 25 markers, whose combination improves the sensitivity and specificity of detecting bladder cancer. It is to be understood that the term BTM does not require that the marker be specific only for bladder tumours. Rather, expression of BTM can be altered in other types of cells, diseased cells, tumours, including malignant tumours.

- 30 The term "under expressing BTM" means a marker that shows lower expression in bladder tumours than in non-malignant bladder tissue.

The term "over expressing BTM" means a marker that shows higher expression in bladder tumours than in non-malignant tissue.

The term "qPCR" means quantitative polymerase chain reaction. The term "qPCR" or "QPCR" refers to quantitative polymerase chain reaction as described, for example, in PCR Technique: Quantitative PCR, J.W. Larrick, ed., Eaton Publishing, 1997, and A-Z of Quantitative PCR, S. Bustin, ed., IUL Press, 2004.

5

The term "TCC" means transitional cell carcinoma of the bladder. TCCs constitute ~95% of all bladder cancers.

10

As used herein "antibodies" and like terms refer to immunoglobulin molecules and immunologically active portions of immunoglobulin (Ig) molecules, i.e., molecules that contain an antigen binding site that specifically binds (immunoreacts with) an antigen. These include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, Fc, Fab, Fab', and Fab2 fragments, and a Fab expression library. Antibody molecules relate to any of the classes IgG, IgM, IgA, IgE, and IgD, which differ from one another by the nature of heavy chain present in the molecule. These include subclasses as well, such as IgG1, IgG2, and others. The light chain may be a kappa chain or a lambda chain. Reference herein to antibodies includes a reference to all classes, subclasses, and types. Also included are chimeric antibodies, for example, monoclonal antibodies or fragments thereof that are specific to more than one source, e.g., a mouse or human sequence. Further included are camelid antibodies, shark antibodies or nanobodies.

15
20

25

The terms "cancer" and "cancerous" refer to or describe the physiological condition in mammals that is typically characterized by abnormal or unregulated cell growth. Cancer and cancer pathology can be associated, for example, with metastasis, interference with the normal functioning of neighbouring cells, release of cytokines or other secretory products at abnormal levels, suppression or aggravation of inflammatory or immunological response, neoplasia, premalignancy, malignancy, invasion of surrounding or distant tissues or organs, such as lymph nodes, etc.

30

The term "tumour" refers to all neoplastic cell growth and proliferation, whether malignant or benign, and all pre-cancerous and cancerous cells and tissues.

35

The term "microarray" refers to an ordered or unordered arrangement of capture agents, preferably polynucleotides (e.g., probes) or polypeptides on a substrate. See, e.g., Microarray Analysis, M. Schena, John Wiley & Sons, 2002; Microarray Biochip Technology, M. Schena, ed., Eaton Publishing, 2000; Guide to Analysis of DNA Microarray Data, S.

Knudsen, John Wiley & Sons, 2004; and Protein Microarray Technology, D. Kambhampati, ed., John Wiley & Sons, 2004.

5 The term "oligonucleotide" refers to a polynucleotide, typically a probe or primer, including, without limitation, single-stranded deoxyribonucleotides, single- or double-stranded ribonucleotides, RNA: DNA hybrids, and double-stranded DNAs. Oligonucleotides, such as single-stranded DNA probe oligonucleotides, are often synthesized by chemical methods, for example using automated oligonucleotide synthesizers that are commercially available, or by a variety of other methods, including in vitro expression systems, recombinant techniques,
10 and expression in cells and organisms.

The term "polynucleotide," when used in the singular or plural, generally refers to any polyribonucleotide or polydeoxyribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. This includes, without limitation, single- and double-stranded DNA,
15 DNA including single- and double- stranded regions, single- and double-stranded RNA, and RNA including single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or include single- and double-stranded regions. Also included are triple-stranded regions comprising RNA or DNA or both RNA and DNA. Specifically included are mRNAs, cDNAs, and genomic DNAs, and
20 any fragments thereof. The term includes DNAs and RNAs that contain one or more modified bases, such as tritiated bases, or unusual bases, such as inosine. The polynucleotides of the invention can encompass coding or non-coding sequences, or sense or antisense sequences. It will be understood that each reference to a "polynucleotide" or like term, herein, will include the full-length sequences as well as any fragments, derivatives, or variants
25 thereof.

"Polypeptide," as used herein, refers to an oligopeptide, peptide, or protein sequence, or fragment thereof, and to naturally occurring, recombinant, synthetic, or semi-synthetic molecules. Where "polypeptide" is recited herein to refer to an amino acid sequence of a
30 naturally occurring protein molecule, "polypeptide" and like terms, are not meant to limit the amino acid sequence to the complete, native amino acid sequence for the full-length molecule. It will be understood that each reference to a "polypeptide" or like term, herein, will include the full-length sequence, as well as any fragments, derivatives, or variants thereof.

35 "Stringency" of hybridization reactions is readily determinable by one of ordinary skill in the art, and generally is an empirical calculation dependent upon probe length, washing

temperature, and salt concentration. In general, longer probes require higher temperatures for proper annealing, while shorter probes need lower temperatures. Hybridization generally depends on the ability of denatured DNA to reanneal when complementary strands are present in an environment below their melting temperature. The higher the degree of desired
5 homologous between the probe and hybridisable sequence, the higher the relative temperature which can be used. As a result, it follows that higher relative temperatures would tend to make the reaction conditions more stringent, while lower temperatures less so. Additional details and explanation of stringency of hybridization reactions, are found e.g., in Ausubel et al., Current Protocols in Molecular Biology, Wiley Interscience Publishers, (1995).

10 "Stringent conditions" or "high stringency conditions", as defined herein, typically: (1) employ low ionic strength and high temperature for washing, for example 0.015 M sodium chloride/0.0015 M sodium citrate/0.1% sodium dodecyl sulfate at 50°C; (2) employ a denaturing agent during hybridization, such as formamide, for example, 50% (v/v) formamide
15 with 0.1% bovine serum albumin/0.1% Ficoll/0.1% polyvinylpyrrolidone/50 mM sodium phosphate buffer at pH 6.5 with 750 mM sodium chloride, 75 mM sodium citrate at 42°C; or (3) employ 50% formamide, 5X SSC (0.75 M NaCl, 0.075 M sodium citrate), 50 mM sodium phosphate (pH 6.8), 0.1% sodium pyrophosphate, 5X, Denhardt's solution, sonicated salmon sperm DNA (50 µg/ml), 0.1% SDS, and 10% dextran sulfate at 42°C, with washes at 42°C in
20 0.2X SSC (sodium chloride/sodium citrate) and 50% formamide at 55°C, followed by a high-stringency wash comprising 0.1X SSC containing EDTA at 55°C.

"Moderately stringent conditions" may be identified as described by Sambrook et al., Molecular Cloning: A Laboratory Manual, New York: Cold Spring Harbor Press, 1989, and
25 include the use of washing solution and hybridization conditions (e. g., temperature, ionic strength, and % SDS) less stringent than those described above. An example of moderately stringent conditions is overnight incubation at 37°C in a solution comprising: 20% formamide, 5X SSC (150 mM NaCl, 15 mM trisodium citrate), 50 mM sodium phosphate (pH 7.6), 5X Denhardt's solution, 10% dextran sulfate, and 20 mg/ml denatured sheared salmon sperm
30 DNA, followed by washing the filters in 1X SSC at about 37-50°C. The skilled artisan will recognize how to adjust the temperature, ionic strength, etc. as necessary to accommodate factors such as probe length and the like.

The practice of the present invention will employ, unless otherwise indicated, conventional
35 techniques of molecular biology (including recombinant techniques), microbiology, cell biology, and biochemistry, which are within the skill of the art. Such techniques are explained

fully in the literature, such as, Molecular Cloning: A Laboratory Manual, 2nd edition, Sambrook et al., 1989; Oligonucleotide Synthesis, MJ Gait, ed., 1984; Animal Cell Culture, R.I. Freshney, ed., 1987; Methods in Enzymology, Academic Press, Inc.; Handbook of Experimental Immunology, 4th edition, D .M. Weir & CC. Blackwell, eds., Blackwell Science Inc., 1987; Gene Transfer Vectors for Mammalian Cells, J.M. Miller & M.P. Calos, eds., 1987; Current Protocols in Molecular Biology, F.M. Ausubel et al., eds., 1987; and PCR: The Polymerase Chain Reaction, Mullis et al., eds., 1994.

Description of Embodiments of the Invention

Using a combination of microarray analysis and quantitative polymerase chain reaction (qPCR), markers for transitional cell carcinoma of the bladder (TCC) that are under-expressed in tumours have been identified. It has surprisingly been found that ratios between these markers and other bladder tumour markers (BTM), especially markers that are over-expressed in tumours, are diagnostic for bladder cancer.

The ratios (rather than measuring an absolute level of a marker) identifies a simple gene expression 'signature' that typifies bladder cancer cells, and surprisingly is more robust to variations in sampling techniques or urine concentration. Moreover, the combination of an under-expressed marker and an over-expressed marker maximizes the differential between samples from patients and non-malignant controls, increasing the test reliability. The under-expressed markers described here have been selected on the basis of (i) strong and consistent down-regulation in TCC, (ii) high expression in normal tissue, and (iii) insignificant expression in whole blood to minimize the risk of false positives in patients presenting with hematuria.

As an alternative to determining the ration of the two BTM,s it has also been found that the under-expressed and over-expressed BTMs can be analysed in regression analyses or classification techniques including linear discriminate analysis, and the results of these analyses are also indicative of the presence of bladder cancer.

The test involves the measuring of at least two TM markers, such as a BTM, in a sample from a patient suspected of having a cancer or at risk of having cancer, wherein at least one of the TMs is an under-expressed TM. The ratio of the under-expressed TM and the other TM is indicative of the presence of cancer. The second TM can be any TM as known in the art, but preferably is an over-expressed BTM. Figure 3 shows a number of under-expressed markers suitable for use in the present invention.

The test is best preformed using an under-expressed TM in combination with an over-expressed TM. Any over-expressed TM can be used, for example. Known over expressed BTMs identified from invasive bladder tumours (defined here as tumours $\geq$ stage 1), are outlined in figure 11, and over-expressed BTMs identified from superficial bladder tumours (defined here as Stage Ta and Tis tumours) are shown in figure 12.

It has also been surprisingly established that preferred under-expressed BTMs for use in the present invention are ones that are not significantly elevated in whole blood, and are present in sufficiently high copy numbers in both tumour cells and non-malignant bladder cells. Preferred under-expressed BTMs are outlined in figure 4.

Cancer markers can be detected in a sample using any suitable technique, and can include, but are not limited to, oligonucleotide probes, qPCR or antibodies raised against cancer markers.

It will be appreciated that the sample to be tested is not restricted to a sample of the tissue suspected of being a tumour. The marker may be secreted into the serum, sloughed from cell membranes, released from lysed cells or associated with cells lost into the urine. Therefore, a sample can include any bodily sample, and includes biopsies, blood, serum, peritoneal washes, cerebrospinal fluid, urine and stool samples.

It will also be appreciate that the present invention is not restricted to the detection of cancer in humans, but is suitable for the detection of cancer in any animal, including, but not limited to dogs, cats, horses, cattle, sheep, deer, pigs and any other animal known to get cancer.

General Approaches to Cancer Detection

The following approaches are non-limiting methods that can be used to measure TMs. Following measurement of individual TMs, ratios between high and low expressing BTM family members are determined. These ratios are used to predict the presence or absence cancer.

Alternatively, the high and low expressing TMs are used in regression or classification analyses. The results of these analyses are also used to predict the presence or absence cancer.

General methodologies for determining expression levels are outlined below, although it will be appreciated that any method for determining expression levels would be suitable.

Quantitative PCR (qPCR)

5 Quantitative PCR (qPCR) can be carried out on tumour samples, on serum, plasma and urine samples using BTM specific primers and probes. In controlled reactions, the amount of product formed in a PCR reaction (Sambrook, J., E Fritsch, E. and T Maniatis, Molecular Cloning: A Laboratory Manual 3rd. Cold Spring Harbor Laboratory Press: Cold Spring Harbor (2001)) correlates with the amount of starting template. Quantification of the PCR product
10 can be carried out by stopping the PCR reaction when it is in log phase, before reagents become limiting. The PCR products are then electrophoresed in agarose or polyacrylamide gels, stained with ethidium bromide or a comparable DNA stain, and the intensity of staining measured by densitometry. Alternatively, the progression of a PCR reaction can be measured using PCR machines such as the Applied Biosystems' Prism 7000 or the Roche
15 LightCycler which measure product accumulation in real-time. Real-time PCR measures either the fluorescence of DNA intercalating dyes such as Sybr Green into the synthesized PCR product, or the fluorescence released by a reporter molecule when cleaved from a quencher molecule; the reporter and quencher molecules are incorporated into an oligonucleotide probe which hybridizes to the target DNA molecule following DNA strand
20 extension from the primer oligonucleotides. The oligonucleotide probe is displaced and degraded by the enzymatic action of the Taq polymerase in the next PCR cycle, releasing the reporter from the quencher molecule. In one variation, known as Scorpion®, the probe is covalently linked to the primer.

Reverse Transcription PCR (RT-PCR)

RT-PCR can be used to compare RNA levels in different sample populations, in normal and tumour tissues, with or without drug treatment, to characterize patterns of expression, to discriminate between closely related RNAs, and to analyze RNA structure.

30 For RT-PCR, the first step is the isolation of RNA from a target sample. The starting material is typically total RNA isolated from human tumours or tumour cell lines, and corresponding normal tissues or cell lines, respectively. RNA can be isolated from a variety of samples, such as tumour samples from breast, lung, colon (e.g., large bowel or small bowel), colorectal, gastric, esophageal, anal, rectal, prostate, brain, liver, kidney, pancreas, spleen,
35 thymus, testis, ovary, uterus, bladder etc., tissues, from primary tumours, or tumour cell lines, and from pooled samples from healthy donors. If the source of RNA is a tumour, RNA can

be extracted, for example, from frozen or archived paraffin-embedded and fixed (e.g., formalin-fixed) tissue samples.

5 The first step in gene expression profiling by RT-PCR is the reverse transcription of the RNA template into cDNA, followed by its exponential amplification in a PCR reaction. The two most commonly used reverse transcriptases are avian myeloblastosis virus reverse transcriptase (AMV-RT) and Moloney murine leukaemia virus reverse transcriptase (MMLV-RT). The reverse transcription step is typically primed using specific primers, random hexamers, or oligo-dT primers, depending on the circumstances and the goal of expression
10 profiling. For example, extracted RNA can be reverse-transcribed using a GeneAmp RNA PCR kit (Perkin Elmer, CA, USA), following the manufacturer's instructions. The derived cDNA can then be used as a template in the subsequent PCR reaction.

15 Although the PCR step can use a variety of thermostable DNA-dependent DNA polymerases, it typically employs the Taq DNA polymerase, which has a 5'-3' nuclease activity but lacks a 3'-5' proofreading endonuclease activity. Thus, TaqMan (q) PCR typically utilizes the 5' nuclease activity of Taq or Tth polymerase to hydrolyze a hybridization probe bound to its target amplicon, but any enzyme with equivalent 5' nuclease activity can be used.

20 Two oligonucleotide primers are used to generate an amplicon typical of a PCR reaction. A third oligonucleotide, or probe, is designed to detect nucleotide sequence located between the two PCR primers. The probe is non-extendible by Taq DNA polymerase enzyme, and is labeled with a reporter fluorescent dye and a quencher fluorescent dye. Any laser-induced emission from the reporter dye is quenched by the quenching dye when the two dyes are
25 located close together as they are on the probe. During the amplification reaction, the Taq DNA polymerase enzyme cleaves the probe in a template-dependent manner. The resultant probe fragments disassociate in solution, and signal from the released reporter dye is free from the quenching effect of the second fluorophore. One molecule of reporter dye is liberated for each new molecule synthesized, and detection of the unquenched reporter dye
30 provides the basis for quantitative interpretation of the data.

TaqMan RT-PCR can be performed using commercially available equipment, such as, for example, ABI PRISM 7700 Sequence Detection System (Perkin-Elmer-Applied Biosystems, Foster City, CA, USA), or Lightcycler (Roche Molecular Biochemicals, Mannheim, Germany).
35 In a preferred embodiment, the 5' nuclease procedure is run on a real-time quantitative PCR device such as the ABI PRISM 7700tam Sequence Detection System. The system consists

of a thermocycler, laser, charge-coupled device (CCD), camera, and computer. The system amplifies samples in a 96-well format on a thermocycler. During amplification, laser-induced fluorescent signal is collected in real-time through fibre optics cables for all 96 wells, and detected at the CCD. The system includes software for running the instrument and for
5 analyzing the data.

5' nuclease assay data are initially expressed as Ct, or the threshold cycle. As discussed above, fluorescence values are recorded during every cycle and represent the amount of product amplified to that point in the amplification reaction. The point when the fluorescent
10 signal is first recorded as statistically significant is the threshold cycle.

Real-time Quantitative PCR (qPCR)

A more recent variation of the RT-PCR technique is the real time quantitative PCR, which measures PCR product accumulation through a dual-labeled fluorogenic probe (i.e., TaqMan
15 probe). Real time PCR is compatible both with quantitative competitive PCR and with quantitative comparative PCR. The former uses an internal competitor for each target sequence for normalization, while the latter uses a normalization gene contained within the sample, or a housekeeping gene for RT-PCR. Further details are provided, e.g., by Held et al., Genome Research 6: 986-994 (1996).

20 Expression levels can be determined using fixed, paraffin-embedded tissues as the RNA source. According to one aspect of the present invention, PCR primers and probes are designed based upon intron sequences present in the gene to be amplified. In this embodiment, the first step in the primer/probe design is the delineation of intron sequences
25 within the genes. This can be done by publicly available software, such as the DNA BLAT software developed by Kent, W. J., Genome Res. 12 (4): 656-64 (2002), or by the BLAST software including its variations. Subsequent steps follow well established methods of PCR primer and probe design.

30 In order to avoid non-specific signals, it is useful to mask repetitive sequences within the introns when designing the primers and probes. This can be easily accomplished by using the Repeat Masker program available on-line through the Baylor College of Medicine, which screens DNA sequences against a library of repetitive elements and returns a query
35 sequence in which the repetitive elements are masked. The masked sequences can then be used to design primer and probe sequences using any commercially or otherwise publicly available primer/probe design packages, such as Primer Express (Applied Biosystems);

MGB assay-by-design (Applied Biosystems); Primer3 (Steve Rozen and Helen J. Skaletsky (2000) Primer3 on the WWW for general users and for biologist programmers in: Krawetz S, Misener S (eds) Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ, pp 365-386).

5

The most important factors considered in PCR primer design include primer length, melting temperature (T_m), and G/C content, specificity, complementary primer sequences, and 3' end sequence. In general, optimal PCR primers are generally 17-30 bases in length, and contain about 20-80%, such as, for example, about 50-60% G+C bases. Melting
10 temperatures between 50 and 80°C, e.g., about 50 to 70°C, are typically preferred. For further guidelines for PCR primer and probe design see, e.g., Dieffenbach, C. W. et al., General Concepts for PCR Primer Design in: PCR Primer, A Laboratory Manual, Cold Spring Harbor Laboratory Press, New York, 1995, pp. 133-155; Innis and Gelfand, Optimization of
15 PCRs in: PCR Protocols, A Guide to Methods and Applications, CRC Press, London, 1994, pp. 5-11; and Plasterer, T. N. Primerselect: Primer and probe design. Methods Mol. Biol. 70: 520-527 (1997), the entire disclosures of which are hereby expressly incorporated by reference.

Microarray Analysis

20 Differential expression can also be identified, or confirmed using the microarray technique. Thus, the expression profile of CCPMs can be measured in either fresh or paraffin-embedded tumour tissue, using microarray technology. In this method, polynucleotide sequences of interest (including cDNAs and oligonucleotides) are plated, or arrayed, on a microchip substrate. The arrayed sequences (i.e., capture probes) are then hybridized with
25 specific polynucleotides from cells or tissues of interest (i.e., targets). Just as in the RT-PCR method, the source of RNA typically is total RNA isolated from human tumours or tumour cell lines, and corresponding normal tissues or cell lines. Thus RNA can be isolated from a variety of primary tumours or tumour cell lines. If the source of RNA is a primary tumour, RNA can be extracted, for example, from frozen or archived formalin fixed paraffin-
30 embedded (FFPE) tissue samples and fixed (e.g., formalin-fixed) tissue samples, which are routinely prepared and preserved in everyday clinical practice.

In a specific embodiment of the microarray technique, PCR amplified inserts of cDNA clones are applied to a substrate. The substrate can include up to 1, 2, 5, 10, 15, 20, 25, 30, 35, 40,
35 45, 50, or 75 nucleotide sequences. In other aspects, the substrate can include at least 10,000 nucleotide sequences. The microarrayed sequences, immobilized on the microchip,

are suitable for hybridization under stringent conditions. As other embodiments, the targets for the microarrays can be at least 50, 100, 200, 400, 500, 1000, or 2000 bases in length; or 50-100, 100-200, 100-500, 100-1000, 100-2000, or 500-5000 bases in length. As further embodiments, the capture probes for the microarrays can be at least 10, 15, 20, 25, 50, 75, 80, or 100 bases in length; or 10-15, 10-20, 10-25, 10-50, 10-75, 10-80, or 20-80 bases in length.

Fluorescently labeled cDNA probes may be generated through incorporation of fluorescent nucleotides by reverse transcription of RNA extracted from tissues of interest. Labeled cDNA probes applied to the chip hybridize with specificity to each spot of DNA on the array. After stringent washing to remove non-specifically bound probes, the chip is scanned by confocal laser microscopy or by another detection method, such as a CCD camera. Quantitation of hybridization of each arrayed element allows for assessment of corresponding mRNA abundance. With dual colour fluorescence, separately labeled cDNA probes generated from two sources of RNA are hybridized pairwise to the array. The relative abundance of the transcripts from the two sources corresponding to each specified gene is thus determined simultaneously. An exemplary protocol for this is described in detail in Example 4.

The miniaturized scale of the hybridization affords a convenient and rapid evaluation of the expression pattern for large numbers of genes. Such methods have been shown to have the sensitivity required to detect rare transcripts, which are expressed at a few copies per cell, and to reproducibly detect at least approximately two-fold differences in the expression levels (Skena et al., Proc. Natl. Acad. Sci. USA 93 (2): 106-149 (1996)). Microarray analysis can be performed by commercially available equipment, following manufacturer's protocols, such as by using the Affymetrix GenChip technology, Illumina microarray technology or Incyte's microarray technology. The development of microarray methods for large-scale analysis of gene expression makes it possible to search systematically for molecular markers of cancer classification and outcome prediction in a variety of tumour types.

RNA Isolation, Purification, and Amplification

General methods for mRNA extraction are well known in the art and are disclosed in standard textbooks of molecular biology, including Ausubel et al., Current Protocols of Molecular Biology, John Wiley and Sons (1997). Methods for RNA extraction from paraffin embedded tissues are disclosed, for example, in Rupp and Locker, Lab Invest. 56: A67 (1987), and De Sandres et al., BioTechniques 18: 42044 (1995). In particular, RNA isolation can be performed using purification kit, buffer set, and protease from commercial

manufacturers, such as Qiagen, according to the manufacturer's instructions. For example, total RNA from cells in culture can be isolated using Qiagen RNeasy mini-columns. Other commercially available RNA isolation kits include MasterPure Complete DNA and RNA Purification Kit (EPICENTRE (D, Madison, WI), and Paraffin Block RNA Isolation Kit (Ambion, Inc.). Total RNA from tissue samples can be isolated using RNA Stat-60 (Tel-Test). RNA prepared from tumour can be isolated, for example, by cesium chloride density gradient centrifugation.

The steps of a representative protocol for profiling gene expression using fixed, paraffin-embedded tissues as the RNA source, including mRNA isolation, purification, primer extension and amplification are given in various published journal articles (for example: T. E. Godfrey et al. J. Molec. Diagnostics 2: 84-91 (2000); K. Specht et al., Am. J. Pathol. 158: 419-29 (2001)). Briefly, a representative process starts with cutting about 10 µm thick sections of paraffin-embedded tumour tissue samples. The RNA is then extracted, and protein and DNA are removed. After analysis of the RNA concentration, RNA repair and/or amplification steps may be included, if necessary, and RNA is reverse transcribed using gene specific promoters followed by RT-PCR. Finally, the data are analyzed to identify the best treatment option(s) available to the patient on the basis of the characteristic gene expression pattern identified in the tumour sample examined.

Immunohistochemistry and Proteomics

Immunohistochemistry methods are also suitable for detecting the expression levels of the proliferation markers of the present invention. Thus, antibodies or antisera, preferably polyclonal antisera, and most preferably monoclonal antibodies specific for each marker, are used to detect expression. The antibodies can be detected by direct labeling of the antibodies themselves, for example, with radioactive labels, fluorescent labels, hapten labels such as, biotin, or an enzyme such as horse radish peroxidase or alkaline phosphatase. Alternatively, unlabeled primary antibody is used in conjunction with a labeled secondary antibody, comprising antisera, polyclonal antisera or a monoclonal antibody specific for the primary antibody. Immunohistochemistry protocols and kits are well known in the art and are commercially available.

Proteomics can be used to analyze the polypeptides present in a sample (e.g., tissue, organism, or cell culture) at a certain point of time. In particular, proteomic techniques can be used to assess the global changes of polypeptide expression in a sample (also referred to as expression proteomics). Proteomic analysis typically includes: (1) separation of individual

polypeptides in a sample by 2-D gel electrophoresis (2-D PAGE); (2) identification of the individual polypeptides recovered from the gel, e.g., by mass spectrometry or N-terminal sequencing, and (3) analysis of the data using bioinformatics. Proteomics methods are valuable supplements to other methods of gene expression profiling, and can be used, alone or in combination with other methods, to detect the products of the proliferation markers of the present invention.

Hybridization Methods Using Nucleic Acid Probes Selective for a Marker

These methods involve binding the nucleic acid probe to a support, and hybridizing under appropriate conditions with RNA or cDNA derived from the test sample (Sambrook, J., E Fritsch, E. and T Maniatis, *Molecular Cloning: A Laboratory Manual* 3rd. Cold Spring Harbor Laboratory Press: Cold Spring Harbor (2001)). These methods can be applied to BTM derived from a tumour tissue or fluid sample. The RNA or cDNA preparations are typically labeled with a fluorescent or radioactive molecule to enable detection and quantification. In some applications, the hybridizing DNA can be tagged with a branched, fluorescently labeled structure to enhance signal intensity (Nolte, F.S., *Branched DNA signal amplification for direct quantitation of nucleic acid sequences in clinical specimens*. *Adv. Clin. Chem.* 33, 201-35 (1998)). Unhybridized label is removed by extensive washing in low salt solutions such as 0.1x SSC, 0.5% SDS before quantifying the amount of hybridization by fluorescence detection or densitometry of gel images. The supports can be solid, such as nylon or nitrocellulose membranes, or consist of microspheres or beads that are hybridized when in liquid suspension. To allow washing and purification, the beads may be magnetic (Haukanes, B-I and Kvam, C., *Application of magnetic beads in bioassays*. *Bio/Technology* 11, 60-63 (1993)) or fluorescently-labeled to enable flow cytometry (see for example: Spiro, A., Lowe, M. and Brown, D., *A Bead-Based Method for Multiplexed Identification and Quantitation of DNA Sequences Using Flow Cytometry*. *Appl. Env. Micro.* 66, 4258-4265 (2000)).

A variation of hybridization technology is the QuantiGene Plex® assay (Genospectra, Fremont) which combines a fluorescent bead support with branched DNA signal amplification. Still another variation on hybridization technology is the Quantikine® mRNA assay (R&D Systems, Minneapolis). Methodology is as described in the manufacturer's instructions. Briefly the assay uses oligonucleotide hybridization probes conjugated to Digoxigenin. Hybridization is detected using anti-Digoxigenin antibodies coupled to alkaline phosphatase in colorimetric assays.

Additional methods are well known in the art and need not be described further herein.

Enzyme-Linked Immunological Assays (ELISA)

Briefly, in sandwich ELISA assays, a polyclonal or monoclonal antibody against the BTM is bound to a solid support (Crowther, J.R. The ELISA guidebook. Humana Press: New Jersey (2000); Harlow, E. and Lane, D., Using antibodies: a laboratory manual. Cold Spring Harbor Laboratory Press: Cold Spring Harbor (1999)) or suspension beads. Other methods are known in the art and need not be described herein further. Monoclonal antibodies can be hybridoma-derived or selected from phage antibody libraries (Hust M. and Dubel S., Phage display vectors for the in vitro generation of human antibody fragments. Methods Mol Biol. 295:71-96 (2005)). Non-specific binding sites are blocked with non-target protein preparations and detergents. The capture antibody is then incubated with a preparation of urine or tissue containing the BTM antigen. The mixture is washed before the antibody/antigen complex is incubated with a second antibody that detects the target BTM. The second antibody is typically conjugated to a fluorescent molecule or other reporter molecule that can either be detected in an enzymatic reaction or with a third antibody conjugated to a reporter (Crowther, Id.). Alternatively, in direct ELISAs, the preparation containing the BTM can be bound to the support or bead and the target antigen detected directly with an antibody-reporter conjugate (Crowther, Id.).

Methods for producing monoclonal antibodies and polyclonal antisera are well known in the art and need not be described herein further.

Immunodetection

The methods can also be used for immunodetection of marker family members in sera or plasma from bladder cancer patients taken before and after surgery to remove the tumour, immunodetection of marker family members in patients with other cancers, including but not limited to, colorectal, pancreatic, ovarian, melanoma, liver, oesophageal, stomach, endometrial, and brain and immunodetection of marker family members in urine and stool from bladder cancer patients.

BTMs can also be detected in tissues or urine using other standard immunodetection techniques such as immunoblotting or immunoprecipitation (Harlow, E. and Lane, D., Using antibodies: a laboratory manual. Cold Spring Harbor Laboratory Press: Cold Spring Harbor (1999)). In immunoblotting, protein preparations from tissue or fluid containing the BTM are electrophoresed through polyacrylamide gels under denaturing or non-denaturing conditions. The proteins are then transferred to a membrane support such as nylon. The BTM is then reacted directly or indirectly with monoclonal or polyclonal antibodies as described for

immunohistochemistry. Alternatively, in some preparations, the proteins can be spotted directly onto membranes without prior electrophoretic separation. Signal can be quantified by densitometry.

- 5 In immunoprecipitation, a soluble preparation containing the BTM is incubated with a monoclonal or polyclonal antibody against the BTM. The reaction is then incubated with inert beads made of agarose or polyacrylamide with covalently attached protein A or protein G. The protein A or G beads specifically interact with the antibodies forming an immobilized complex of antibody-BTM-antigen bound to the bead. Following washing the bound BTM can
10 be detected and quantified by immunoblotting or ELISA.

Threshold Determination

- For tests using down-regulated BTMs in either ratios or regression analyses, thresholds will be derived that will enable a sample to be called either positive or negative for TCC. These
15 thresholds will be determined by the analysis of cohorts of patients who are being investigated for the presence of TCC. Thresholds may vary for different test applications; for example, thresholds for use of the test in population screening will be determined using cohorts of patients who are largely free of urological symptoms, and these thresholds may be different to those used in tests for patients who are under surveillance for TCC recurrence, or
20 those being investigated for the presence of urological symptoms such as hematuria. A threshold could be selected to provide a practical level of test specificity in the required clinical setting; that is, a specificity that allows reasonable sensitivity without excessive numbers of patients receiving false positive results. This specificity may be within the range of 80-90%. An alternative method to obtain a test threshold is to plot sensitivity against
25 specificity for different test thresholds (ROC curves) then select the point of inflexion of the curve.

- As an alternative to single thresholds, the test may use test intervals which provide different degrees of likelihood of presence of disease and which have different clinical consequences
30 associated with them. For example, a test may have three intervals; one associated with a high (eg 90%) risk of the presence of TCC, a second associated with a low risk of TCC and a third regarded as being suspicious of disease. The "suspicious" interval could be associated with a recommendation for a repeat test in a defined period of time.

Methods for Detecting Bladder Cancer Markers in Urine

In several embodiments, assays for BTM can be desirably carried out on urine samples. In general, methods for assaying for oligonucleotides, proteins and peptides in these fluids are known in the art. However, for purposes of illustration, urine levels of a BTM can be quantified using a sandwich-type enzyme-linked immunosorbent assay (ELISA). For plasma or serum assays, a 5 μ L aliquot of a properly diluted sample or serially diluted standard BTM and 75 μ L of peroxidase-conjugated anti-human BTM antibody are added to wells of a microtiter plate. After a 30-minute incubation period at 30°C, the wells are washed with 0.05% Tween 20 in phosphate-buffered saline (PBS) to remove unbound antibody. Bound complexes of BTM and anti-BTM antibody are then incubated with o-phenyldiamine containing H₂O₂ for 15 minutes at 30°C. The reaction is stopped by adding 1 M H₂SO₄, and the absorbance at 492 nm is measured with a microtiter plate reader. It can be appreciated that anti-BTM antibodies can be monoclonal antibodies or polyclonal antisera.

Because many proteins are either (1) secreted by cells, (2) cleaved from cell membranes, (3) lost from cells upon cell death or (4) contained within sloughed cells, it will be appreciated that BTMs may also be detected in the urine. Additionally, diagnosis of bladder cancer can be determined by measuring either expression of BTMs in a sample, or accumulation of BTMs in a sample. Prior art methods of diagnosis include cystoscopy, cytology and examination of cells extracted during these procedures. Such methods have relied upon identification of tumour cells in the urine or in a brush sample of urothelium, or in other cases, in biopsy specimens of the bladder wall. These methods suffer from several types of errors, including sampling error, errors in identification between observers, and the like.

Antibodies to Bladder Tumour Markers

In additional aspects, this invention includes manufacture of antibodies against BTMs. Using methods described herein, novel BTMs can be identified using microarray and/or qPCR methods. Once a putative marker is identified, it can be produced in sufficient amount to be suitable for eliciting an immunological response. In some cases, a full-length BTM can be used, and in others, a peptide fragment of a BTM may be sufficient as an immunogen. The immunogen can be injected into a suitable host (e.g., mouse, rabbit, etc) and if desired, an adjuvant, such as Freund's complete adjuvant, Freund's incomplete adjuvant can be injected to increase the immune response. It can be appreciated that making antibodies is routine in the immunological arts and need not be described herein further. As a result, one can produce antibodies against BTMs identified using methods described herein.

In yet further embodiments, antibodies can be made against the protein or the protein core of the tumour markers identified herein or against an oligonucleotide sequence unique to a BTM. Although certain proteins can be glycosylated, variations in the pattern of glycosylation can, in certain circumstances, lead to mis-detection of forms of BTMs that lack usual glycosylation patterns. Thus, in certain aspects of this invention, BTM immunogens can include deglycosylated BTM or deglycosylated BTM fragments. Deglycosylation can be accomplished using one or more glycosidases known in the art. Alternatively, BTM cDNA can be expressed in glycosylation-deficient cell lines, such as prokaryotic cell lines, including *E. coli* and the like.

Vectors can be made having BTM-encoding oligonucleotides therein. Many such vectors can be based on standard vectors known in the art. Vectors can be used to transfect a variety of cell lines to produce BTM-producing cell lines, which can be used to produce desired quantities of BTM for development of specific antibodies or other reagents for detection of BTMs or for standardizing developed assays for BTMs.

Kits

Based on the discoveries of this invention, several types of test kits can be envisioned and produced. First, kits can be made that have a detection device pre-loaded with a detection molecule (or "capture reagent"). In embodiments for detection of BTM mRNA, such devices can comprise a substrate (e.g., glass, silicon, quartz, metal, etc) on which oligonucleotides as capture reagents that hybridize with the mRNA to be detected is bound. In some embodiments, direct detection of mRNA can be accomplished by hybridizing mRNA (labeled with cy3, cy5, radiolabel or other label) to the oligonucleotides on the substrate. In other embodiments, detection of mRNA can be accomplished by first making complementary DNA (cDNA) to the desired mRNA. Then, labeled cDNA can be hybridized to the oligonucleotides on the substrate and detected.

Antibodies can also be used in kits as capture reagents. In some embodiments, a substrate (e.g., a multiwell plate) can have a specific BTM capture reagent attached thereto. In some embodiments, a kit can have a blocking reagent included. Blocking reagents can be used to reduce non-specific binding. For example, non-specific oligonucleotide binding can be reduced using excess DNA from any convenient source that does not contain BTM oligonucleotides, such as salmon sperm DNA. Non-specific antibody binding can be reduced using an excess of a blocking protein such as serum albumin. It can be appreciated that numerous methods for detecting oligonucleotides and proteins are known in the art, and any

strategy that can specifically detect BTM associated molecules can be used and be considered within the scope of this invention.

5 In addition to a substrate, a test kit can comprise capture reagents (such as probes), washing solutions (e.g., SSC, other salts, buffers, detergents and the like), as well as detection moieties (e.g., cy3, cy5, radiolabels, and the like). Kits can also include instructions for use and a package.

BTM ratios Used for Detection of Bladder Cancer I

10 In one series of embodiments, reagents for the testing the BTM LTBDH4 in combination with over-expressing BTMs can be incorporated into a kit for the testing of unfractionated urine or urine cell sediments to detect bladder cancer. The urine samples could be collected from patients with diagnosed bladder cancer who require monitoring for disease progression or treatment response, individuals with urological symptoms including macroscopic or
15 microscopic hematuria, or asymptomatic individuals. For patients or individuals being tested with a kit that measures the BTMs in unfractionated urine, approximately 2mls of urine can be taken for testing. For tests on the urine pellet, >10mls of urine can be collected.

A suitable kit includes: (i) instructions for use and result interpretation, (ii) software for
20 interpretation of multiple gene analyses, including any regression analysis classifier or formula (iii) reagents for the stabilization and purification of RNA from unfractionated urine or urine pellets, (iv) reagents for the synthesis of cDNA including dNTPs and reverse transcriptase, and (v) reagents for the quantification of the BTM cDNA. In one form, these reagents would be used for quantitative PCR and would include specific exon-spanning
25 oligonucleotide primers, a third oligonucleotide labeled with a probe for detection, Taq polymerase and the other buffers, salts and dNTPs required for PCR. The kit can also use other methods for detection of the transcripts such as direct hybridization of the BTM RNA with labeled probes or branched DNA technology; and (vi) oligonucleotides and probe for the detection of transcripts from a highly transcribed gene, such as β actin, to serve as a quality
30 control measure.

Evaluation of Progression of Bladder Cancer Using BTM ratios

To evaluate the progression of bladder tumours, samples of tissue are obtained by biopsy of bladder wall or samples of urine are collected over time from a patient having bladder cancer.
35 Evaluation of the ratio of BTMs or combinations thereof are made for samples taken at

different times. BTM ratios within a specified range are indicative of progression of bladder cancer.

Evaluation of Therapy of Bladder Cancer Using BTM ratios

- 5 To evaluate the efficacy of therapy for bladder tumours, samples of tissue and/or urine are obtained before treatment is initiated. The baseline levels of one or more BTMs are determined, as are ratios of various BTMs with respect to each other. Treatment is initiated, and can include any therapy known in the art, including surgery, radiation therapy or chemotherapy as appropriate to the type and stage of the disease. During the course of
- 10 therapy, samples of tissue and/or urine are collected and analyzed for the presence and amount of BTMs. Ratios of various BTMs are determined and results are compared to: (1) the patient's baseline levels before treatment or (2) normal values obtained from a population of individuals not having bladder cancer.

15 Use of BTM Ratios to Monitor the Progression of TCC Therapies

- In addition to the rapid diagnosis and early detection of TCC, BTM marker ratios detected in either tissue, serum or urine can be used to monitor a patient's response to therapy. In these applications, urine and/or serum samples can be taken at intervals following the initiation of systemic, intravesicular or intravascular chemotherapy, radiotherapy or immunotherapy. A
- 20 change in marker ratio can indicate a reduction in tumour size, indicative of effective treatment. The rate of change can be used to predict the optimum therapeutic dose for each patient or treatment.

Use of BTM Regression Analyses

- 25 In addition to the BTM ratios, regression or classification analyses that include high and low expressing BTM family members can be used for the applications described above.

- Markers evaluated are selected from known human genes. The genes evaluated are indicated in Figures 3 and 4. Included in Figures 3 and 4 are the name of the gene, the
- 30 HUGO identifier, MWG oligo number, NCBI mRNA reference sequence number and the protein reference number. The full length sequences can be found at <http://www.ncbi.nlm.nih.gov/entrez/>.

EXAMPLES

- 35 The examples described herein are for purposes of illustrating embodiments of the invention and are not intended to limit the scope of the invention. Other embodiments, methods and

types of analyses are within the scope of persons of ordinary skill in the molecular diagnostic arts and need not be described in detail hereon. Other embodiments within the scope of the art that are based on the teachings herein are considered to be part of this invention.

5 **Methods**

Tumour Collection

Bladder tumour samples and non-malignant urothelium samples were collected from surgical specimens resected at Kyoto University Hospital, Japan.

10 **Urine Collection**

Urine samples from non-malignant controls and bladder cancer patients were obtained from Kyoto University Hospital, Japan. Healthy control samples were obtained from Japanese volunteers (figure 1).

15 **RNA Extraction**

Tumour tissues were homogenized in a TriReagent: water (3:1) mix, then chloroform extracted. Total RNA was then purified from the aqueous phase using the RNeasy™ procedure (Qiagen). RNA was also extracted from 16 cancer cell lines and pooled to serve as a reference RNA.

20

RNA was extracted from urine by mixing the urine sample with an equal volume of lysis buffer (5.64M guanidine-HCl, 0.5% sarkosyl, 50mM sodium acetate (pH 6.5) and 1mM β-mercaptoethanol; pH adjusted to 7.0 with 1.5M Hepes pH 8). Due to the low amounts of RNA in urine, 7.5ug of total bacterial RNA was added to the urine/lysis buffer mix to act as a carrier. Total RNA was then extracted using Trizol and the RNeasy™ procedure. RNA preparations were further purified prior to cDNA synthesis using the Qiagen QIAquick™ PCR purification kit.

25

RNA was extracted from the blood of three healthy volunteers by performing a Trizol/RNeasy™ extraction on cells enriched from whole blood using sedimentation in 3.6% dextran.

30

Microarray Slide Preparation

Epoxy coated glass slides (MWG Biotech) were printed with ~30,000 50mer oligonucleotides (MWG Biotech) using a Gene Machines microarraying robot, according to the manufacturer's protocol.

35

RNA Labeling and Hybridization

cDNA was transcribed from 5µg total RNA using Superscript II™ reverse transcriptase (Invitrogen) in reactions containing 5-(3-aminoallyl)- 2' deoxyuridine -5'-triphosphate. The reaction was then de-ionised in a Microcon column before being incubated with Cy3 or Cy5 in bicarbonate buffer for 1 hour at room temperature. Unincorporated dyes were removed using a Qiaquick column (Qiagen) and the sample concentrated to 15 µl in a SpeedVac. Cy3 and Cy5 labeled cDNAs were then mixed with Ambion ULTRAhyb™ buffer, denatured at 100°C for 2 min and hybridized to the microarray slides in hybridisation chambers at 42°C for 16 hours. The slides were then washed and scanned twice in an Axon 4000A™ scanner at two power settings.

Microarray Analysis of Cancer Marker Genes

RNA from 53 bladder tumours and 20 non-malignant ("normal") bladder tissue samples were labeled with Cy5 and hybridized in duplicate or triplicate with Cy3 labeled reference RNA. After normalization, the change in expression in each of 29,718 genes was then estimated by fold change and statistical probability.

Normalisation Procedure

Median fluorescence intensities detected by Genepix™ software were corrected by subtraction of the local background intensities. Spots with a background corrected intensity of less than zero were excluded. To facilitate normalization, intensity ratios and overall spot intensities were log-transformed. The logged intensity ratios were corrected for dye and spatial bias using local regression implemented in the LOCFIT™ package. Logged intensity ratios were regressed simultaneously with respect to overall spot intensity and location. The residuals of the local regression provided the corrected logged fold changes. For quality control, ratios of each normalized microarray were plotted in respect to spot intensity and localization. The plots were subsequently visually inspected for any remaining artifacts. Additionally, an ANOVA model was applied for the detection of pin-tip bias. All results and parameters of the normalization were inserted into a Postgres-database for statistical analysis.

Statistical Analysis

To improve the comparison of measured fold changes between arrays, log2 (ratios) were scaled to have the same overall standard deviation per array. This standardization reduced the average within-tissue class variability. A rank-test based on fold changes was then used to improve the noise robustness. This test consists of two steps: (i) calculation of the rank of

fold change (Rfc) within arrays and ii) subtraction of the median (Rfc) for normal tissue from the median (Rfc) for tumour tissue. The difference of both median ranks defines the score of the fold change rank. Three additional statistical tests were also performed on standardized data: 1) Two sample student's t-test, 2) the Wilcoxon test and 3) Statistical Analysis of
5 Microarrays (SAM). The most significantly down-regulated genes determined by each of the statistical methods (rank fold change, t-test, Wilcoxon test, and SAM) were given a rank score for each test. All rank scores were then added into one summated rank score.

cDNA Synthesis from Urine RNA

10 Total urine RNA was annealed to gene-specific primers for each of the bladder tumour markers by incubating at 70°C then cooling on ice for 2 mins in 50ul reactions containing forward primers at 0.01µg/µl. Each cDNA reaction contained annealed RNA and 4µl of 5x Superscript II reverse transcriptase buffer (Invitrogen, USA), 2µl of 0.1M DTT (Invitrogen, USA), 0.5µl of RNase out (40U/µL), (Invitrogen, USA), 4µl of 10 mM dNTP (Invitrogen, USA)
15 and 0.5µl of Superscript II reverse transcriptase (200U/µl), (Invitrogen, USA) in a final volume of 20µl. Reactions were incubated at 42°C for 1 hour, 10 minutes at 70°C and 1 minute at 80°C. Reactions were cleaned prior to qPCR with Qiagen QIAquick PCR purification columns (Qiagen, Victoria, Australia) and stored at -80°C.

20 Quantitative Real-Time PCR

Real-time or quantitative PCR (qPCR) is used for absolute or relative quantitation of PCR template copy number. Taqman™ probe and primer sets were designed using Primer Express V 2.0™ (Applied Biosystems). Where possible, all potential splice variants were included in the resulting amplicon, with amplicon preference given to regions covered by the
25 MWG-Biotech-derived microarray oligonucleotide. Primer and probe sequences are shown in Figure 2. Alternatively, if the target gene was represented by an Assay-on-Demand™ expression assay (Applied Biosystems) covering the desired amplicons, these were used. In the in-house designed assays, primer concentration was titrated using a SYBR green labeling protocol and cDNA made from the reference RNA. Amplification was carried out on
30 an ABI Prism™ 7000 sequence detection system under standard cycling conditions. When single amplification products were observed in the dissociation curves, standard curves were generated over a 625 fold concentration range using optimal primer concentrations and 5'FAM - 3'TAMRA phosphate Taqman™ probe (Proligo) at a final concentration of 250 nM. Assays giving standard curves with regression coefficients over 0.98 were used in
35 subsequent analyses.

Assays were performed in 96 well plates. Each plate contained a reference cDNA standard curve, over a 625-fold concentration range. For the urine qPCR, total RNA extracted from ~0.5mls unfractionated urine was used in each reaction. The ΔCt (target gene Ct – mean reference cDNA Ct) was calculated for each marker, and used in subsequent ratios, regression or classification analysis.

Expression of Markers in Blood

The expression of the markers shown in figures 3 and 4 in whole blood was determined *in silico*. Microarray probes were linked to UniGene clusters via the GenBank accession numbers of their target mRNAs, and the tissue expression profile from UniGene used to determine the number of expressed sequence tags (ESTs) in blood libraries. Only genes with 0 or 1 expressed sequence tags (EST) are shown in figure 4. To confirm the low expression of LTB4DH in whole blood, RT-qPCR was carried out on total RNA extracted from whole blood using the primers and probes shown in Figure 2. No significant expression was observed (results not shown).

Identification of Down Regulated Bladder Cancer Markers

To identify down-regulated markers of bladder cancer, we performed microarray studies on RNA from 53 bladder tumours and 20 non-malignant bladder tissue samples using 30,000 oligonucleotide chips. Figure 3 shows the statistical analysis of microarray data for 300 genes that show significant downregulation in bladder cancer tissue compared to non-malignant tissue. Figure 3 includes the HUGO gene name and symbol, the protein reference sequence number, the NCBI mRNA reference sequence number, the MWG Biotech probe oligonucleotide number, the median fold change in gene expression between tumour and non-malignant tissue, the results of an original unadjusted Student's t-test, the results of the 2-sample Wilcoxon test, the results of the SAM test, and the summated rank score.

Identification of Preferred Under-Expressed Bladder Tumour Markers for Use in Urine Tests for Bladder Cancer

Because urinary hematuria is a common co-occurrence with bladder cancer, it is an advantage that bladder cancer markers are not significantly elevated in whole blood. In addition, because the downregulated markers are being used in ratios, regression or classification analysis, it is an advantage that they be present in sufficiently high copy numbers in both tumour cells and non-malignant bladder cells to enable reliable detection in urine. To identify suitable markers, we screened the genes in Figure 3 for a subset that had little or no representation in blood EST libraries, and had higher than median expression in

non-malignant tissue. Median expression was estimated by ranking the 30,000 oligonucleotides on the array by their median intensity in the samples analysed in the microarray study. Markers that met the criteria are shown in Figure 4. Figure 4 includes the HUGO gene name and symbol, the protein reference sequence number, the NCBI mRNA reference sequence number, the median fold change, the rank score, the median rank of microarray spot intensity in tumour tissue and non-malignant tissue, and the number of ESTs present in blood EST libraries.

The down regulation observed in the array data was validated by qPCR for three genes shown in Figure 4, LTB4DH, BAG1 and FLJ21511. These genes were tested on total RNA from 10 tumour samples and 10 non-malignant samples. LTB4DH, BAG1 and FLJ21511 showed an average downregulation in bladder tumours compared to bladder non-malignant tissue of 2.5 fold, 1.4 fold and 6.1 fold, respectively, in these samples.

qPCR Analysis of Urine using LTB4DH

Urine from TCC patients and controls with non-malignant urological conditions was collected by either voiding or catheterisation. Total RNA was extracted from the voided urine of 42 controls and the voided or catheterised urine of 37 TCC patients and used in quantitative RT-PCR using primers and probes for LTB4DH and three over-expressed markers, IGFBP5, MDK and HoxA13. The ΔC_t ratios were determined for IGFBP5/LTB4DH, MDK/LTB4DH and HoxA13/LTB4DH. This data is illustrated by the box plots in Figure 5, which show a clear difference in the spread of data between the urine samples from controls and TCC patients for each of the three tests. The most accurate test was IGFBP5/LTB4DH which demonstrated sensitivity and specificity of 87% and 88% in this sample cohort, respectively (Fig. 6a). To illustrate the correspondence between sensitivity and specificity for each of these tests, ROC curves are shown in Figure 7. The areas under the curve for IGFBP5/LTB4DH, MDK/LTB4DH and HoxA13/LTB4DH are 0.9223, 0.9199, and 0.7497, respectively. These areas, which measure test accuracy, indicate that all three ratios with LTB4DH are useful tests, in particular IGFBP5/LTB4DH and MDK/LTB4DH.

To increase the sensitivity and specificity of TCC detection, combinations of two tests were used. The optimal sensitivities and specificities of these test combinations are shown in Figure 6b. Figure 8a-f shows the separation of data in 2 dimensional space for each of the three tests using LTB4DH and BAG1. This data shows that combinations of two or more tests that include either of the downregulated BTMs LTB4DH or BAG1, are able to achieve sensitivities and specificities of over 90%. Moreover, because these tests are measuring

simple gene expression signatures and not absolute levels of markers, they will be robust to variations in urine concentration.

To demonstrate the robustness of tests involving ratios with LTB4DH to urine concentration, the levels of IGFBP5 alone (Δ Ct) and IGFBP5/LTB4DH were plotted as a function of urine concentration (Figure 9a-b) and trendlines fitted to the data. It can be seen that for both urine samples from non-malignant controls and patients with TCC, there is a decrease in the IGFBP5 Δ Ct with increasing urine concentration that is absent in the IGFBP5/LTB4DH ratio. The effect is most pronounced with the non-malignant samples because of the absence of other influences such as tumour size and tumour heterogeneity in the expression of IGFBP5 and LTB4DH.

In some instances, when single markers are used in bladder cancer assays, the method of urine sample collection can affect the amount of marker detected due to variations in the number of exfoliated bladder cells collected. This bias could lead to false positive or false negative results in a small proportion of samples. The use of ratios including LTB4DH or other low-expressing genes should provide a method to compensate for different methods. To test this hypothesis, samples collected from TCC patients by either simple voiding (nine samples) or catheterisation (28 samples) were tested for the presence of TCC markers and LTB4DH. Analysis of the TCC markers alone showed that the voided samples were more heavily represented at the lower end of the range of data (higher Ct), consistent with a lower average number of exfoliated cells in these samples compared to the catheterised samples. This is illustrated in the self-self scatter plots for IGFBP5, MDK and HoxA13 in figure 10a-c. In contrast, when ratios between these markers and LTB4DH were calculated, the voided and catheterised samples were spread over similar ranges of Ct ratios (figure 10d-f), illustrating that the calculation of gene expression signatures between high expressing markers and low expressing markers such as LTB4DH compensate for variations in marker levels introduced by different urine sampling methodologies.

Urine samples from patients with low grade tumours are often borderline in their accumulation of BTMs due to the presence of only small numbers of exfoliated cells in these samples. These samples are therefore at high risk of being incorrectly classified due to variations in sampling method or urine concentration.

The utility of gene expression ratios that incorporate down-regulated genes for the detection of TCC is therefore likely to be pronounced when applied to the detection of low grade TCC.

To demonstrate this effect, a cohort of voided 43 urine samples from patients with low grade TCC and 123 controls were tested with the markers IGFBP5, HoxA13 and LTB4DH. The clinical characteristics of the cohort are summarised in Fig. 13. The qPCR data for IGFBP5 and HoxA13 were analysed alone and in ratios with LTB4DH using the area under the ROC curve as a measure of test accuracy (STATA statistics package). The results are summarized in Fig. 14. Using the IGFBP5 marker, LTB4DH increased the accuracy of detection of low grade (grade 1-2) stage Ta TCCs by 9% and low grade TCCs of any stage by 8%. The accuracy of HoXA13 testing of low grade stage Ta TCCs was increased by 3%.

Linear Discriminate Analysis of qPCR Data using LTB4DH

Linear discriminate analysis (LDA) is a statistical technique (Fisher R.A. "The Use of Multiple Measurements in Taxonomic Problems", Annals of Eugenics 7 179 (1936)) in which a linear combination of variables is generated, such that there is maximal separation between two or more groups. This linear combination of variables is termed the "linear discriminant", which is a linear function of the input variables that maximally separates the classes of the data set. The ability of LDA (or any other classification technique) to characterise a particular dataset, such as qPCR data, can be tested using cross-validation. In this method, part of the dataset is used to generate a classifier, and part of the dataset is used to measure the effectiveness of that classifier. The partitioning of the dataset into training and testing sets can be repeated multiple times (each time generating a new classifier). In k-fold cross-validation, the dataset is split k-wise, and each subset is used as the testing set in one of k rounds of training and validation. This can be extended to leave-one-out cross-validation (LOOCV) where each sample is classified according to a classifier generated from the remaining samples in the dataset ("leaving one out"; leaving out the sample which is being tested).

LDA and LOOCV were used to illustrate the utility of the downregulated BTM, LTB4DH, in improving the diagnosis of TCC. qPCR was first carried out on the cohort of control and TCC urine samples described in Figure 13 which were supplemented with an additional 30 grade 3 tumours (5 > stage 1, 13 =stage 1, 4=Tis, and 8=Ta). Combinations of the six genes LTB4DH, MDK, IGFBP5, HOXA13, TOP2a and CDC2 were tested for classifier performance, as judged by LOOCV. The posterior probability (that the sample "left out" was a TCC sample) was used to generate ROC curves using the ROCR package of the R statistical programming environment. The sensitivity of the classifier for a given specificity was obtained by reference to the appropriate ROC curve.

The sensitivity of detection of TCC using combinations of upregulated BTMs with and without LTB4DH was determined at a specificity of 85%. The results of this analysis are shown in Figure 15. It can be seen that the addition of LTB4DH to assays including combinations of the upregulated BTMs MDK, IGFBP5, Top2a, cdc2 and HoxA13 increased the overall
5 sensitivity by 1-2% and the sensitivity of detection of Stage Ta tumours, grade 1-2 tumours and grade 3 tumours by up to 3%.

Wherein in the foregoing description reference has been made to integers or components having known equivalents, such equivalents are herein incorporated as if individually set
10 fourth.

Although the invention has been described by way of example and with reference to possible embodiments thereof, it is to be appreciated that improvements and/or modifications may be made without departing from the scope or the spirit thereof.

15 INDUSTRIAL APPLICABILITY

Methods for detecting BTM family members include detection of nucleic acids, proteins and peptides using microarray and/or real time PCR methods. The compositions and methods of this invention are useful in diagnosis of disease, evaluating efficacy of therapy, and for
20 producing reagents and test kits suitable for measuring expression of BTM family members.

CLAIMS

1. A method for determining the presence of a cancer in a subject, comprising:
 - (a) providing a sample from the subject;
 - (b) detecting the expression level of at least two TM family members in said
 - 5 sample, wherein at least one TM is an under-expressed TM;
 - (c) establishing whether the patient has cancer according to a predetermined threshold, wherein the predetermined threshold is established by determining the ratio of expression of said TMs, or by performing regression or classification analysis on the TM expression values.
- 10 2. The method of claims 1, wherein the TM is a BTM.
3. The method of claim 1, wherein said cancer is bladder cancer.
- 15 4. The method of claim 1, wherein at least one of the TMs is an over-expressed TM.
5. The method according to claim 1, wherein the at least one under-expressed TM is a BTM selected from the group outlined in Figure 3.
- 20 6. The method according to claim 1, wherein the at least one under-expressed TM is a BTM selected from the group outlined in Figure 4.
7. The method according to claim 4, wherein the over-expressed TM is BTM selected from the group outlined in Figure 11 or Figure 12.
- 25 8. The method of claim 1, wherein said step of detecting is carried out by detecting expression of TM mRNA.
9. The method of claim 1, wherein said step of detecting is carried out by detecting
- 30 expression of a TM protein.
10. The method of claim 1, wherein said step of detecting is carried out by detecting expression of a TM peptide.
- 35 11. The method of claim 1, wherein the sample is any one of biopsy, blood, serum, peritoneal washes, cerebrospinal fluid, urine and stool samples.

12. The method of claim 5, wherein the under exposed TM is selected from the group LTB4DH and BAG1.

13. The method of claim 7, wherein the over exposed TM is selected from the group
5 CDC2, MDK, HOXA13 and IGFBP5.