TRIAGING OF PATIENTS HAVING ASYMPTOMATIC HEMATURIA USING GENOTYPIC AND PHENOTYPIC BIOMARKERS

Claim of Priority

This application claims priority to United States Provisional Patent Application No. 61/907,013 filed 21 November 2013; Inventors David Darling, Satish Kumar, Mark Dalphin, and Paul O'Sullivan. This Provisional Application is herein incorporated fully by reference.

Field of the Invention

This invention relates to the detection of patients not having disease. Specifically, this invention relates to the use of genetic markers and phenotypic markers for triaging patients that present with hematuria without cancer. Particularly, this invention relates to analysis of genetic markers and phenotypic markers in triaging patients with either macroscopic or microscopic hematuria. More particularly, this invention relates to use of genetic and phenotypic markers in combination to triage patients with asymptomatic macroscopic or microscopic hematuria and to predict whether a patient's condition warrants further clinical procedures.

BACKGROUND

Survival of cancer patients is greatly enhanced when the cancer is treated early. In the case of bladder cancer, patients diagnosed with disease that is confined to the primary site have a 5 year survival rate of 73%, compared to 6% for patients with metastatic disease (Altekruse et al). Therefore, developments that lead to early and accurate diagnosis of bladder cancer can lead to an improved prognosis for the patients. To aid in early detection of cancer a number of cancer specific markers have been identified. However the use of these markers can result in false positive results in patients having inflammatory bladder diseases, and not bladder cancer.

Asymptomatic hematuria ("AH") is one of the most frequent urological findings, with incidence rates of between 2% and 30% depending on the population (Schwartz G: Proper evaluation of asymptomatic microscopic hematuria in the era of evidence-based medicine-progress is being made. Mayo Clin Proc. 2013, 88(2); 123-125, McDonald M, Swagerty D, Wetzel L: Assessment of Microscopic hematuria in adults. AFP 2006 73:10, Grossfield G,

Wolf J, Litwan M, Hricak H, Shuler C, Agerter D, et al. Asymptomatic microscopic hematuria in adults: summary of AUA best practice policy recommendations. AFP 2001:63:1145-54).

AH is, however, indicative of broad range of pathologies with urinary tract malignancy incidences in the AH population ranging from 1.9-7%. Full diagnostic work up on all confirmed AH patients puts a considerable burden on many healthcare systems. Use of phenotypic indicators to segregate high and low risk patients has been explored in a recent study by Loo et al. (Loo R, Lieberman S, Slezak J, Landa H, Mariani A, Nicolaisen G, Aspera A and Jaconsen S: Stratifying risk of urinary tract malignant tumors in patients with asymptomatic microscopic hematuria. Mayo Clin Proc. 2013, 88(2); 129-138).

The above-mentioned study of 4414 patients presenting with confirmed AH showed that 73% of patients had no cause identified, while 26% of patients warranted some form of urological work up to identify the cause. Approximately 2.5% of patients presenting with AH were diagnosed with urothelial malignancy, with other conditions such as urinary tract infection (UTI) (2.3%), kidney stones (16.2%), prostatic bleeding (4%), and contamination (0.4%) making up the alterative diagnoses (Loo et al., Id.).

SUMMARY

We have identified a new problem in the field, namely how to identify patients presenting with hematuria who do not have or are at low risk for having bladder cancer. This solves the problem that many patients with hematuria and without bladder cancer may undergo expensive and invasive further workup when such workups are not needed. Thus, this invention is useful to exclude individuals from the hazards and costs associated with full work-up for bladder cancer when a combination of genetic information and phenotypic information provides identification of patients that do not have, or are at low risk of having bladder cancer, and to effectively triage patients having no cancer, from those having cancerous conditions, including urothelial carcinomas, transitional cell carcinoma (TCC) and non-cancerous conditions, including inflammatory disease. This invention represents a new approach to a new problem, in that it is unexpectedly useful, not to diagnose cancer, but rather to diagnose non-cancers. The use of combinations of genetic and phenotypic criteria provide unexpectedly better discrimination than either genetic or phenotypic variables alone. Data were obtained from 541 observations from validated under CLIA standards and CURT +

North Shore product trial using bootstrap procedures for internal validation. Phenotypic variables included presence of: (1) smoking history, (2) hematuria, (3) gender, and (4) age. Genetic variables included analysis of expression of IGF, HOXA13, MDK, CDC, and IL8R. The genetic + phenotypic model ("G+P") performed unexpectedly better than either genetic or phenotypic variables alone.

Macrohematuria, or finding of visually identified blood in the urine is a common finding in patients with bladder cancer. For those patients, it is often standard practice to perform additional diagnostic procedures to diagnose bladder cancer. However, readily identifying patients with microhematuria and understanding the implications of microhematuria in urothelial carcinomas, remained a problem.

We herein provide improved methods for determining whether a patient presenting with either macrohematuria or microhematuria could avoid invasive and expensive further clinical procedures to detect urothelial carcinomas (UC) including bladder cancer, if such patient is at a sufficiently low probability of having bladder cancer to warrant not carrying out additional procedures.

Factors attributable to high probability of urothelial carcinoma (UC) are described. Demographic factors such as gender, race and age in addition to environmental factors such as smoking history and occupational exposure to aromatic amines contribute significantly to the risk of developing UC. Characterization of patients in healthcare assessment based on these factors is used routinely on an ad-hoc basis. For example, it is well accepted that a 60 year old male with smoking history presenting with hematuria has a higher probability of being positive for UC than a 35 year old non-smoking female presenting with the same symptoms, however these differences have not been quantitated to contribute to the overall probability of the patient having UC. Attribution of specific weights to various genotypic and phenotypic factors and combining these with a diagnostic test output can add significantly to the accuracy of the diagnostic power of non-invasive tests and provide clinicians with greater certainty in segregating patients on the basis of their probability of having UC as defined by the clinical and biomarker test results.

Although there are methods available to detect the presence of bladder cancer, there are no reliable and accurate methods to determine whether a patient does not have, or is at low risk of having bladder cancer. To address this need, we have developed new analytical methods for distinguishing between cancerous conditions from non-cancerous ones in patients

presenting with hematuria, either macrohematuria or microhematuria. In some aspects of this invention, we combine quantified phenotypic variables and quantified expression of genetic markers to form a combined segregation index (the "G + P INDEX") in order to effectively triage out AH patients with a low probability of having UC from those AH patients that have high probability of UC. This segregation defines those that don't need a complete urological workup from those that do require a complete workup and thereby avoids unnecessary work-ups on patients of low probability of UC.

Phenotypic Assays

Phenotypic variables evaluated in the G+P INDEX include frequency of hematuria (HFREQ), age, gender, smoking history, and red blood count (RBC). These terms are defined herein below. Phenotypic variables are defined herein to include clinical findings and observations.

Genotypic Assays

In general, preferred genotypic assays developed by Pacific Edge Ltd. include quantification of expression of the genetic markers CDC2, HOXA13, MDK and IGFBP5 (a "4-marker assay"). In another preferred assay, the above 4 markers and a fifth marker, IL8R, is quantified (a "5-marker" or Cxbladder® assay; a trademark of Pacific Edge Ltd., Dunedin, New Zealand) (Holyoake A, O'Sullivan P, Pollock R et al: Development of a multiplex RNA urine test for the detection and stratification of transitional cell carcinoma of the bladder. Clin Cancer Res 2008;14: 742, and O'Sullivan P, Sharples K, Dalphin M et al: A Multigene Urine Test for the Detection and Stratification of Bladder Cancer in Patients Presenting with Hematuria. J Urol 2012, Vol. 188 No 3; 746), and International Patent Application No. PCT/NZ2011/000238, entitled "Novel Markers for Detection of Bladder Cancer." Each of these publications and patent application are herein incorporated fully by reference as if separately so incorporated.

In preferred embodiments, a 4-marker assay can be performed on unfractionated urine using PCR amplification to quantify four mRNA markers (for CDC2, HOXA13, MDK and IGFBP5), which are overexpressed in urothelial carcinoma. IL8R is highly overexpressed in neutrophils and is consequently elevated in non-malignant inflammatory conditions. Inclusion of this 5th mRNA marker significantly reduced the risk of false positive detection of

transitional cell carcinoma (TCC). From the patient's perspective, the test is non-invasive and very simple. A single sample of urine often mid-stream urine but not exclusively, is taken, and this can often be done at home without coming into the clinic.

The Cxbladder® assay has been shown to be considerably more sensitive than cytology in patients presenting with macroscopic hematuria. Most notably, the Cxbladder® assay achieved a sensitivity of 100% (at a pre-specified specificity of 85%) for all urothelial carcinomas with a stage greater than Ta, and 97% for all high-grade tumors. The Cxbladder® assay attributes a single value score that combines the quantitative gene expression of five genes represented in the patients urine. The score segregates patients into three classes based on the probability that the patient has a urothelial carcinoma.

For patients presenting with hematuria, (either macrohematuria or microhematuria), this invention has been shown to enhance these genotypic tools (either the 4-marker assay or Cxbladder® assay) with the addition of phenotypic variables collected from the patient over the same time period, and to combine these into a new tool, an index that can be used to segregate patients into three defined risk classes relative to the patient's probability of having urothelial carcinoma ("UC").

Aspects

Aspects of this invention are illustrated below. It can be understood that these are not the only aspects or embodiments of this invention. Persons of ordinary skill can combine one or more aspects together to produce additional aspects or embodiments.

One aspect includes a method for determining, in a patient presenting with hematuria,or the level of risk for having urothelial cancer, comprising:

providing a sample of urine from said patient;

quantifying a value, MI, comprising quantifying the levels of expression of human MDK, CDC2, HOXA13, and IGFBP5 in said sample;

assessing the phenotypic variables HFREQ, AgeGT, sex, SMK, and RBC of said patient;

calculating G+P INDEX according to either:

formula (i), G+P INDEX = (1*HFREQ+3*Gender+4*SMK) + (5*M1+2*IL-8),

or

formula (ii), G+P INDEX=

(w1*HFREQ+w2*AgeGT50+w3*Gender+w4*SMK+w5*RBC) + (w6*M1+w7*IL-8),

or

G+P INDEX = -8.46 + 0.79 IGF -1.60 HOXA + 2.10 MDK + 0.95 CDC – 0.38 IL8R + 0.98 SNS + 0.56 Hfreq +1.11 Gender + 0.64 Age; and

determining whether the G+P INDEX is greater than a threshold indicating the level of risk that the patient has urothelial cancer.

Additional aspects include the method of the other aspect, where said threshold is selected from the group of G+P INDEX values of from 0 to 5, from 6 to 10, or from 11-15, where said value of from 0 to 5 indicates Low Risk, 6 to 10 indicates Moderate Risk, and 11-15 indicates High Risk.

Further aspects include the method of any other aspect, where if said threshold is a G+P INDEX value of from 6-10, said patient undergoes additional clinical or laboratory tests.

Yet further aspects include the method of any prior aspect, where if said threshold is a G+P INDEX value of from 11-15, said patient undergoes additional clinical or laboratory tests.

Still further aspects include the method of any other aspect, where if said threshold is a G+P INDEX value of from 0-5, the patient is placed on a watch list for further clinical or laboratory tests.

Additional aspects include the method of any other aspect, where the threshold is established using a statistical method.

Still further aspects include the method of any other aspect, wherein the statistical method is any one of Linear Discriminant Analysis (LDA), Logistic Regression (LogReg), Support Vector Machine (SVM), K-nearest 5 neighbors (KN5N), and Partition Tree Classifier (TREE).

Additional aspects include the method of any other aspect, further comprising quantifying expression of one additional genotypic marker selected from Figure 6 or Figure 7.

Other aspects include the method of any previous aspect, where said step of quantifying genetic expression is carried out by detecting the levels of mRNA.

Further aspects include the method of any other aspect, wherein said step of quantifying genetic expression is carried out by detecting the levels of cDNA.

Yet further aspects include the method of any of any other aspect, where said step of quantifying genetic expression is carried out using an oligonucleotide complementary to at least a portion of said cDNA.

Additional aspects include the method of any other aspect, where said step of quantifying genetic expression is carried out using qRT-PCR method using a forward primer and a reverse primer.

Yet additional aspects include the method of any other aspect, where said step of quantifying genetic expression is carried out by detecting the levels of a protein.

Still other aspects include the method of any other aspect, where said step of quantifying genetic expression is carried out by detecting the levels of a peptide.

Additional aspects include the method of any of any other aspect, where said step of quantifying genetic expression is carried out using an antibody directed against said marker.

Yet further additional aspects include the method of any of claims 1 to 8 or 13-15, where said step of quantifying genetic expression is carried out using a sandwich-type immunoassay method, or using an antibody chip.

Still further aspects include the method of any other aspect, where said quantifying genetic expression is carried out using a monoclonal antibody.

Other aspects include the method any of any other aspect, where said quantifying genetic expression is carried out using a polyclonal antiserum.

G+P INDEX

Phenotypic and genotypic variables described above are combined into a G+P INDEX according to the following relationship:

G+P INDEX = (w1*HFREQ+w2*AgeGT50+w3*Gender+w4*SMK+w5*RBC) + (w6*M1+w7*IL-8),

where HFREQ means the frequency of finding 3 or more red blood cells per high power field in a 6-month period; if frequency is low then HFREQ = 0, and if higher than 3 red blood cells per high power field, then 1. AgeGT50 refers to subject's age, if greater than 50 years then AgeGT50 = 1, and if less than 50 years, then 0. Gender is assigned a value of 1 for male, and 0 for female. SMK means whether the subject is a current or ex-smoker; if non-smoker then

SMK = 0 and if a smoker, then 1. RBC means red blood cell count; if 25 or more then RBC is set to 1, and if less than 25, then 0. MI is a combination of expression of the genetic markers MDK, CDC, IGFBP5, and HOXA13; if M1 > 4.5 then set it to 1, if less than 4.5, 0. IL-8refers to expression level of RNA for IL-8; if IL-8 > 2.5 then IL-8 is set to 1, if ness than 2.5, 0. The symbols "*" means the multiplication operator, and weighting factors, w1 – w7 are respectively the weights assigned to each of the variables listed above in the G+P INDEX.

In other preferred embodiments, (AgeGT50 and RBC) may be dropped from the model as shown below:

G+P INDEX = (1*HFREQ+3*Gender+4*SMK)+(5*M1+2*IL-8)

The G+P INDEX produces a value between 0 and 15. A patient with G+P INDEX value of 11 to 15 is considered to be at "High Risk" for bladder cancer, and indicates the need for additional work up for bladder cancer. A patient with a G+P INDEX value of 6 to 10 is considered to be at "Moderate Risk" for developing bladder cancer, and additional work up is indicated. A patient with a G+P INDEX of 0 to 5 is considered to be at "Low Risk" for developing bladder cancer. Patients in the "Low Risk" group are placed on a watchful waiting list, and if additional symptoms appear, or if recurrent episodes of microhematuria occur, they are reevaluated for possible further work up.

As described more fully in Example 3 (**FIGs. 18** and **19**), we found that the ROC curve for the quantified phenotypic variables alone produced a modest level of diagnostic power. The ROC curve for the quantified genotypic markers alone produced a significant level of diagnostic power. We found an unexpectedly better diagnostic power when both genotypic and phenotypic variables were combined into a G+P INDEX.

Quantification of Genetic Expression

Proteins or nucleic acids that are secreted by or cleaved from the cell, or lost by apoptotic mechanisms, either alone or in combination with each other, have utility as serum or body fluid markers for the diagnosis of disease, including inflammatory disease in bladder and/or bladder cancer or as markers for monitoring the progression of established disease. Detection of protein and cell markers can be carried out using methods known in the art, and

include the use of RT-PCT, qRT-PCR, monoclonal antibodies, polyclonal antisera and the like.

Specifically the present invention provides methods for triaging patients presenting with hematuria, (either macrohematuria or microhematuria), comprising: (i) providing a biological sample; (ii) detecting one or more bladder tumor markers (BTMs) in said sample. Bladder tumor markers of particular interest include MDK, CDC2, HOXA13, and IGFBP5 (a "4-marker assay"). Optionally, one can also detect the levels of human neutrophil marker interleukin 8 receptor B (IL8Rb) in the sample (Cxbladder® assay). The presence of cancer can be established by comparing the levels of the one or more ~~BTMx~~ BTMs with the levels in normal patients, patients having early stage bladder cancer, and/or patients having an inflammatory disease. For example, the presence of cancer can be established by comparing the expression of BTMs against a threshold of expression. The threshold may be in the order of expression that is at least 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2, 3, 4, 5, 6, 7, 8, 9, or 10, 100, 1000, or up to 10,000 times the level of expression in a group of patients not having cancer. In other aspects, a high expression of IL8Rb without altered expression of a bladder tumor marker can be indicative of an inflammatory disease rather than cancer.

The methods of the present invention can be used in conjunction with any suitable marker for detecting bladder cancer. Examples of suitable markers for use in the invention are outlined in **FIGs. 6** or **7**. The present invention includes the use of any one or more of the markers outlined in **FIGs. 6** or **7**.

Optionally, in other preferred embodiments, the present invention can include any combination of IL8Rb with one or more of the markers MDK, CDC2, HOXA13, and IGFBP5, which can also be in combination with one or more other marker suitable for detecting bladder cancer, for example, any one of more of the markers outlined in **FIGs. 6** or **7**. More specifically, the present invention includes quantification of expression of any one or more combination of markers: IL8Rb / MDK, IL8Rb / CDC2, IL8Rb / HOXA13, IL8Rb / IGFBP5, IL8Rb / MDK / CDC2, IL8Rb / MDK / HOXA13, IL8Rb / MDK / IGFBP5, IL8Rb / CDC2 / HOXA13, IL8Rb / CDC2 / IGFBP5, IL8Rb / HOXA13 / IGFBP5, IL8Rb / MDK / CDC2 / HOXA13, IL8Rb / MDK / CDC2 / IGFBP5, IL8Rb / CDC2 / HOXA13 / IGFBP5, and IL8Rb / MDK / CDC2 / HOXA13 / IGFBP5. These combinations can optionally include one or more further markers suitable for detecting bladder cancer, for example any one of more of the markers outlined in **FIGs. 6** or **7**.

The present invention also provides for a method for detecting inflammatory conditions of the bladder, comprising: (i) providing a biological sample from a patient; and (ii) detecting the levels of human neutrophil marker interleukin 8 receptor B (IL8Rb) in said sample. The presence of inflammatory conditions of the bladder is established by comparing the levels of IL8Rb with the levels in normal patients, patients having hematuria, and patients having an inflammatory condition of the bladder. For example, the presence of an inflammatory condition of the bladder can be established by comparing the expression of the marker IL8Rb against a threshold, The threshold may be in the order of expression that is at least 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2, 3, 4, 5, 6, 7, 8, 9, 10 , 100, 1000, or up to 10,000 times the level of expression in another group of patients.

Preferred genotypic methods of the present invention can be carried out by detecting any suitable marker of gene expression, for example determining the levels of mRNA, cDNA, a protein or peptide utilizing any suitable method.

The establishment of a diagnosis can be established through the use classifier system, for example Linear Discriminant Analysis (LDA), Logistic Regression (LogReg), Support Vector Machine (SVM), K-nearest 5 neighbors (KN5N), and Partition Tree Classifier (TREE).

BRIEF DESCRIPTION OF THE FIGURES

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

FIG. 1 depicts protein and mRNA sequences of IL8Rb (also known as CXCR2).

FIG. 2 depicts a graph of scatter plots showing the effect of *IL8Rb* on the separation of TCC from non-malignant disease (prostate disease, cystitis, urinary tract infection and urolithiasis). IL8Rb has been substituted for different bladder cancer RNA markers in **FIGs. 2c** and **2f**. (a). MDK/IGFBP5; (b). MDK/HOXA13; (c) MDK/IL8Rb; (d) CDC2/IGFBP5; (e). CDC2/HOXA13; (f). CDC2/IL8Rb.

FIG. 3 depicts ROC curve analysis (sensitivity vs specificity) showing the effect of including IL8Rb in diagnostic algorithms derived using linear discriminate analysis (LD) and linear regression (LR). The ROC curves were derived from patients with TCC and upper urinary tract cancers (n=61), and the non-malignant diseases cystitis, urinary tract infection

and urolithiasis (n=61). (a). LD1 (solid) and LD2 (dashed). (b) LR1 (solid) and LR2 (dashed). IL8Rb is included in LD2 and LR2.

FIG. 4 depicts extended ROC curve analysis showing the effect of including IL8Rb in diagnostic algorithms derived using linear discriminate analysis (LD) and linear regression (LR). The ROC curves are derived from patients with TCC (n=56) and, unlike **FIG. 3**, any non-malignant disease in the cohort (n=386). (a). LD1 (solid) and LD2 (dashed). (b) LR1 (dashed) and LR2 (solid). IL8Rb is included in LD2 and LR2.

FIG. 5 depicts box plots showing the accumulation of IL8Rb mRNA in the urine of patients with non-malignant urological disease. The RNA has been quantified by qRT-PCR using the delta-Ct method (Holyoake et al, 2008). With this method a lower Ct reflects higher RNA levels. BPH: benign prostatic hyperplasia; UTI: urinary tract infection; NS prostate: non-specific prostate diseases; Vasc. Prostate: vascular prostate; warfarin: hematuria secondary to warfarin use. The observations in patients with cystitis/UTI are significantly different (p=0.001) to the other non-malignant presentations shown.

FIG. 6 depicts markers known to be over expressed in bladder cancer, and are suitable for use in the present invention.

FIG. 7 depicts markers known to be under expressed in bladder cancer, and are suitable for use in the present invention.

FIG. 8 depicts a flow chart for the patient recruitment procedures and numbers for Example 2.

FIG. 9 depicts baseline clinical and demographic characteristics of the patients by disease status at 3 months.

FIG. 10 depicts overall sensitivity and specificity of the urine tests.

FIG. 11 depicts various ROC curves; **FIG. 11a** depicts ROC curves for NMP22 ELISA and uRNA-D (test comprising the four markers MDK+CDC2+IGFBP5+HOXA13); and **FIG. 11b** depicts ROC curve for the five markers MDK, CDC2, HOXA13, IGFBP5 and IL8Rb.

FIG. 12 depicts the sensitivity of urine tests by stage, grade, location of tumour, multiplicity of tumor, hematuria status, creatinine of urine sample and sex. Tables show numbers and percent with a positive urine test among those with TCC.

FIG. 13 depicts specificity of urine tests by diagnosis, macrohematuria or, creatinine and sex. Tables show number and % with a negative urine test result among those without TCC.

FIG. 14 depicts ROC curves for the combinations of markers: (a) MDK, (b) CDC, (c) IGFBP5, (d) HOXA13, (e) MDK+CDC2, (f) MDK+IGFBP5, (g) MDK+HOXA13, (h) CDC2+IGFBP5, (i) CDC+HOXA13, (j) IGF+HOXA13, (k) MDK+CDC2+IGFBP5, (l) MDK+CDC2+HOXA13, (m) MDK+IGFBP5+HOXA13, (n) CDC2+IGFBP5+HOXA13, (o) MDK+CDC2+IGFBP5+HOXA13, plus or minus IL8Rb, using five different classifier models (i) Linear Discriminant Analysis (LDA), (ii) Logistic Regression (LogReg), (iii) Support Vector Machine (SVM), (iv) K-nearest 5 neighbors (KN5N), and (v) Partition Tree Classifier (TREE).

FIG. 15 depicts results of sensitivity selectivity studies. **FIG. 15a** depicts "Area Under the Curve" (AUC) for up to 20% false positive rate (at 80% specificity) of the ROC curves from **FIG. 14** and **FIG. 15b** showing the difference the AUC resulting from the inclusion of IL8Rb.

FIG. 16 depicts graphs of the sensitivity of the combinations of the four markers MDK, CDC2, IGFBP5, and HOXA13, plus or minus IL8Rb, using five different classifier models (i) Linear Discriminant Analysis (LDA), (ii) Logistic Regression (LogReg), (iii) Support Vector Machine (SVM), (iv) K-nearest 5 neighbors (KN5N), and (v) Partition Tree Classifier (TREE), at different set specificities; (a) 80%, (b) 85%, (c) 90%, (d) 95%, (e) 98%.

FIG. 17 depicts the gains in sensitivity from adding IL8Rb at different set specificities (a) 80%, (b) 85%, (c) 90%, (d) 95%, (e) 98%, and the resulting gains in specificity from adding IL8Rb at different set specificities (f) 80%, (g) 85%, (h) 90%, (i) 95%, (j) 98%.

FIG. 18 depicts ROC curves for testing of patients having hematuria studied using either genetic testing alone, phenotype evaluation alone, and/or both genetic testing and phenotypic evaluation.

FIG. 19 depicts a graph of odds ratios (horizontal axis) for variables gender, smoking history and HFREQNEW of this invention.

FIGs. 20 A and **20B** depict flow charts for standards of reporting diagnostic accuracy. FIG. 20A depicts a flow chart for patients with macrohaematuria across all three cohorts in this study.

FIG. 20B depicts flow chart for reporting diagnostic accuracy in patients with

microhematuria included in this study.

FIG. 21 depicts ROC curves representing the three classification models. P INDEX (dotted line), G INDEX (dashed line) and G+P INDEX (solid line).

FIG. 22 depicts NPV versus proportion of patients with haematuria testing negative according each model. P INDEX (dotted line), G INDEX (dashed line), and G+P INDEX (solid line).

FIG. 23 depicts a graph of the relationship between detect results (horizontal axis) versus Triage result (vertical axis).

FIG. 24 depicts a graph of G2 INDEX (horizontal axis) versus G1 +P INDEX (vertical axis).

DETAILED DESCRIPTION

Definitions

Before describing the embodiments of the invention in detail, it will be useful to provide some definitions of terms as used herein.

The term "marker" refers to 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 or gene fragment, whether coding or non-coding, DNA or DNA fragment RNA or RNA fragment, whether coding or non-coding; or a gene product, including a polypeptide such as a peptide, oligopeptide, protein, or protein fragment; or any related metabolites, by products, or any other identifying molecules, such as antibodies or antibody fragments, whether related directly or indirectly to a mechanism underlying the phenomenon. 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, any fragments, any possible probes (e.g., created across an exon-exon boundary), including those with capture motifs, hairpins or fluorophores, or any complements thereof, and any measurable marker thereof as defined above.

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 Fab_2 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 IgGl, 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.

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 neighboring cells, release of cytokines or other secretory products at abnormal levels, suppression or aggravation of inflammatory or immunological response, neoplasia, pre-malignancy, malignancy, invasion of surrounding or distant tissues or organs, such as lymph nodes, etc.

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

The term "bladder cancer" refers to a tumor originating in the bladder. These tumors are able to metastasize to any organ.

The term “BTM” or “bladder tumor marker” or “BTM family member” means a tumor marker (TM) that is associated with urothelial cancers, bladder cancer, transitional cell carcinoma of the bladder (TCC), squamous cell carcinomas, and adenocarcinomas of the bladder. The term BTM also includes combinations of individual 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 tumors. Rather, expression of BTM can be altered in other types of cells, diseased cells, tumors, including malignant tumors.

The term “under expressing BTM” means a marker that shows lower expression in bladder tumors than in non-malignant bladder tissue.

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

The terms "differentially expressed," "differential expression," and like phrases, refer to a gene marker whose expression is activated to a higher or lower level in a subject (e.g., test

sample) having a condition, specifically cancer, such as melanoma, relative to its expression in a control subject (e.g., reference sample). The terms also include markers whose expression is activated to a higher or lower level at different stages of the same condition; in diseases with a good or poor prognosis; or in cells with higher or lower levels of proliferation. A differentially expressed marker may be either activated or inhibited at the polynucleotide level or polypeptide level, or may be subject to alternative splicing to result in a different polypeptide product. Such differences may be evidenced by a change in mRNA levels, surface expression, secretion or other partitioning of a polypeptide, for example.

Differential expression may include a comparison of expression between two or more markers (e.g., genes or their gene products); or a comparison of the ratios of the expression between two or more markers (e.g., genes or their gene products); or a comparison of two differently processed products (e.g., transcripts or polypeptides) of the same marker, which differ between normal subjects and diseased subjects; or between various stages of the same disease; or between diseases having a good or poor prognosis; or between cells with higher and lower levels of proliferation; or between normal tissue and diseased tissue, specifically cancer, or melanoma. Differential expression includes both quantitative, as well as qualitative, differences in the temporal or cellular expression pattern in a gene or its expression products among, for example, normal and diseased cells, or among cells which have undergone different disease events or disease stages, or cells with different levels of proliferation.

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 the 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 is the binding of a binding ligand, such as a hybridization probe or antibody, to a gene or other polynucleotide or oligonucleotide, a polypeptide or a protein fragment, and the visualization of the binding ligand. Thus, the intensity 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.

The terms "gene expression threshold," and "defined expression threshold" are used interchangeably and refer to the level of a marker in question, outside which the expression level of the polynucleotide or polypeptide serves as a predictive marker for a condition in the patient. For example, the expression of IL8Rb above a certain threshold is diagnostic that the patient has an inflammatory condition. A threshold can also be used when testing a patient for suspected bladder cancer, using bladder cancer makers. Expression levels above a threshold indicates that the patient has an inflammatory bladder condition, likely to cause a false positive test for cancer, whereas an expression level of IL8Rb below a threshold is predictive that the patient does not have an inflammatory bladder condition. By including the measurement of IL8Rb any result from the expression of the bladder tumor markers can be relied upon if the levels of IL8Rb is below the threshold (i.e. a positive result is likely to be positive for the patient having cancer rather than increased levels of the bladder tumor markers actually resulting from exfoliation of non-malignant cells from the mucosa from inflammation).

The term "diagnostic threshold" refers to a threshold in which a patient can be said to have been diagnosed either with or without a given condition, for example bladder cancer. A diagnostic threshold is generally set to achieve a desired sensitivity and specificity, depending on factors such as population, prevalence, and likely clinical outcome. In general the diagnostic threshold can be calculated and/or established using algorithms, and/or computerized data analysis.

The exact threshold will be dependent on the population and also any model being used to predict disease (predictive model). A threshold is established experimentally from clinical studies such as those described in the Examples below. Depending on the prediction model used, the expression threshold may be set to achieve maximum sensitivity, or for maximum specificity, or for minimum error (maximum classification rate). For example a higher threshold may be set to achieve minimum errors, but this may result in a lower sensitivity. Therefore, for any given predictive model, clinical studies will be used to set an expression threshold that generally achieves the highest sensitivity while having a minimal error rate. In general the threshold is likely to be in the order of expression that is at

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.

The term "Receiver Operating Characteristic" ("ROC curve") means a plot of the true positive rate (sensitivity) against the false positive rate (specificity) for different cut off points for a particular marker or test. Each point on the ROC curve represents a specific sensitivity/specificity point that will correspond to a given threshold. ROC curves can be important to establish a threshold to give a desired outcome. The area under a ROC curve represents (expressed as an Area Under the Curve (AUC) analysis, can be a measure of how well a given marker or test consisting of a number of markers, can distinguish between two or more diagnostic outcomes. ROC curves can also be used to compare the accuracy of two different tests.

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, and expression in cells and organisms.

The term "overexpression" or "overexpressed" refers to an expression level of a gene or marker in a patient that is above that seen in normal tissue. Expression may be considered to be overexpressed if it is 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2, 10, 100, 1000, or up to 10,000 times the expression in normal tissue or in tissues from another group of patients.

The term "polynucleotide," when used in the singular or plural, generally refers to any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. This includes, without limitation, single- and double-stranded DNA, 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 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 thereof.

The term "phenotypic," means a trait that is observable in a clinical setting, or in a clinical interview, or in a patient's history. When used in a formula for calculating G+P index, "phenotypic" or "P" means the patient's age, sex, incidence of hematuria, and smoking history.

"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 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.

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.

The term "Reverse Transcription" means a process in which an oligoribonucleotide, including a messenger RNA ("mRNA") is used as a template for biochemical synthesis of a complementary oligodeoxyribonucleotide ("cDNA") using an enzyme ("Reverse Transcriptase"), which binds to the template RNA, and catalyzes a series of addition reactions that sequentially attaches deoxyribonucleotide bases to form an oligodeoxyribonucleotide strand that is complementary to the RNA template.

The term "Hematuria" is defined as the presence of blood in the urine. It may present as macroscopic hematura (visible traces of blood cells) or microscopic hematuria (microscopic traces of blood) within the urine. A confirmed indication of microhematuria is defined as 3 or more red blood cells present per microscopic high-powered field (HPF) on a minimum of 3 properly collected urine samples. Microhematuria may also be detected by urine dipstick (colorimetric comparison estimate) at clinic. Hematuria (either microscopic or macroscopic) may be asymptomatic (no additional symptoms associated with hematuria) or symptomatic. Additional symptoms include dysuria (painful urination), a feeling of incomplete emptying of the bladder or increased frequency or urination.

"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 re-anneal when complementary strands are present in an environment below their melting temperature. The higher the degree of desired homology between the probe and hybridizable 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).

"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 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 ug/ml), 0.1% SDS, and 10% dextran sulfate at 42°C, with washes at 42°C in 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 include the use of washing solution and hybridization conditions (e. g., temperature, ionic strength, and % SDS) less stringent that 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 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 term "IL8Rb" means neutrophil marker interleukin 8 receptor B (also known as chemokine (C-X-C motif) receptor 2 [CXCR2]) (**FIG. 1**; SEQ ID NOs. 1 and 2), and includes the marker IL8Rb. The term includes a polynucleotide, such as a gene or gene fragment, RNA or RNA fragment; or a gene product, including a polypeptide such as a peptide, oligopeptide, protein, or protein fragment; or any related metabolites, by products, or any other identifying molecules, such as antibodies or antibody fragments.

The term "reliability" includes the low incidence of false positives and/or false negatives. Thus, with higher reliability of a marker, fewer false positives and/or false negatives are associated with diagnoses made using that marker.

"Accuracy" is the proportion of true results (true positives plus true negatives) divided by the number of total cases in the population, according to the formula:

$$\frac{\text{True Positives + True Negatives}}{\text{Total Number of Measurements}}$$

The term "triage" means to differentiate patients with hematuria that have a low probability of having bladder cancer from those patients with hematuria that have a reasonable probability of having bladder cancer and requiring further clinical work up, including cystoscopy or other clinical procedure.

Embodiments

Therefore, in certain preferred embodiments, a combination of genetic markers and phenotypic markers are provided that permit differentiating patients having a low probability of having bladder cancer from those patients with a sufficient risk of having bladder to warrant further clinical work up, possibly including cystoscopy or other procedures. In other embodiments, markers are provided that have reliability greater than about 70%; in other embodiments, greater than about 73%, in still other embodiments, greater than about 80%, in yet further embodiments, greater than about 90%, in still others, greater than about 95%, in yet further embodiments greater than about 98%, and in certain embodiments, about 100% reliability.

For genetic analysis, the practice of the present invention will employ, unless otherwise indicated, conventional 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.

It is to be understood that the above terms may refer to protein, DNA sequence and/or RNA sequence. It is also to be understood that the above terms also refer to non-human proteins, DNA and/or RNA having homologous sequences as depicted herein.

Embodiments of this Invention

Often patients referred to a urologist with hematuria are booked to be seen in a dedicated hematuria clinic. Patients with macroscopic hematuria are often prioritized over those with micro hematuria. The information provided at the time of referral from the primary care provider can be highly variable, making accurate stratification of patients by probability of having a urothelial cancer difficult. Often urine cytology is routinely requested before the patient is seen and, if positive, is used to increase the patients priority of receiving a full clinical work-up, however urine cytology, whilst highly specific, has a very low sensitivity and hence is of little practical value with its high rate of false-negatives[6].

Phenotypic and Genotypic Analysis of Patients With Hematuria Not Having Bladder Cancer

G+P INDEX

Phenotypic and genotypic variables described above are combined into a G+P INDEX according to the following relationship:

G+P INDEX = (w1*HFREQ+w2*AgeGT50+w3*Gender+w4*SMK+w5*RBC) + (w6*M1+w7*IL-8),

where HFREQ means the frequency of finding 3 or more red blood cells per high power field

in a 6-month period; if frequency is low then HFREQ = 0, and if higher than 3 red blood cells per high power field, then 1. AgeGT50 refers to subject's age, if greater than 50 years then AgeGT50 = 1, and if less than 50 years, then 0. Gender is assigned a value of 1 for male, and 0 for female. SMK means whether the subject is a current or ex-smoker; if non-smoker then SMK = 0 and if a smoker, then 1. RBC means red blood cell count; if 25 or more then RBC is set to 1, and if less than 25, then 0. MI is a combination of expression of the genetic markers MDK, CDC, IGFBP5, and HOXA13; if M1 > 4.5 then set it to 1, if less than 4.5, 0. IL-8refers to expression level of RNA for IL-8; if IL-8 > 2.5 then IL-8 is set to 1, if ness than 2.5, 0. The symbols "*" means the multiplication operator, and weighting factors, w1 – w7 are respectively the weights assigned to each of the variables listed above in the G+P INDEX.

In other preferred embodiments, (AgeGT50 and RBC) may be dropped from the model as shown below:

G+P INDEX = (1*HFREQ+3*Gender+4*SMK)+(5*M1+2*IL-8)

The G+P INDEX produces a value between 0 and 15. A patient with G+P INDEX value of 11 to 15 is considered to be at "High Risk" for bladder cancer, and indicates the need for additional work up for bladder cancer. A patient with a G+P INDEX value of 6 to 10 is considered to be at "Moderate Risk" for developing bladder cancer, and additional work up is indicated. A patient with a G+P INDEX of 0 to 5 is considered to be at "Low Risk" for developing bladder cancer. Patients in the "Low Risk" group are placed on a watchful waiting list, and if additional symptoms appear, or if recurrent episodes of microhematuria occur, they are reevaluated for possible further work up.

Genotypic variables useful for differentiating patients without and with bladder cancer include expression of RNA markers "M1" being a combination of MDK, CDC, IGFBP5 (IGBP5), and HOXA13. Another genotypic variable is expression of RNA for IL8R. Coefficients for these genotypic variables are shown in **Table** 7 below. A threshold of 4.5 and 2.5 was used for M1 and IL8R, respectively, and a coefficient of 5 and 2, respectively, were assigned.

The G+P INDEX produces a value between 0 and 15. A G+P INDEX value of 11 to 15 is considered "High Risk" for bladder cancer, and indicates the need for additional work up for bladder cancer. A G+P INDEX value of 6 to 10 is considered "Moderate Risk" for

developing bladder cancer, and additional work up is indicated. A G+P INDEX of 0 to 5 is considered "Low Risk" for developing bladder cancer, and these patients are placed on a waiting list, and if additional symptoms appear, or if recurrent episodes of microhematuria occur, are reevaluated for possible fuller work up.

As described more fully in Example 3 (**FIGs. 18** and **19**), we found that the ROC curve for phenotypic data alone produced a modest level of diagnostic power. The ROC curve for genotypic data alone produced a significant level of diagnostic power. We found an unexpectedly better diagnostic power when both genotypic and phenotypic data was combined into a G+P INDEX.

Genetic Analysis of Patients Not Having Bladder Cancer

In some preferred embodiments, this invention combines use of a 4-marker assay or a Cxbladder® assay (genotypic variables) and one or more of five key risk factors (phenotypic variables) to produce a selection index that can be used to triage patients with microscopic or macroscopic hematuria in terms of their potential risk of having urothelial cancer. While not precluding the need for a flexible cystoscopy, patients deemed at high risk of urothelial cancer based on phenotypic variables and genotypic variables may be seen earlier, potentially improving overall patient outcome.

Genotypic markers can be used as tools to detect cancer-free patients or to select patient groups that are at low, medium or high risk of having a disease. The markers can, for example, be differentially expressed between disease tissue and corresponding non-disease tissue. In this situation, the detection of differential expression is associated with the presence of the disease. Alternatively, the marker can be associated directly with changes occurring in the disease tissues, or changes resulting from the disease. Inflammatory diseases are associated with an increase in neutrophils. It has been found that the neutrophil marker interleukin 8 receptor B (IL8Rb; **Fig. 1**; SEQ ID NOs 1 and 2), can provide a good marker for the presence of neutrophils in a sample, and therefore can be used as a diagnostic marker for the detection of inflammatory disease in a sample, and in particular, in the detection of inflammatory disease of the bladder.

As shown in **FIG. 5**, accumulation of IL8Rb in urine is indicative of the presence of inflammatory disease of the bladder. Specifically, **FIG. 5** shows the accumulation of IL8Rb in the urine of patients having the conditions; benign prostatic hyperplasia, urinary tract

infection, non-specific prostate diseases, vascular prostate and secondary warfarin use. It will be appreciated however, that the use of IL8Rb is not be limited to the detection of these diseases only, but that these examples show that IL8Rb does increase in samples from patients having an inflammatory disease of the bladder. That is, IL8Rb can be used as a marker of inflammation associated with bladder disease and therefore is suitable for use in detecting any condition associated with inflammation. Therefore, the detection of the amount of IL8Rb can be used as a marker for inflammatory disease of the bladder. More particularly, IL8Rb can be used to detect inflammatory disease of the bladder associated with the accumulation of neutrophils.

Urine tests for TCC rely largely on the presence of markers in the urine derived from exfoliated tumor cells. The ability to detect these cells can be masked by the presence of large numbers of contaminating cells, such as, blood and inflammatory cells. Moreover, inflammation of the bladder lining can result in the increased exfoliation of non-malignant cells from the mucosa. As a result, urine tests that use markers derived from bladder transitional cells have a higher likelihood of giving a false positive result from urine samples taken from patients with cystitis, urinary tract infection or other conditions resulting in urinary tract inflammation or transitional cell exfoliation, such as, urolithiasis (Sanchez-Carbayo et al).

One way to try and avoid such false positive results has been to select markers with low relative expression in blood or inflammatory cells. The use of such markers results in fewer false positives in TCC patients presenting with non-malignant, inflammatory conditions. However, low expression of the markers in hematologically-derived cells fails to compensate for the enhanced rate of exfoliation of non-malignant transitional cells.

It has been discovered that the negative impact of exfoliated transitional cells from inflamed tissue has on the accuracy of bladder cancer urine tests can be minimized by improving the identification of patients with inflammatory conditions of the urinary tract. Here it has been surprisingly found that using the marker IL8Rb in combination with one or more bladder tumor markers (BTM's) provides for a more accurate detection of bladder cancer. In particular, a marker based test for bladder cancer that includes the marker IL8Rb is less susceptible to false positive results, which can result in patients suffering from an inflammatory non-cancer condition.

In general, the presence or absence of an inflammatory condition is established by having a threshold of gene expression, above which expression of IL8Rb is indicative of an inflammatory condition. For example, the expression of IL8Rb above a certain threshold is diagnostic that the patient has an inflammatory condition (see thresholds described above)

When IL8Rb is used in conjunction with one or more markers predictive for the presence of bladder cancer, the presence of elevated expression of the bladder tumor marker(s), and expression of IL8Rb, above a certain threshold, is predictive of the patient having an inflammatory condition and not cancer. Furthermore, if the test is preformed on urine from the patient, then this result is predictive of the patient having an inflammatory bladder condition. The high levels of the bladder tumor markers are most likely the result of non-malignant cells coming from the mucosa as a result of the inflammation. That is, the patient, although having high levels of the bladder tumor marker(s) does not actually have bladder cancer – a false positive.

Alternatively, if the patient has abnormally high levels or diagnostic levels of one or more bladder tumor markers, but the level of IL8Rb is below a threshold, then the patient is likely to have cancer, and in particular bladder or urothelial cancer. This is especially so, if the test is preformed on urine from the patient. This result is of significant benefit to the health provider because they can be sure that the patient does have cancer, and can start treatment immediately, and not be concerned that the result is actually caused an inflammatory condition giving a false positive result.

It has been surprisingly shown that the quantification of RNA from the gene encoding the neutrophil marker interleukin 8 receptor B (IL8Rb) improves the overall performance of detecting patients with TCC, using known TCC or BTM markers. The reference sequences for IL8Rb are shown in **FIG. 1** and SEQ ID NOs 1 and 2). In addition to its role in TCC detection, it has been explored whether IL8Rb could be used as a urine marker to aid in the diagnosis of inflammatory disease (**FIG. 5**).

The use of IL8Rb marker can be used in isolation for the detection of inflammatory conditions of the bladder utilizing known methods for detecting gene expression levels. Examples of methods for detecting gene expression are outlined below.

Alternatively, IL8Rb can be combined with one of more BTMs to detect bladder cancer. It has been shown that by utilizing the inflammatory disease marker IL8Rb as part of the test for bladder cancer, the influence of inflamed tissue on creating a false positive result is

minimized. The marker IL8Rb can be used in association with any bladder cancer markers, or alternatively can be used with two or more markers, as part of a signature, for detecting bladder cancer.

Reducing the number of false positive results means that fewer patients not having bladder cancer are subjected to potentially unnecessary procedures, including cystoscopy, which carries its own risks. Reducing the number of false negative results means that it is more likely that a patient with bladder cancer is detected, and can therefore be further evaluated for cancer.

The action of IL8Rb to improve the detection of bladder cancer results from the ability to separate non-malignant conditions from patients having bladder cancer. This is achieved because an increase of IL8Rb is indicative of an increase in the presence of neutrophils in a sample. Therefore, the ability of IL8Rb is not dependent on the bladder tumor marker used. As shown in **FIGs. 2**, and **12** to **15**, when combined with a variety of bladder tumor markers and combinations of bladder tumor markers, IL8Rb had the general effect of increasing the specificity of the ability of the marker(s) to detect cancer in the subjects.

One example of a signature according to the present invention is the use of IL8Rb in combination with MDK, CDC2, IGFBP5 and HOXA13, which may also be in combination with one or more other marker suitable for detecting bladder cancer, for example any one of more of the markers outlined in **FIGs. 6** or **7**. As shown in **FIGs 14** and **15**, IL8Rb can be used in any combination of the markers, specifically the combinations IL8Rb / MDK, IL8Rb / CDC2, IL8Rb / HOXA13, IL8Rb / IGFBP5, IL8Rb / MDK / CDC2, IL8Rb / MDK / HOXA13, IL8Rb / MDK / IGFBP5, IL8Rb / CDC2 / HOXA13, IL8Rb / CDC2 / IGFBP5, IL8Rb / HOXA13 / IGFBP5, IL8Rb / MDK / CDC2 / HOXA13, IL8Rb / MDK / CDC2 / IGFBP5, IL8Rb / CDC2 / HOXA13 / IGFBP5, and IL8Rb / MDK / CDC2 / HOXA13 / IGFBP5. As shown in **FIGs. 14** and **15**, the inclusion of IL8Rb increased the ability of the marker, or the combination of markers to accurately diagnose bladder cancer in a subject. The present invention is not to be limited to these specific combinations but can optionally include one or more further markers suitable for detecting bladder cancer, for example any one of more of the markers outlined in **FIGs. 6** or **7**. **Table 1** below shows the identifiers for the specific markers MDK, CDC2, IGFBP5 and HOXA13 and IL8Rb.

Table 1: Identifiers for Bladder Tumor Markers

PE Gene Name	HGNC Gene Name (Official)	NCBI RefSeq	NCBI Entrez Gene ID	HGNC URL
MDK	MDK	NM_002391	4192	http://www.genenames.org/data/hgnc_data.php?hgnc_id=6972
CDC	CDK1	NM_001170406	983	http://www.genenames.org/data/hgnc_data.php?hgnc_id=1722
IGF	IGFBP5	NM_000599	3488	http://www.genenames.org/data/hgnc_data.php?hgnc_id=5474
HOXA	HOXA13	NM_000522	3209	http://www.genenames.org/data/hgnc_data.php?hgnc_id=5102
IL8Rb	CXCR2	NM_001168298	3579	http://www.genenames.org/data/hgnc_data.php?hgnc_id=6027

FIGs. 2 to **4** and **12** to **17** show the effect of using IL8Rb in combination with four known, representative, markers of bladder cancer; MDK, CDC2, IGFBP5 and HOXA13. The results show that by incorporating the use if IL8Rb individually with each marker (**FIGs. 2**, **14** and **15** to **17**), but also when used with all possible combinations of the four BTM markers as a signature, there is an improvement in the ability to separate the samples of patients with TCC and those with non-malignant conditions.

As shown in **FIGs. 10** to **13**, the inclusion of IL8Rb with the four markers MDK, CDC2, IGFBP5 and HOXA13 (uRNA-D) not only increased the overall performance of the test compared to the four markers alone, the test also compared extremely favorably with other known tests, NMP22® "a registered trademark of Matritech, Inc., of Massachusetts, United States" Elisa, NMP22 BladderChek® (a registered trademark of Matritech, Inc., of Massachusetts, United States), and cytology. **FIGs. 14** through **17** also show the effect of IL8Rb in the various combinations of the four markers MDK, CDC2, IGFBP5 and HOXA13.

FIG. 14 shows the ROC curves for all the combinations of the four markers MDK, CDC2, IGFBP5 and HOXA13, with and without IL8Rb, calculated using five different classifier models (i) Linear Discriminant Analysis (LDA), (ii) Logistic Regression (LogReg), (iii) Support Vector Machines (SVM), (iv) K-nearest 5 neighbors (KN5N), and (v) Partition Tree Classifier (TREE). **FIG. 15** tabulates the Area Under the Curve (AUC) for all 5

classifiers and all 15 combinations of the 4 biomarkers, with and without IL8Rb. This AUC calculation is restricted to the area from a false positive rate of 0 to a false positive rate of 20%, covering the useful ranges of specificity (80 – 100%). The AUC quantifies the visible differences on the ROC curves of **FIG. 14**. **FIG. 16** shows the sensitivity of all combinations of the four markers measured with and without IL8Rb at specificities of (a) 80%, (b) 85%, (c) 90%, (d) 95%, and (e) 98%. **FIG. 17** tabulates the changes in either sensitivity (vertical direction on the ROC curves; better is "up") or specificity (horizontal direction on the ROC curve; better is to the left) at the fixed specificities of (a, f) 80%, (b, g) 85%, (c, h) 90%, (d, I) 95%, and (e, j) 98%, respectively.

These results show that IL8Rb, in general, improves the ability of the biomarkers (MDK, CDC, IGFBP5, and HOXA13), singly or in combination, to classify tumor from normal samples.

These results generally show that the IL8Rb was able to increase the accuracy at which the test could detect bladder or urothelial cancer. The biggest gains were seen with either markers that did not perform as well without the inclusion of IL8Rb or with classifiers that did not perform as well. Smaller gains were seen for markers and/or classifiers that performed well prior to adding IL8Rb and therefore there was less room for improvement. It is important to note that the results show a population based analysis and the benefit of incorporating IL8Rb could be greater when diagnosis individual patients, especially those whose diagnosis on the expression of the BTM markers maybe unclear.

These results show that not only can IL8Rb be used to detect inflammatory disease of the bladder, but also when used in combination with markers for bladder cancer, results in an improved detection of bladder cancer, arising from a reduction in "false positive" results.

These results also show the utility of IL8Rb in that it affects the overall performance of the various markers combinations, and confirms the ability of IL8Rb to improve the performance of one or more bladder cancer markers to accurately detect cancer in a patient. Further, **FIGs. 14** and **15** show that the same results can be achieved using a range of classifier models, and shows that the result is not dependent on a classifier model or algorithm, but rather the combination of markers used. These results confirm that any suitable classifier model or algorithm could be used in the present invention. In particular, **FIGs. 14** and **15** show that IL:8Rb has a greater effect at the higher specificities, and in particular in the most clinically applicable ranges.

Therefore, using the G+P Index of this invention, we are able to accurately triage patients with hematuria based on phenotypic and genetic variables into groups being at "High Risk" for having urothelial cancer and warrant immediate further work-up, "At Risk" and warrant immediate work-up, and those with "Low Risk" who may be placed on a watch list for later evaluation.

Detection of Genetic Markers in Body Samples

In several preferred embodiments, assays for cancer can be desirably carried out on samples obtained from blood, plasma, serum, peritoneal fluid obtained for example using peritoneal washes, or other body fluids, such as urine, lymph, cerebrospinal fluid, gastric fluid or stool samples. For the detection of inflammatory conditions of the bladder or bladder cancer the test is ideally preformed on a urine sample.

Specifically, present methods for detecting inflammatory bladder disease or bladder cancer can be conducted on any suitable sample from the body that would be indicative of the urine, but ideally the level of IL8Rb, and any further cancer marker is established directly from a urine sample.

A test can either be performed directly on a urine sample, or the sample may be stabilized by the addition of any suitable compounds or buffers known in the art to stabilize and prevent the breakdown of RNA and/or protein in the sample so that it can be analyzed at a later date, or even to ensure that the RNA and/or protein is stabilized during the analysis.

The determination of either the protein and/or RNA level in the subject's urine can be performed directly on the urine, or the urine can be treated to further purify and/or concentrate the RNA and/or protein. Many methods for extracting and/or concentrating proteins and/or RNA are well known in the art and could be used in the present invention.

It can be appreciated that many methods are well known in the art for establishing the expression level of a particular gene, either at the RNA and/or protein level, and any suitable method can be used in the present invention. Some common methods are outlined below, however, the invention is not restricted to these methods and any method for quantifying protein and/or RNA levels is suitable for use in the present invention.

General Approaches to Disease and Cancer Detection Using Genotypic Markers

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)

Quantitative PCR (qPCR) can be carried out on tumor samples, on serum and plasma using 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 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™ (a trademark of Applera Corporation, Connecticut, United States) or the Roche LightCycler™ (a trademark of Roche Molecular Systems, Inc., California, United States), 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 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 tumor tissues, with or without drug treatment, to characterize patterns of expression, to discriminate between closely related RNAs, and to analyze RNA structure.

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 tumors or tumor cell lines, and

corresponding normal tissues or cell lines, respectively. RNA can be isolated from a variety of samples, such as tumor samples from breast, lung, colon (e.g., large bowel or small bowel), colorectal, gastric, esophageal, anal, rectal, prostate, brain, liver, kidney, pancreas, spleen, thymus, testis, ovary, uterus, bladder etc., tissues, from primary tumors, or tumor cell lines, and from pooled samples from healthy donors. If the source of RNA is a tumor, RNA can be extracted, for example, from frozen or archived paraffin-embedded and fixed (e.g., formalin-fixed) tissue samples.

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 leukemia 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 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.

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® qPCR (a registered trademark of Roche Molecular Systems, Inc., California, United States) 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.

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 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 provides the basis for quantitative interpretation of the data.

TaqMan® RT-PCR (a registered trademark of Roche Molecular Systems, Inc., California, United States) can be performed using commercially available equipment, such as, for example, ABI PRISM 7700™ Sequence Detection System (a trademark of Applera Corporation, Connecticut, United States) (Perkin-Elmer-Applied Biosystems, Foster City, CA, USA), or Lightcycler™ (a registered trademark of Roche Molecular Systems, Inc., California, United States((Roche Molecular Biochemicals, Mannheim, Germany). In a preferred embodiment, the 5' nuclease procedure is run on a real-time quantitative PCR device such as the ABI PRISM 7700™ 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 fiberoptic cables for all 96 wells, and detected at the CCD. The system includes software for running the instrument and for analyzing the data.

5' nuclease assay data are initially expressed as Cp, 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 signal is first recorded as statistically significant is the threshold cycle, Cp.

Real-time Quantitative PCR (qRT-PCR)

A more recent variation of the RT-PCR technique is the real time quantitative PCR, which measures PCR product accumulation through a dual-labeled fluorigenic probe (i.e., TaqMan® 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).

Expression levels can be determined using fixed, paraffin-embedded tissues as the RNA source. According to one aspect of the present invention, PCR primers are designed to flank 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 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.

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 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 VIMNV 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).

The most important factors considered in PCR primer design include primer length, melting temperature (Tm), and G/C content, specificity, complementary primer sequences, and 3' end sequence. In general, optimal PCR primers are generally 1730 bases in length, and contain about 20-80%, such as, for example, about 50-60% G+C bases. Melting 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 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

Differential expression can also be identified, or confirmed using the microarray technique. Thus, the expression profile of disease specific markers can be measured in either fresh or paraffin-embedded tumor 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 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 tumors or tumor cell lines, and corresponding normal tissues or cell lines. Thus RNA can be isolated from a variety of primary tumors or tumor cell lines. If the source of RNA is a primary tumor, RNA can be extracted, for example, from frozen or archived formalin fixed paraffin-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, 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.

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 (Schena 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 tumor 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® "a registered trademark of Qiagen GmbH, Hilden, Germany" 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 micron thick sections of paraffin-embedded tumor 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 tumor 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 horseradish 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 polyacrylamide 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 markers 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-1 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 (a registered trademark of Panomics, of California, United States) (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 colorometric 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 marker 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)). Nonspecific binding sites are blocked with non-target protein preparations and detergents. The capture antibody is then incubated with a preparation of sample or tissue from the patient containing the antigen. The mixture is washed before the antibody/antigen complex is incubated with a second antibody that detects the target marker. 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 marker 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.

Disease markers can also be detected in tissues or samples 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 marker are electrophoresed through polyacrylamide gels under denaturing or non-denaturing conditions. The proteins are then transferred to a membrane support such as nylon. The marker 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.

In immunoprecipitation, a soluble preparation containing the marker is incubated with a monoclonal or polyclonal antibody against the marker. 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-marker-antigen bound to the bead. Following washing the bound marker can be detected and quantified by immunoblotting or ELISA.

Establishing a Diagnosis Based on Genotypic Analysis

Once the level of expression of IL8Rb, and optionally one or more further cancer markers, has been obtained then a diagnosis for that subject can be established. If the expression of IL8Rb is above the expression seen in subjects that do not have an inflammatory bladder disease, and/or is consistent with the level of expression in subjects known to have an inflammatory bladder disease, then the subject will be diagnosed as having an inflammatory bladder disease. Alternatively, if the expression is not above the expression seen in subjects that do not have an inflammatory bladder disease, and/or is below the levels of expression in subjects known to have an inflammatory bladder disease, then the subject will be diagnosed as not an inflammatory bladder disease.

In the situation where IL8Rb is used in conjunction with one or more markers for Bladder cancer, then the expression level of IL8Rb will be compared with the level of expression of subjects without an inflammatory bladder disease, and/or subjects known to have an inflammatory bladder disease. The one or more cancer markers are compared to the expression level in subjects without bladder cancer and/or subjects known to have bladder cancer. If the expression level of the IL8Rb is consistent with a subject that does not have an inflammatory bladder disease (less than a subject having an inflammatory bladder disease) and the expression level of the one or more bladder cancer markers are consistent with a subject having bladder cancer (differential to a subject that does not have bladder cancer), then the subject is diagnosed as having bladder cancer. If the expression level of the IL8Rb is greater than a subject that does not have an inflammatory bladder disease (consistent with a subject having an inflammatory bladder disease) and the expression level of the one or more bladder cancer markers are consistent with a subject having bladder cancer (differential to a subject that does not have bladder cancer), then the subject is diagnosed as having an inflammatory bladder disease. If the expression level of the IL8Rb is consistent with a subject that does not have an inflammatory bladder disease (less than a subject having an inflammatory bladder disease) and the expression level of the one or more bladder cancer markers are consistent with a subject that does not have bladder cancer (differential to a subject that does have bladder cancer), then the subject is diagnosed as having neither bladder cancer or an inflammatory bladder disease.

Because there is often an overlap in expression levels between the normal and disease expression of a diagnostic marker, in order to establish a diagnosis for a subject it is typical to

establish a classifying threshold. A classifying threshold is a value or threshold which distinguishes subjects into disease or non disease categories. A threshold is commonly evaluated with the use of a Receiver Operating Characteristic (ROC) curve, which plots the sensitivity against specificity for all possible thresholds.

Determination of Diagnostic Thresholds

For tests using disease markers, diagnostic thresholds can be derived that enable a sample to be called either positive or negative for the disease, e.g., bladder cancer. These diagnostic thresholds are determined by the analysis of cohorts of patients that are investigated for the presence of bladder cancer or inflammatory bladder disease. Diagnostic thresholds may vary for different test applications; for example, diagnostic thresholds for use of the test in population screening are determined using cohorts of patients who are largely free of urological symptoms, and these diagnostic thresholds may be different to those used in tests for patients who are under surveillance for bladder cancer recurrence. A diagnostic threshold can 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 80-100%.

A diagnostic threshold is determined by applying an algorithm that combines the genotypic expression levels of each marker to each sample from a prospective clinical trial.

Samples used are from patients with bladder cancer and a range of non-malignant urological disorders. A diagnostic threshold is selected by determining the score of the algorithm that resulted in the desired specificity. For example, in some applications a specificity of 85% is desired. ~~The~~ A diagnostic threshold is then set by selecting an algorithm score that results in 85% of patients without bladder cancer being correctly classed as negative for cancer. In other applications (such as population screening), higher specificity, such as 90%, is favoured. To set a threshold for this application, an algorithm score that results in 90% of patients without bladder cancer being correctly classed as negative for cancer is selected. Examples of the use of an algorithm is outlined in the Examples.

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 associated with them. For example, a test may have three intervals; one associated with a high (e.g. 90%) risk of the presence of bladder cancer, a second associated

with a low risk of bladder cancer 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.

Data Analysis

Once the method to test for the amount of RNA and/or protein has been completed, the data then has to be analyzed in order to determine the distribution of biomarker values associated with tumor and non-tumor samples. This typically involves normalizing the raw data, i.e., removing background "noise" etc and averaging any duplicates (or more), comparison with standards and establishing cut-offs or thresholds to optimally separate the two classes of samples. Many methods are known to do this, and the exact method will depend on specific method for determining the amount of RNA and/or protein used.

Below is an example of how the data analysis could be performed when using qRT-PCR. However, it will be appreciated the general process could be adapted to be used for other methods of establishing the RNA and/or protein content, or other methods could be established by someone skilled in the art to achieve the same result.

Data

Measurements of fluorescence are taken at wavelengths ω_i i=1,2 at each cycle of the PCR. Thus for each well we observe a pair of fluorescence curves, denoted by $f_t(\omega_i)$, where $t=1,\cdots,k$ denotes cycle number and $i = 1,2$ indexes the wavelengths.

Fluorescence curves have a sigmoidal shape beginning with a near horizontal baseline and increasing smoothly to an upper asymptote. The location of a point C_p where the fluorescence curve departs from the linear baseline will be used to characterize the concentration of the target gene. A precise definition of C_p follows later.

The following is an example of a scheme to process these data.

- Compensate for fluorescence overlap between frequency bands,
- Estimate a smooth model for each fluorescence curve in order to estimate C_p
- Combine data from replicated wells.
- Estimate standard curves
- Compute a concentration relative to the standard.

Each biological sample yields relative concentrations of 5 genes, which are the inputs to the discriminant function.

Color compensation

Denote the level of fluorescence of dye $_j$ at cycle t and frequency ω by $W_{tj}(\omega)$. In a multiplexed assay the measured response at any frequency ω is the sum of contributions from all dyes at that frequency, so for each cycle.

$$f_t(\omega) = W_{t1}(\omega) + W_{t2}(\omega) + \ldots$$

The purpose of color compensation is to extract the individual contributions $W_{tj}(\omega)$, from the observed mixtures $f_t(\omega)$.

In the ideal situation, fluorescence $W_{tj}(\omega_0)$, due to dye j at a frequency ω is proportional to its fluorescence $W_{tj}(\omega_0)$ at reference frequency ω_0, regardless of the level of $W_{tj}(\omega_0)$. This suggests the linear relationship

$$\begin{bmatrix} f_t(\omega_1) \\ f_t(\omega_2) \end{bmatrix} = \begin{bmatrix} W_{t1}(\omega_1) + W_{t2}(\omega_1) \\ W_{t1}(\omega_2) + W_{t2}(\omega_2) \end{bmatrix} = \begin{bmatrix} 1 & A_{12} \\ A_{21} & 1 \end{bmatrix} \begin{bmatrix} W_{t1}(\omega_1) \\ W_{t2}(\omega_2) \end{bmatrix}$$

for some proportionality constants A_{12} and A_{21} that are to be determined.

In reality, there are additional effects, which are effectively modeled by introducing linear terms in this system, so

$$\begin{bmatrix} f_t(\omega_1) \\ f_t(\omega_2) \end{bmatrix} = \begin{bmatrix} 1 & A_{12} \\ A_{21} & 1 \end{bmatrix} \begin{bmatrix} W_{t1}(\omega_1) \\ W_{t2}(\omega_2) \end{bmatrix} + \begin{bmatrix} a_1 + b_1 t \\ a_2 + b_2 t \end{bmatrix}$$

After estimating the "color compensation" parameters A_{12} and A_{21} we can recover $W_{t1}(\omega_1)$ and $W_{t2}(\omega_2)$, albeit distorted by a linear baseline, by matrix multiplication:

$$\begin{bmatrix} W_{t1}(\omega_1) \\ W_{t2}(\omega_2) \end{bmatrix} = \begin{bmatrix} 1 & A_{12} \\ A_{21} & 1 \end{bmatrix}^{-1} \begin{bmatrix} f_t(\omega_1) \\ f_t(\omega_2) \end{bmatrix} + \begin{bmatrix} a_1^* + b_1^* t \\ a_2^* + b_2^* t \end{bmatrix}$$

$W_{t1}(\omega_1)$ and $W_{t2}(\omega_2)$ are called "color compensated" data. The linear distortions $a_i^* + b_i^* t$ in the last term of this expression will be accommodated in the baseline estimate when estimating a model for the colour compensated data below 2. It has no influence on the estimate of C_p.

Estimation of the color compensation coefficients requires a separate assay using single (as opposed to duplex) probes. Then $W_{t2}(\omega_2) = 0$ giving:

$$\begin{bmatrix} f_t(\omega_1) \\ f_t(\omega_2) \end{bmatrix} = \begin{bmatrix} 1 & A_{12} \\ A_{21} & 1 \end{bmatrix} \begin{bmatrix} W_{t1}(\omega_1) \\ 0 \end{bmatrix} + \begin{bmatrix} a_1 + b_1 t \\ a_2 + b_2 t \end{bmatrix}$$

Thus,

$$f_t(\omega_2) = A_{21} f_t(\omega_1) + a^* + b^* t$$

The coefficient A_{21} can be estimated by ordinary linear regression of $f_t(\omega_2)$ on $f_t(\omega_1)$ and PCR cycle t for $t = 1,\cdots,k$.

Model Estimation

In this section, let y_t $t=1,\cdots,k$ denote a color compensated fluorescence curve.

Amplification

Models are only estimated for fluorescence curves that show non-trivial amplification. We define the term "amplification" as a non-trivial departure from the linear baseline of the color compensated fluorescence curve. Use signal to noise ratio (SNR) to quantify amplification. Here SNR is defined as the ratio of signal variance to noise variance. Noise variance is set as part of calibration of the assay procedure and remains unchanged: for this purpose, use the residual variance from a linear model for the baseline from wells that can have no amplification, i.e., wells without RNA. For each fluorescence curve, estimate the signal variance as the residual variance from the best fitting straight line ("best" is meant in the least squares sense.)

- If SNR is less than a specified threshold, the fluorescence curve is close to linear and no amplification is present. Then there is no point of departure from the baseline and the concentration in the sample may be declared as zero.
- If the SNR is above the threshold, amplification is present and a concentration can be estimated.

Thresholds for the (dimensionless) SNR are selected to provide clear discrimination between "amplified" and "non amplified" curves. For example, the following ranges for thresholds are effective for the markers.

Fluor	Gene	Range
JOE	MDK	40-120
JOE	CDC	35-70
JOE	IL8R	30-60
FAM	IGF	50-80
FAM	HOXA	50-150
FAM	XENO	50-80

Model

Estimate a sigmoidal model for each fluorescence curve. Any suitable parametric form of model can be used, but it must be able to model the following features:

- linear baseline that may have a non-zero slope,
- asymmetries about the mid point.
- asymptotes at lower and upper levels
- smooth increase from baseline to upper asymptote

An example of a model that achieves these requirements is

$$g_t(\theta) = A + A_s t + \frac{D}{\left(1 + \left(\frac{t}{B}\right)^E\right)^F}$$

We call this the "6PL model". The parameter vector $\theta = [A,As,D,B,E,F]$ is subject to the following constraints to ensure that $g_t(\theta)$ is an increasing function of t and has the empirical properties of a fluorescence curve.

$$D > 0, B > 0, E < 0, F < 0$$

The other two parameters determine the base line $A + A_s t$, and these parameters do not need explicit constraints though A is always positive and the slope parameter A_s is always small. The parameter D determines the level of amplification above the baseline. The remaining parameters $B.E.F$ have no intrinsic interpretation in themselves but control the

shape of the curve. These parameters are also the only parameters that influence the estimate of C_p. When $A_s = 0$ this is known as the five-parameter logistic function (5PL) and if, in addition, $F = 1$ this model reduces four-parameter logistic model (4PL), Gottschalk and Dunn (2005), Spiess et al. (2008).

Initialization

Initial values for non-linear estimation are set as

- $A_s = 0$, $F = 1$
- $A = \text{mean}(y_1, \cdots, y_5)$
- $D = \text{range}(y_1, \cdots, yk)$
- B = cycle corresponding to half height
- E is initialized by converting $g_t(\theta)$ into a linear form having set the values of the remaining parameters to their initial values defined above. Linearization obtains

$$E \log\left(1 + \frac{t}{B}\right) = \log\left(\frac{D}{y_t - A}\right)$$

Now estimate E by regression of $\log\left(\frac{D}{y_t - A}\right)$ on $\log\left(1 + \frac{t}{B}\right)$ for t selected so that

$$A + \frac{D}{10} < y_t < A + \frac{9D}{10}$$

An alternative form of this model that leads to an almost identical analysis (with its own initialization) is:

$$A + A_s t + \frac{D}{\left(1 + \exp\left(-\frac{t - B}{E}\right)\right)^F}$$

When $A_s = 0$ this is sometimes known as the Richards equation, Richards (1959).

Estimation criterion

Estimate parameters to minimize a penalized sum of squares criterion:

$$\sum_t (y_t - g_t(\theta))^2 + \lambda(\theta)$$

Here $\lambda(\theta)$ is a non-negative function that penalizes large values of some (or all) of the parameters in θ. This method is known as regularization or ridge regression (Hoerl, 1962) and may be derived from a Bayesian viewpoint by setting a suitable prior distribution for the parameter vector θ. A satisfactory choice for the penalty is:

$$\lambda(\theta) = \lambda(B^2 + D^2 + E^2 + F^2)$$

Large values of λ bias the parameter estimates towards zero and reduce the variance of the parameter estimates. Conversely, small (or zero) λ leads to unstable parameter estimates and convergence difficulties in minimization algorithms. The choice of λ is a compromise between bias and variance or stability. Empirical evidence shows that a satisfactory compromise between bias and variance may be achieved if λ is chosen in the range:

$$0.01 > \lambda > 0.0001.$$

This choice also ensures convergence of the optimization algorithm.

Algorithm choice

For any choice of λ in the above range, the description in the previous paragraph completely defines the parameter estimates. A non-linear least squares procedure based on the classical Gauss-Newton procedure (such as the Levenberg-Marquardt algorithm as implemented in More, 1978) has been successfully used and is a suitable approach. General purpose optimizing algorithms such as Nelder and Mead, 1965, or Broyden-Fletcher algorithm as implemented by Byrd, et al., 1995) have also been successfully trialed in this context.

C_p estimate

C_p is the point at time t that maximizes the second derivative of $g_t(\theta)$. Each fluorescence curve yields a C_p that characterizes the concentration of the target gene. The average of the estimated C_ps for each set of technical replicates is computed and used in the subsequent analysis.

Standard curves

Absolute or relative concentrations are derived from a comparison with standard curves on the same PCR plate. Model a dilution series using the linear model:

$$C_p = R + S \log_{10} Conc$$

where *Conc* is an absolute or relative concentration of the standard. The intercept and slope parameters are plate specific. Model the between-plate variability in the intercept and slope parameters by setting population models

$$R \sim N(\mu_R, \sigma_R^2)$$
$$S \sim N(\mu_S, \sigma_S^2)$$

where the parameters $\mu_R, \sigma_R^2, \mu_S, \sigma_S^2$ are set on the basis of prior data as described below. Then for a given plate R and S can be interpreted as observations from these populations.

For replicate i of standard at concentration $Conc_j$ the following model can be used:

$$C_p(i,j) = R + S \log_{10} Conc_j + \epsilon_{ij}$$

where $\epsilon_{ij} \sim N(0, \sigma_j^2)$. Note that the variance of the residuals depends on C_p. Empirical estimates of $Var(\epsilon_{ij})$ are given in **Table 2**. Estimate the parameters R and S using by maximizing the likelihood function. Interpret the slope parameter in terms of the efficiency of the PCR process through the expression:

$$S = -\frac{1}{\log_{10} Efficiency}$$

This model has a Bayesian interpretation: Give vague (non-informative) prior distributions to the parameters $\mu_R, \sigma_R^2, \mu_S, \sigma_S^2$. Then the population models for R and S and for $C_p(i,j)$ fully determine a probability model for the prior data. A Markov chain Monte Carlo (MCMC) algorithm (Lunn et al., 2009) allows estimation of $\mu_R, \sigma_R^2, \mu_S, \sigma_S^2$. If the prior distribution is omitted, a traditional frequentist interpretation results. Following this estimation procedure it is possible to obtain the gene-dependent population parameter estimates in **Table 3**.

Table 2: Variance of Residuals

C_p	σ^2	C_p	σ^2
12	0.0100	21	0.0466
13	0.0108	22	0.0625
14	0.0119	23	0.0860
15	0.0134	24	0.1212
16	0.0155	25	0.1750
17	0.0184	26	0.2591
18	0.0224	27	0.3931
19	0.0279	28	0.6112
20	0.0356	29	0.9741

Table 3: Population Parameters for Slopes and Intercepts of Standard Curves

	μ_R	σ_R^2	μ_S	σ_S^2
MDK	19.49	0.5112	-3.426	0.0481
CDC	18.91	0.2343	-3.414	0.0198
IL8R	31.43	0.0919	-3.192	0.0017
IGF	20.63	0.3835	-3.275	0.0247
HOXA	22.51	0.1544	-3.270	0.0037

The estimates of intercept and slope of the standard curve are denoted by $\hat{R}$ and $\hat{S}$.

Relative concentrations ΔC_p

Use the standard curve to compute $C_{p(REF)}$ at the concentration $Conc_{REF}$ from the expression: $C_{p(REF)} = \hat{R} + \hat{S}\log_{10} Conc_{REF}$. The relative concentration of a sample is given by the expression:

$$\Delta C_p = \frac{C_p - C_{p(REF)}}{\hat{S}} = \log_{10} \frac{Conc_{SAMPLE}}{Conc_{REF}}$$

Alternatively $\hat{S}$ may be approximated at a fixed level corresponding to a PCR efficiency of 2. Then $\hat{S} = -1/log_{10}(2) = -3.32$. Use the same notation ΔC_P for either choice. The resulting ΔC_P estimates, one for each gene, are inputs to the discriminant function in the next step.

Discriminant Function

The ΔC_P values correspond to a relative biomarker value with plate-to-plate variation removed. Examination of the 5 ΔC_P values in comparison with each other (for example, see **FIG. 2**), shows how tumor samples typically have different biomarker values than non-tumor samples. Furthermore, while there is overlap in the areas for tumor and normal, a large number of samples are effectively well separated. Under these circumstances, many different statistical classifiers could be used to separate the normal from the tumor samples. We show here that a sample of several classifiers do work to separate these samples. We used 5 different classification methods: 1) Linear Discriminant Analysis (LDA), 2) Logistic Regression (LogReg); 3) Support Vector Machines (SVM); 4) K-nearest-neighbor (KNN) based on 5 neighbors (KN5N); and 5) Recursive partitioning trees (TREE) (Cite: Venables & Ripley and Dalgaard).

Creation of a classifier requires a dataset containing the biomarker values for a large number of samples which should represent the ultimate population to be tested by the classifier. For example, if a classifier is to be used for screening an at-risk population (eg age 50 and older, smokers), then the set of data required for creating the classifier (called the "training set") should mirror that population and contain only samples from people older than 50 who smoke. Typically to obtain measurement precision of smaller than 10% error for parameters like sensitivity and specificity, the training set needs to be larger than 300 samples.

Estimation of the effectiveness of a classifier can be made using cross-validation. In cross-validation (Wikipedia: Cross-validation), the dataset is divided into a small number of equally sized partitions (typically 3 to 10). One section is left out and the remaining sections used to build a classifier; then the left out section is tested by the new classifier and its predictions noted. This is done for each section in turn and all the predictions combined and analysed to compute the characteristics of the classifier: Sensitivity, Specificity, etc. If the cross-validation is performed by partitioning the data into 10 parts, it is called 10-fold cross-validation; similarly, 3 parts would be 3-fold cross-validation. If the data are partitioned into

as many classes are there are samples, this is called "leave one out cross-validation". By testing on data not used to build the classifier, this method provides an estimate of the classifier performance in the absence of additional samples.

We have built classifiers using all 15 combinations of 4 biomarkers, MDK, IGFBP5, CDC2, and HOXA13, all with and without the IL8Rb biomarker, using the clinical trial dataset described elsewhere in this document (Example 1) and tested those 30 classifiers using 10-fold cross-validation. This was done for each of the 5 classifier types listed above and the ROC curves computed. All work was erformed using the R Statistical Programming Environment (CITE).These results (**FIG. 14**) show that in most cases, the classifier with IL8Rb is more sensitive for values of specificity which are useful diagnostically (False Positive Rate of 0 to 20%; Specificity from 100 to 80%). The Area Under the Curve (AUC) for the region with diagnostic utility of specificities is used to quantify how well classifiers perform with larger values indicating better classifier performance. **FIG. 15a** tabulates the AUC for each classifier and biomarker combination, while Figure 15b shows the amount of increase in AUC for each condition when IL8Rb is added. In most cases, the addition of IL8Rb improves the ability to make accurate diagnoses. Specific sensitivity values for diagnostically useful specificity values are tabulated for all the classifiers in **FIG. 16**. In addition, **FIG. 17** tabulates the amount of gain in sensitivity or specificity which the addition of IL8Rb provides.

The utility of the classifier is created when, having created it and tested it, it is used to test a new sample. To simplify the interpretation of results, a cut-off score or threshold is established; samples on one side of the cut-off are considered positive and on the other side, negative for tumors. Additional cut-offs may be established for example to indicate increasing levels of certainty of results. In this case, we have established a cut-off which gives a false positive rate of 15% in our training set. Using our cross-validated ROC curves, we can then estimate our sensitivity. Typically, we also establish a cut-off at a positive predictive value of 75%. To use these cut-offs we establish a "negative" result for scores less than the cut-off established by the 85% specificity. Scores greater than the 75% PPV are called "positive" and score between the two are called "indeterminate" " or "suspicious".

Antibodies to IL8Rb

In additional aspects, this invention includes manufacture of antibodies against IL8Rb. The marker IL8Rb can be produced in sufficient amount to be suitable for eliciting an immunological response. In some cases, a full-length IL8Rb can be used, and in others, a peptide fragment of a IL8Rb 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 or 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, including monoclonal or phage-display antibodies, against IL8Rb.

In yet further embodiments, antibodies can be made against the protein or the protein core of the tumor markers identified herein or against an oligonucleotide sequence unique to a IL8Rb. Although certain proteins can be glycosylated, variations in the pattern of glycosylation can, in certain circumstances, lead to mis-detection of forms of IL8Rb that lack usual glycosylation patterns. Thus, in certain aspects of this invention, IL8Rb immunogens can include deglycosylated IL8Rb or deglycosylated IL8Rb fragments. Deglycosylation can be accomplished using one or more glycosidases known in the art. Alternatively, IL8Rb cDNA can be expressed in glycosylation-deficient cell lines, such as prokaryotic cell lines, including E. coli and the like.

Expression vectors can be made having IL8Rb-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 IL8Rb -producing cell lines, which can be used to produce desired quantities of IL8Rb for development of specific antibodies or other reagents for detection of IL8Rb or for standardizing developed assays for IL8Rb.

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 IL8Rb 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 multi-well plate) can have a specific IL8Rb and BTM capture reagents 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 IL8Rb and 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 marker associated molecules can be used and be considered within the scope of this invention.

Antibodies can also be used when bound to a solid support, for example using an antibody chip, which would allow for the detection of multiple markers with a single chip.

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.

Detection of IL8Rb and BTMs in a sample can be performed 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 an inflammatory disease or tumor. The marker may be secreted into the serum or other body fluid. 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 Tests for Inflammatory Disease or Cancer Markers in Body Fluids

In general, methods for assaying for oligonucleotides, proteins and peptides in these fluids are known in the art. Detection of oligonucleotides can be carried out using hybridization methods such as Northern blots, Southern blots or microarray methods, or qPCR. Methods for detecting proteins include such as enzyme linked immunosorbent assays (ELISA), protein chips having antibodies, suspension beads radioimmunoassay (RIA), Western blotting and lectin binding. However, for purposes of illustration, fluid levels of a disease markers can be quantified using a sandwich-type enzyme-linked immunosorbent assay (ELISA). For plasma assays, a 5 uL aliquot of a properly diluted sample or serially diluted standard marker and 75 uL of peroxidaseconjugated anti-human marker 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 marker and anti- marker antibody are then incubated with o-phenylendiamine containing H_20_2 for 15 minutes at 30°C. The reaction is stopped by adding 1 M H_2SO_4, and the absorbance at 492 nm is measured with a microtiter plate reader.

It can be appreciated that anti- IL8Rb antibodies can be monoclonal antibodies or polyclonal antisera. It can also be appreciated that any other body fluid can be suitably studied.

It is not necessary for a marker to be secreted, in a physiological sense, to be useful. Rather, any mechanism by which a marker protein or gene enters the serum can be effective in producing a detectable, quantifiable level of the marker. Thus, normal secretion of soluble proteins from cells, sloughing of membrane proteins from plasma membranes, secretion of alternatively spliced forms of mRNA or proteins expressed therefrom, cell death (either apoptotic) can produce sufficient levels of the marker to be useful.

There is increasing support for the use of serum markers as tools to diagnose and/or evaluate efficacy of therapy for a variety of cancer types.

EXAMPLES

The examples described herein are for purposes of illustrating embodiments 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 are considered to be part of this invention.

Example 1: Genotypic Analysis of Bladder Cancer

Methods

Patients: Between April 2008 and September 2009, 485 patients presenting with macroscopic hematuria, but no prior history of urinary tract malignancy, were recruited at eleven urology clinics in New Zealand and Australia. Each patient provided a urine sample immediately prior to undergoing cystoscopy and any additional diagnostic procedures. A diagnosis was made by three months following enrollment in the study. Of these 485 patients, gene expression data on all five study genes was successfully obtained for 442 patients using the methods described below. The characteristics of these patients are shown in **Table 4**.

Table 4: Characteristics of the Study Population 1

Diagnosis	**Number**
Benign prostatic hyperplasia	18
Cystitis	18
Exercise-induced hematuria	3
Non-specific kidney disease	3
Non-specific neoplasia	3
Non-specific prostate disease	63
Vascular prostate	49
Other urological cancer (non-TCC)	5
Superficial vessels	3
Urethral stricture	6
Urinary tract infection	18
Urolithiasis	25
Warfarin use	10
Unknown etiology	155
Miscellaneous	7
TCC	56
Total	**442**

Table 4 shows the number of patients in each of the main diagnostic categories at three months after the patient's initial presentation with gross hematuria.

Urine analysis: Urine samples were analyzed by central review cytology (Southern Community Laboratories, Dunedin, New Zealand). The diagnostic tests NMP22

BladderChek® (Matritech) and NMP22 ELISA (Matritech) were carried out according to the manufacturer's instructions at the clinical site (BladderChek®) or by Southern Community Laboratories (NMP22 ELISA).

RNA quantification: 2mls or urine from each patient was mixed with RNA extraction buffer containing 5.64M guanidine thiocyanate, 0.5% sarkosyl and 50mM NaoAc pH6.5. Total RNA was then extracted by Trizol extraction (Invitrogen) and the RNeasy procedure (Qiagen), as previously described1. RNA was eluted from the columns in 35ul water and 3ul was used in each subsequent monoplex or duplex quantitative reverse transcription polymerase chain reaction (qRT-PCR) assay. Each 16ul qRT-PCR reaction contained 0.3U RNAse-OUT (Invitrogen), 0.225uM each Taqman probe, 1.25U Superscript III (Invitrogen), 0.275uM each primer, 1.5U Fast Start Taq polymerase (Roche), 10mM DTT, 0.375mM dNTPs, 4.5mM MgSO4, 1.6ul 10x Fast Start PCR buffer (Roche) and 2.6ul GC Rich solution (Roche). Primers and fluorescently dual-labeled probes were obtained from Integrated DNA Technologies (Coralville USA) for each of the five study genes: MDK, CDC2, HOXA13, IGFBP5 and IL8Rb. Primer/probe sequences are shown in Table 2. Reactions were set up in 96 well plates and cycled as follows on a Roche Light Cycler® 480: 50°C, 15 mins; 95°C 8 mins; 10 cycles of 95°C 15 sec, 60°C 2 mins and 30 cycles of 95°C 15 secs, 60°C 1 min. Standard curves of 1/16 serial dilutions of a reference RNA (derived from pooled cell line RNAs) were included on each plate to generate range of 0.3 pg/µl to 20 ng/µl. Data was collected at the extension phase of the final 30 thermocycles and exported as a raw text file. Table 5 below depicts primers and probe sequences used for qRT-PCR quantification of the five RNA markers.

Table 5

Marker	Forward Seq	Reverse Seq	Probe
MDK	TGC ACC CCC AAG ACC AAA (SEQ ID NO. 3)	TGA TTA AAG CTA ACG AGC AGA CAG AA (SEQ ID NO. 4)	CCT TCC CTT TCT TGG CTT TGG CCT TT (SEQ ID NO. 5)
IGFBP5	CGT TGT ACC TGC CCA ATT GTG A (SEQ ID NO. 6)	GGG ACG CAT CAC TCA ACG TT (SEQ ID NO. 7)	AAG AGA AAG CAG TGC AAA CCT TCC CGT (SEQ ID NO. 8)
CDC2	GCC GCC GCG GAA TAA T (SEQ ID NO. 9)	TGT CTA CCC TTA TAC ACA ACT CCA TAG G (SEQ ID NO. 10)	AGC CGG GAT CTA CCA TAC CCA TTG ACT AAC T

			(SEQ ID NO. 11)
HOXA13	TGG AAC GGC CAA ATG TAC TG (SEQ ID NO. 12)	TGG CGT ATT CCC GTT CAA GT (SEQ ID NO.13)	ACT CTG CCC GAC GTG GTC TCC CA (SEQ ID NO.14)
IL8Rb	CCT TGA GGC ACA GTG AAG ACA TC (SEQ ID NO. 15)	CCT GTA GGA CAC CTC CAG AAG AG (SEQ ID NO.16)	TGG CCA CTC CAA TAA CAG CAG GTC ACA (SEQ ID NO. 17)

qRT-PCR Data Analysis

Raw fluorescence data was exported from the Roche LightCycler® 480 as a tab-delimited file containing cycle number versus two channels of fluorescence data for all wells on the plate. The data were processed using an R program that applied color compensation ([Bernard1999]) to the data to correct for bleed over from one fluorescent channel into another. It then fitted a 5-point logistic model to estimate the C_P using the second derivative maximum ([Spiess2008]).

All samples and controls were applied in duplicate to the PCR plates. The C_P values from the duplicate wells were averaged before use. If the difference between the two C_P values exceeded 3 units, that sample was repeated. To provide standardization across PCR plates, C_P's were expressed as ΔC_P's relative to a reference RNA (derived from pooled cell line RNAs) at 20 ng/µl:

$$\Delta C_P = C_P \text{(sample)} - C_P \text{(reference RNA)}$$

Statistical Analysis

qRT-PCR ΔC_P values from MDK, CDC2, HOXA13, IGFBP5 and IL8Rb were used to generate classifiers to separate samples containing TCCs from samples containing no TCCs, based on Linear Discriminant Analysis or Logistic Regression ([Venables2002]). In both cases, interactions between genes were permitted in the classifier models. The generation of the LDA followed standard procedures, as described, for example in "Modern Applied Statistics with S, 4th edition" by W.N. Venables and B.D. Ripley (2002), Springer. The dataset from the study was cleaned of any incomplete data then the R Statistical Environment (R Development Core Team (2009) and the function "lda" from the package MASS (Venables and Ripley (2002)) were used to generate and test the linear discriminant on the clinical trial

data.

The generation of the Logistic Regression classifier was performed in a similar manner to the generation of the LDA. Again, the study data was cleaned of incomplete data. A logistic regression classifier was created using R; no additional packages were required. Logistic regression was performed as described by Dalgaard (2008). Comparison among classifiers was made using ROC curves, using the R package, ROCR (Sing et al. 2009). Confidence intervals for ROC curves were generated using the methods of Macskassy et al ([Macskassy2005]). The following algorithms were generated:

Linear Discriminant Classifier

The first classifier, a linear discriminant, (called LDA-3), Is based on five gene values (normalized to a Reference value by subtracting the reference value) allowing for multiway interactions between the genes. The classifier was built in R using the 'lda()' function from the package called "MASS". (R version 2.9.1; MASS version 7.2-49). The classifier was built using the following equation:

lda3 <- lda(TCC.YN ~ MDK * IGF * CDC * HOXA * IL8R, data=uRNA.Trial)

Where lda3 is the created model; TCC.YN is the true value for "presence of TCC in urine"(Yes or No) as determined by cystoscopy; MDK, IGF, CDC, HOXA and IL8R are the normalized gene Cp value; and uRNA Trial is a data file containing the Cp values for the each of the 5 genes and TCC.YN (yes or no) from the clinical trial. Use of the ASTERISK, '*', in the formula signifies multiplication. Evaluation of the classifier score takes as input a new data frame containing the five gene values as well as the classifier, lda3, to output a classifier score:

score <- c(predict(lda3, new.data)$x),

where "score" is the output used from the classifier to predict the presence of TCCs; "lda3" is the classifier created above and "new.data" is a data FILE containing the measured values of the five genes called by the same names as used in classifier creation. The syntax, '$x' and "c(...)" is present to extract the score specifically from the large amount of information returned by the predict function. Setting the score cut off to 0.112 and above, sets our specificity to 85% for presence of TCCs in the urine sample. The coefficients for LDA-3 are

shown in **Table 6** : **Table 6**

MDK.d.R100	5.333639e+00
IGF.d.R100	3.905978e+00
CDC.d.R100	6.877143e-01
HOXA.d.R100	6.073742e+00
IL8R.d.R100	-1.229466e+00
MDK.d.R100:IGF.d.R100	-7.420480e-01
MDK.d.R100:CDC.d.R100	-2.611158e-01
IGF.d.R100:CDC.d.R100	-1.965410e-01
MDK.d.R100:HOXA.d.R100	-8.491556e-01
IGF.d.R100:HOXA.d.R100	-4.037102e-01
CDC.d.R100:HOXA.d.R100	-3.429627e-01
MDK.d.R100:IL8R.d.R100	1.903118e-01
IGF.d.R100:IL8R.d.R100	2.684005e-01
CDC.d.R100:IL8R.d.R100	-1.229809e-01
HOXA.d.R100:IL8R.d.R100	2.909062e-01
MDK.d.R100:IGF.d.R100:CDC.d.R100	4.108895e-02
MDK.d.R100:IGF.d.R100:HOXA.d.R100	7.664999e-02
MDK.d.R100:CDC.d.R100:HOXA.d.R100	4.832034e-02
IGF.d.R100:CDC.d.R100:HOXA.d.R100	2.116340e-02
MDK.d.R100:IGF.d.R100:IL8R.d.R100	-3.750854e-02
MDK.d.R100:CDC.d.R100:IL8R.d.R100	1.664612e-02
IGF.d.R100:CDC.d.R100:IL8R.d.R100	2.089442e-03
MDK.d.R100:HOXA.d.R100:IL8R.d.R100	-1.539486e-02
IGF.d.R100:HOXA.d.R100:IL8R.d.R100	-3.894153e-02
CDC.d.R100:HOXA.d.R100:IL8R.d.R100	6.295032e-03
MDK.d.R100:IGF.d.R100:CDC.d.R100:HOXA.d.R100	-4.359738e-03
MDK.d.R100:IGF.d.R100:CDC.d.R100:IL8R.d.R100	-2.019317e-04
MDK.d.R100:IGF.d.R100:HOXA.d.R100:IL8R.d.R100	3.746882e-03
MDK.d.R100:CDC.d.R100:HOXA.d.R100:IL8R.d.R100	-2.902150e-03
IGF.d.R100:CDC.d.R100:HOXA.d.R100:IL8R.d.R100	4.799489e-04
MDK.d.R100:IGF.d.R100:CDC.d.R100:HOXA.d.R100:IL8R.d.R100	7.512308e-05

Logistic Regression Classifier

A second classifier based on Logistic Regression was derived from the same cleaned dataset as LDA-3. Instead of using the lda() function, however, we used the glm() function from the package stats (included with a base install of R) as shown below:

```
lr1 <- glm(TCC.YN ~ CDC * IGF * HOXA * IL8R * MDK,
    family=binomial("logit"), data=uRNA.Trial),
```

where "lr1" is the classifier created and the other parameters are as described for the linear discriminant. Once again, full interaction is specified using the '*' operator. Classification is

performed in a manner very similar to that for LDA-3:

score <- predict(lr1, new.data, type='response'),

where "score" is the value used to classify urine samples based on the measurement of the five genes in "new.data", as above. The cut off for lr1 is set to 0.102 to achieve a specificity of 85%; values about the cut off are considered to be positive to TCCs. The coefficients for the classifier are:

-103.0818143 +
3.9043769 * CDC.d.R100 +
13.1120675 * IGF.d.R100 +
17.4771819 * HOXA.d.R100 +
-10.7711519 * IL8R.d.R100 +
21.1027595 * MDK.d.R100 +
-0.5938881 * CDC.d.R100 * IGF.d.R100 +
-1.0736184 * CDC.d.R100 * HOXA.d.R100 +
-1.3340189 * IGF.d.R100 * HOXA.d.R100 +
0.3126461 * CDC.d.R100 * IL8R.d.R100 +
1.4597355 * IGF.d.R100 * IL8R.d.R100 +
1.8739459 * HOXA.d.R100 * IL8R.d.R100 +
-1.035054 * CDC.d.R100 * MDK.d.R100 +
-2.5885156 * IGF.d.R100 * MDK.d.R100 +
-2.7013483 * HOXA.d.R100 * MDK.d.R100 +
1.4546134 * IL8R.d.R100 * MDK.d.R100 +
0.0767503 * CDC.d.R100 * IGF.d.R100 * HOXA.d.R100 +
-0.0663361 * CDC.d.R100 * IGF.d.R100 * IL8R.d.R100 +
-0.1015552 * CDC.d.R100 * HOXA.d.R100 * IL8R.d.R100 +
-0.2110656 * IGF.d.R100 * HOXA.d.R100 * IL8R.d.R100 +
0.1361215 * CDC.d.R100 * IGF.d.R100 * MDK.d.R100 +
0.1601118 * CDC.d.R100 * HOXA.d.R100 * MDK.d.R100 +
0.259745 * IGF.d.R100 * HOXA.d.R100 * MDK.d.R100 +
-0.0106468 * CDC.d.R100 * IL8R.d.R100 * MDK.d.R100 +
-0.1947899 * IGF.d.R100 * IL8R.d.R100 * MDK.d.R100 +
-0.185286 * HOXA.d.R100 * IL8R.d.R100 * MDK.d.R100 +
0.0136603 * CDC.d.R100 * IGF.d.R100 * HOXA.d.R100 * IL8R.d.R100 +
-0.0151368 * CDC.d.R100 * IGF.d.R100 * HOXA.d.R100 * MDK.d.R100 +
0.0056651 * CDC.d.R100 * IGF.d.R100 * IL8R.d.R100 * MDK.d.R100 +
0.0030538 * CDC.d.R100 * HOXA.d.R100 * IL8R.d.R100 * MDK.d.R100 +
0.0232556 * IGF.d.R100 * HOXA.d.R100 * IL8R.d.R100 * MDK.d.R100 +
-0.000867 * CDC.d.R100 * IGF.d.R100 * HOXA.d.R100 * IL8R.d.R100 * MDK.d.R100

Results

qRT-PCR analysis of urine samples

To obtain an overview of the effect of IL8Rb on TCC detection, two dimensional

scatter plots were constructed using qRT-PCR data obtained from the urine of patients with either TCC (n= 56) or the non-malignant conditions urolithiasis (n=25), urinary tract infection (n=18) or cystitis (n=18). The scatter plots were constructed using pairs of genes from a four gene signature (MDK, CDC2, HOXA13, IGFBP5). IL8Rb was then substituted for one gene of each pair and the data re-plotted. These plots are shown in **FIG. 2a-f**. Substitution of IL8Rb for IGFBP5 and HOXA13 in plots with MDK (**FIG. 2a-c**) showed improved separation between samples from patients with TCC and those with non-malignant conditions. The same trend was observed in plots with CDC2 in which IL8Rb was substituted for IGFBP5 and HOXA13 (**FIG. 2d-f**).

The contribution of IL8Rb to the correct diagnosis of TCC in patients presenting with gross hematuria was then quantified by ROC curve analysis. qRT-PCR data for each gene in the signature (MDK, CDC2, IGFBP5 and HOXA13) and IL8Rb was used to develop linear discriminate algorithms that maximized the discrimination between the patients with TCC and those without. Two linear discriminate algorithms were developed using the entire cohort of 442 samples: LD1, which used the qRT-PCR data from MDK, CDC2, HOXA13 and IGFBP5 and LD2, which used MDK, CDC2, HOXA13, IGFBP5 and IL8Rb. LD1 and LD2 were then used to generate ROC curves showing the sensitivity and specificity of TCC detection in the group of patients with confirmed TCC (n=56) or the non-malignant conditions urolithiasis (n=25), urinary tract infection (n=18) or cystitis (n=18). Figure 3a shows the ROC curves for LD1 and LD2. The area under the ROC curve for LD1 was 78% compared to 84% for LD2.

As an alternative to linear discriminate analysis, logistic regression was used as an independent method to develop an algorithm for the discrimination between patients with TCC and those with non-malignant disease. As for the linear discriminate analysis, the logistic regression algorithms were developed using the entire cohort of 442 samples. The ROC curves obtained using logistic regression and the 56 TCC and 61 non-malignant samples described above are shown in **FIG. 3b**. The area under the ROC curve for LR1 (obtained using qRT-PCR data from MDK, CDC2, HoxA13 and IGFBP5) was 80% compared to 86% for LR2 (obtained using qRT-PCR data from MDK, CDC2, HOXA13, IGFBP5 and IL8Rb). This data clearly illustrates that inclusion of IL8Rb in methods for the detection of TCC using urine samples can lead to improved discrimination between patients with TCC and non-malignant diseases such as cystitis, urinary tract infection and urolithiasis.

To confirm the improved accuracy afforded by IL8Rb for the discrimination between patients with TCC and urolithiasis, urinary tract infection or cystitis was maintained in an unselected cohort of patients comprising a larger number and diversity of non-malignant patients, the ROC curve analyses were repeated with the entire cohort of 442 samples described in **Table 1**. In this analysis, the area under the curve for LD1 and LD2 was 86 and 89%, respectively (**FIG. 4a**). Similarly, the area under the curve for LR1 was 87% and for LR2 91% (**FIG. 4b**). This result confirms that IL8Rb leads to improved accuracy in the detection of TCC using urine samples.

This improvement in cancer detection due to the inclusion of IL8Rb was further illustrated by applying LD1/LD2 and LR1/LR2 to the 442 patient cohort and then determining the sensitivity of detection of stage Ta TCC alone. Stage Ta tumors are smaller, more differentiated tumors that are typically more difficult to detect than higher stage tumors. LD1 detected 18/31 (58%) of the Ta tumors compared to 19/31 (61%) for LD2 at a specificity of 85%. LR1 detected 21/31 (68%) compared to 24/31 (77%) for LR2 (specificity of 85%). This data shows that the inclusion of IL8Rb into the LD and LR algorithms increased the sensitivity of detection of stage Ta tumors by up to 9%. In comparison to these RNA tests, the three other bladder cancer tests in this study showed markedly lower accuracy for the detection of Ta tumors: urine cytology (39% sensitivity, 94% specificity), NMP22 ELISA (35% sensitivity, 88% specificity) and NMP22 (BladderChek® "a registered trademark of Matritech, Inc. of Massachusetts, United States") (39% sensitivity, 96% specificity).

IL8Rb as an Aid in the Diagnosis of Inflammation of the Urinary Tract

To determine the ability of IL8Rb to be used in the diagnosis of patients with inflammation of the urinary tract due to causes such as cystitis or urinary tract infections, the urine levels of IL8Rb mRNA in hematuria patients diagnosed with benign prostate hyperplasia, non-specific prostate disease, vascular prostate, hematuria secondary to warfarin use, and cystitis/urinary tract infection were determined by qRT-PCR. The mean IL8Rb ΔCt levels for each of these conditions were -3.12, -3.10, -2.84, -1.98 and -5.27, respectively. The difference between the mean of the IL8Rb level in patients with cystitis/urinary tract infection and the other non-malignant states combined was determined to be significant ($p=0.001$) using the Wilcoxon rank sum test. Box plots portraying this data are shown in **FIG. 5**. This data shows an elevation of IL8Rb levels in the majority of patients diagnosed

with either cystitis or urinary tract infection compared to the other non-malignant conditions examined. Overlap between plots is likely to be explained by a combination of three factors: (i) the inability of standard clinical practice to correctly diagnose each condition, (ii) co-morbidity (e.g., infection and benign prostate hyperplasia), and (iii) the normal association of high urine neutrophil counts in a subset of patients with benign prostate hyperplasia, non-specific prostate disease, vascular prostate or hematuria secondary to warfarin use. Regardless, given the strict association between inflammation and neutrophil numbers, the quantification of IL8Rb in urine provides an accurate method of detecting inflammation of the urinary tract, be it as a consequence of infection or in association with other non-malignant conditions.

Example 2

Methods

Study Population

A consecutive series of patients without a prior history of TCC were recruited prospectively from nine urology clinics in New Zealand and two in Australia between 28th April 2008 and 11th August 2009. The patient set included the patients used in example 1, but included an additional 46 patients, whose data was not available for the first analysis. The further studying also includes further analysis of the results obtained. The samples were collected and RNA collected and tested as described in **Example 1**.

RNA Test Development

uRNA® consists of four mRNA markers, *CDC2*, *HOXA13*, *MDK* and *IGFBP5*. These markers were selected on the basis of their low expression in blood and inflammatory cells and over-expression in TCC.[2] In this cohort study, we prospectively specified a linear discriminate algorithm (uRNA-D) that combined the four markers into a single score. uRNA-D was independent, being developed on an earlier dataset. It was not however, derived using a strictly characterized patient group representing the intended target population for the test. As a consequence, the study protocol also defined the development of a new algorithm (*Classifier*-D) for the use of the five markers *CDC2*, *HOXA13*, *MDK*, *IGFBP5* and *IL8Rb* using data obtained from the patients recruited to the current cohort study.

In addition to *Classifier*-D, a second algorithm (*Classifier*-S) was derived using the cohort study data to enable identification of tumors that were either of advancing stage (≥ stage 1) or high grade (WHO/ISUP 1998 classification). Algorithm-S comprised all five markers, including *CDC2* and *HOXA13* which had previously been shown to be differentially expressed between Stage Ta tumors and those ≥ stage 1.

Classifier development

Development of two classifiers for the use of the five markers *CDC2*, *HOXA13*, *MDK*, *IGFBP5* and *IL8Rb* (*Classifier-D* and *Classifier-S)* were based on data obtained in this study, in accordance with the methods outlined in this specification. Briefly, logistic regression models were made using the statistical programming environment, R (R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http:// www.R-project.org/). Models made using ΔC_P values for each of the five markers and their two way interactions (e.g., *MDK* x *CDC2*, *MDK* x *IGFBP5*, etc) were evaluated for their ability to classify; those with the lowest AIC values were evaluated in a leave-one-out cross validation procedure for their sensitivity when the specificity was set to 85%. Several models demonstrated comparable performance for each of *Classifier-D* and *Classifier-S*, with the model with the fewest numbers of parameters being selected.

Statistical Methods

Where a diagnostic test was specified in the protocol, proportions and 95% confidence intervals were calculated for sensitivity and specificity. Receiver operating characteristic (ROC) curves were plotted and compared using the Stata roctab and roccomp commands (Statacorp and Delong). For *Classifier-D* confidence intervals are not appropriate, but Fishers exact or Chi squared tests (where sample sizes allow) were used to test for an association between TCC or patient characteristics and chances of true positive or false positive results. Logistic regression models were used to explore factors associated with false positive and false negative results. All analyses were carried out in Stata version 11.2.

Results

A total of 517 patients were initially recruited to the study; 4% of patients were

excluded because they were found to be ineligible (n=10), did not undergo cystoscopy (n=9), TCC status was not stated (n=2) or they did not provide an acceptable urine sample (n=2) (Figure 8). A further 10 patients were excluded from the analysis because they did not have results for one or more of the urine tests. The baseline demographic and clinical characteristics of the 485 remaining patients are shown in **FIG. 9**.

The prevalence of TCC in the cohort was 13.6%. Two were missing a review stage (both were Ta by local review) and two were not given a review grade (one was grade 1 by the local pathologist, the other low). Of the 66 tumors, 55 were superficial (stage Ta, T1 or Tis) and 11 were muscle invasive (T2). No patients had detectable metastases or involvement of regional lymph nodes. Using the 1973 grading system, 24 were classified as grade 3, 38 grade 2, three grade 1 and one unknown. With the WHO98 system, 29 were classified as high grade, four mixed, 32 low grade and one unknown. In addition to the TCCs, two patients were diagnosed with a papilloma, and seven with other neoplasms (five of these urological).

The cutoff for the uRNA-D test was determined on the study cohort, with specificity set at 85%. With this cutoff, uRNA-D detected 41 of the 66 TCC cases (sensitivity of 62%), compared with NMP22™ ELISA (50%), Bladderchek® (38%) and cytology (56%). The RNA test developed on the cohort data Classifier-*D* detected 54 of the TCC cases (82%) at a specificity of 85% and 48 (73%) at a specificity of 90%. uRNA-D and NMP22™ ELISA values can be directly compared as both tests were fully specified prior to the study. **FIG. 21**shows the ROC curves; the area under the curves (AUCs) are 0.81 and 0.73 respectively (p=0.03). The ROC curve for *Classifier-D* was 0.87 (**FIG. 21**), and the improvement in performance relative to uRNA-D appears to be mostly in the range of clinically relevant specificities (above 80%).

Overall, Classifier-*D* detected 97% of the high/grade 3 tumors, compared to uRNA-D (83%), cytology (83%), NMP22 ELISA (69%) and Bladderchek® (38%). *Classifier-D* was also more sensitive for the detection of low-grade tumors (69%), with the other tests ranging from 28-41% (**FIG. 12**). *Classifier-D* was positive for all the TCC cases of Stage ≥1 plus both Tis, but the sensitivity was 68% for stage Ta (p=0.016, **FIG. 12**). This was still substantially higher than the other tests, with uRNA-D being the next highest at 41%. TCC patients with macrohematuria or microhematuria evident in their urine sample were more likely to have their TCC detected by including IL8Rb than those without macrohematuria or microhematuria (p<0.0005), though this is likely to be at least partially a result of the higher proportion of

high stage and grade TCCs among those with macrohematuria or microhematuria. Numbers were insufficient to explore this further in regression analyses.

Of the 12 cases missed by Classifier-*D*, all were stage Ta and all except one were low grade (WHO ISUP 1998). Only two of the twelve (both low grade, stage Ta TCC) were picked up by another test (one by both NMP22™ ELISA and BladderChek® and one by uRNA-D). Of the 12 cases missed by *Classifier-D*, all were stage Ta and all except one were low grade (WHO ISUP 1998). Only two of the twelve (both low grade, stage Ta TCC) were picked up by another test (one by both NMP22™ ELISA and BladderChek® and one by uRNA-D). Cytology did not pick up any TCCs that *Classifier -D* missed.

Patient A : High Grade renal pelvic T2 tumour, no concurrent Tis, no size given.

Pateint B: High grade Bladder T3a no concurrent Tis, 2x 3cm

Pateint C : a high grade tumour measuring 4.8 x 5.6 cm with extensive stromal and muscularis propria invasion, extending to the perivescical fat with no evidence of metastasis.

The specificity of the urine tests among those with alternative diagnoses and according to urine sample characteristics are shown in **FIG. 13**. Control patients with macrohematuria or microhematuria were more likely to have false positive tests than those without macrohematuria or microhematuria (p=0.002), and there was a trend that patients with calculi may as well, although the differences in specificity by diagnosis were not statistically significant overall (p=0.12). There were five patients with other urological cancers; only one of these gave a positive Classifier-*D* test result. Results from fitting logistic regression models were similar. In a logistic regression model with diagnosis and macrohematuria or microhematuria, the association with macrohematuria or microhematuria status remained significant (p=0.006) and, when compared directly to no diagnosis those with calculi had a 2.7 fold increased odds of a false positive test (95% CI (1.1 to 6.4), p=0.03). Age did not affect the specificity of the test.

Macrohematuria or microhematuria detected in the urine sample was the only factor clearly associated with test sensitivity. The predictive value of a positive test in this cohort was 63% for those with macrohematuria or microhematuria and 24% for those without, largely reflecting the greater prevalence of TCC in the patients with macrohematuria or microhematuria (39% vs 6%).

There were 54 patients with TCC in whom the Classifier-*D* test was positive. These patients were classified into severe and less severe TCC using Classifier-*S*. Severe TCC was

defined as stage ≥1 or grade 3 at any stage. At a specificity of 90%, Classifier-*S* correctly classified 32/35 (91%) of the severe TCC cases.

Example 3: Combined Genotype and Phenotype Analysis of Patients with Hematuria I

This study focuses on patients presenting with confirmed asymptomatic microscopic hematuria (AH) who are undergoing a full clinical work-up for the investigation of possible urothelial cancer (UC). Approximately 500 patients are enrolled to participate in the study.

As used herein, terms are defined below in additional examples.

Objectives

Objectives are to determine the: (1) efficacy of a genotypic and phenotypic algorithm in patients presenting with micro hematuria who are scheduled for a full urological clinical work-up, (2) performance characteristics (sensitivity, specificity, area under the ROC curve, positive and negative predictive values) of the genotypic and phenotypic algorithm G+P INDEX for the detection of primary UC in patients presenting with confirmed microscopic hematuria, and (3) number of patients correctly diagnosed as negative for UC by a genotypic and phenotypic tool, the G+P INDEX and therefore do not require investigative cystoscopy.

Study Population

The study population consists of patients presenting with confirmed micro-hematuria who fulfill study requirements. Patients are recruited from general practices that refer patients to the urological clinics.

Informed Consent

Patients scheduled for investigative cystoscopy are contacted to discuss possible participation. Patients are informed of the nature of the study and consent is obtained. Patients provide demographic, occupational, and smoking history information, and ensure that they fully understand the patient information and consent forms prior to provision of their urine samples. Study coordinators complete a CRF page detailing the relevant inputs to the genotypic and phenotypic index and transfer the data for analysis.

Inclusion Criteria

Patients undergoing cystoscopic investigation for Urothelial Carcinoma following a confirmed clinical finding of microscopic hematuria (Minimum of 3 RBC per high power field (HPF)) on 2 or 3 properly collected urine specimens [3].

Patients are willing to comply with study requirements.

Patients are over 18 years of age.

Exclusion Criteria

Prior history of urothelial caarcinoma (UC).

Current presentation of macroscopic hematuria

Prior history (past 12 months of an episode of Macroscopic Hematuria with confirmed diagnosis (malignant or otherwise).

G+P Index

We developed a novel index, the "G+P Index," which comprises of a combination of both genotypic and phenotypic data. The Genotypic ("G") component utilizes RNA biomarker expression information in conjunction with five clinical factors collected from the patient in the same time window (Phenotypic data "P") to determine the risk of UC in AH patients.

All patients receive a standard clinical work-up to determine true clinical outcome and the outputs from this study are simulated outputs based on the clinical data collected and the genotypic data collected from the patients' urine samples. As such, patient care is not altered as a result of the study output. Patients provide urine samples, which are sent for genetic analysis.

Patient Triage

No change to overall standard of care is made for patients participating in the study. All patients scheduled for a full urological work-up undertake the appropriate investigations according to the current standard of care.

Study Data

Demographic and risk factor information are inputs to the genotypic and phenotypic

index. Final disease state (as determined by flexible cystoscopy and follow up) are collated with (ϕ) results and demographic information and subjected to statistical analysis.

Determination of the G+P Index

Using datasets obtained from samples collected from a large number of sites in New Zealand and Australia from approximately 500 patients, a training model to predict the probability of 'TCC = Yes' was developed.

Data Collected for Variables Used in the Training and Validation Populations

Phenotypic Variables

Clinical findings are: Gender, Age, Smoking history, and HFREQNEW (<=1 denoted as Low; >1 denoted as High).

Genotypic Variables

Genotypic Variables include expression of RNA markers: M1 (= MDK + CDC + IGBP5 – HOXA13) and IL8R. **Table 7** below shows estimates of the coefficients of each of the factors in the validation G+P INDEX:

Table 7

Analysis of Maximum Likelihood Estimates							
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Exp(Est)
Intercept		1	-4.8445	0.6532	55.0092	<.0001	0.008
gender	2	1	-1.8544	0.6760	7.5484	0.0060	0.157
smoke	2	1	-0.9049	0.4537	3.9775	0.0461	0.405
smoke	3	1	-1.3407	0.4936	7.3756	0.0066	0.262
HFREQNEW	Low	1	-0.7015	0.3478	4.0681	0.0437	0.496
M1		1	1.2221	0.1553	61.9328	<.0001	3.394
IL8R		1	-0.2408	0.1789	1.8112	0.1784	0.786

FIG. 18 depicts ROC curves for the G+P Index. **FIG. 18** is a graph of sensitivity (vertical axis) versus 1-specificity (horizontal axis) for results according to an embodiment of this invention. For comparison, a diagonal line depicts the model. The outcomes based solely upon Phenotypic information is shown as the dash-dotted line, the outcome based solely upon Genotypic information is shown as the dashed line, and the outcomes based on the G+P Index are shown as the solid line. These results indicate that the combination of Genotypic and Phenotypic information provides an unexpected, substantial improvement in prediction of outcome.

Exploratory models considered seven phenotypic variables, but AgeGT50 showed insignificant effect while there was insufficient data for RBC, so both of these variables were dropped from the final model. Based on the significance level of the remaining five phenotypic variables and the two RNA markers, an index was constructed. Using relationship between M1, IL8R and TCC (=yes) in the training dataset, a threshold of 4.5 and 2.5 was used for M1 and IL8R respectively. A score of 5, 4, 3, 2 and 1 was assigned to M1, Smokers, Male, IL8R, and HFREQ – which result in an index score ranging from 0 to 15. The integrated algorithm based on co-efficient is given below as the combined G+P index:

G+P INDEX = (1*HFREQ+3*Gender+4*SMK)+(5*M1+2*IL-8)

Odds ratios for different clinical factors that were retained in the final model are shown in **FIG. 19**. An odds ratio can be interpreted as having a harmful or protective effect upon the subject depending on how far it deviates from 1 (i.e., no effect). Odds ratios whose confidence limits exclude 1 are statistically significant. Generally, the factors with higher odd ratio (e.g. SMK, Gender) are assigned larger weights compared to factors with small odds ratio (e.g. HFREQ).

The classification table for the full model is presented below in **Table 8**.

Table 8

Classification Table									
	Correct		Incorrect		Percentages				
Prob Level	Event	Non-Event	Event	Non-Event	Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.000	66	0	415	0	13.7	100.0	0.0	86.3	.
0.050	61	245	170	5	63.6	92.4	59.0	73.6	2.0
0.100	54	319	96	12	77.5	81.8	76.9	64.0	3.6
0.150	47	358	57	19	84.2	71.2	86.3	54.8	5.0
0.200	45	376	39	21	87.5	68.2	90.6	46.4	5.3
0.250	40	381	34	26	87.5	60.6	91.8	45.9	6.4
0.300	37	391	24	29	89.0	56.1	94.2	39.3	6.9
0.350	36	398	17	30	90.2	54.5	95.9	32.1	7.0
0.400	35	402	13	31	90.9	53.0	96.9	27.1	7.2
0.450	35	403	12	31	91.1	53.0	97.1	25.5	7.1
0.500	31	407	8	35	91.1	47.0	98.1	20.5	7.9
0.550	29	408	7	37	90.9	43.9	98.3	19.4	8.3
0.600	29	408	7	37	90.9	43.9	98.3	19.4	8.3
0.650	28	409	6	38	90.9	42.4	98.6	17.6	8.5
0.700	24	409	6	42	90.0	36.4	98.6	20.0	9.3
0.750	20	409	6	46	89.2	30.3	98.6	23.1	10.1
0.800	12	410	5	54	87.7	18.2	98.8	29.4	11.6
0.850	11	411	4	55	87.7	16.7	99.0	26.7	11.8
0.900	6	412	3	60	86.9	9.1	99.3	33.3	12.7
0.950	1	413	2	65	86.1	1.5	99.5	66.7	13.6
1.000	0	415	0	66	86.3	0.0	100.0	.	13.7

Preliminary Validation Study of G+P Index

To further test the use of the G+P Index, we carried out another study. Based on the statistical significance of various clinical and RNA markers, an index was constructed. There were 98 subjects whose TCC status (yes or no) as well as G+P INDEX variables were available.

A score of 5, 4, 3, 2 and 1 was assigned to MI (genetic tests), Smokers, Male, IL8R, and HFREQ – which results in an index score ranging from 0 to 15. The number of true positives and true negatives were 6 and 84 respectively. Similarly, the number of false positives and false negatives were 5 and 3 respectively. Thus, the overall accuracy of the proposed index was 0.92.

Implications and Follow Up Based on the G+P Index

If the G+P INDEX result indicates a "High Risk" of UC defined as **a score of** 11-15 or above, the patient **is** prioritized for a flexible cystoscopy and abdominal ultrasound as clinically indicated.

If the G+P INDEX result indicates an "Moderate Risk" of UC, defined as a score of from 6-10, the patient is reviewed and followed up as per clinical practice. Consideration may be given to the use of cytology, uretoscopy and/or a CT scan.

If the G+P INDEX result indicates a "Low Risk" of UC, defined as a score of from 0-5, the patient will receive the normal standard of care and be placed on the appropriate waiting list.

Refernces

The following references relate to the disclosure above.

Agresti, A. and Coull, B. A. (1998) Approximate is better than "ex- act" for interval estimation of binomial proportions. The American Statistician, 52, 119-126.

Altekruse SF, Kosary CL, Krapcho M, Neyman N, Aminou R, Waldron W, Ruhl J, Howlader N, Tatalovich Z, Cho H, Mariotto A, Eisner MP, Lewis DR, Cronin K, Chen HS, Feuer EJ, Stinchcomb DG, Edwards BK (eds). *SEER Cancer Statistics Review, 1975-2007,* National Cancer Institute. Bethesda, MD, http:// seer.cancer.gov/csr/1975_2007/, based on November 2009 SEER data submission, posted to the SEER web site, 2010.

Brown, L. D. Cai, T. T. and DasGupta, A. (2001) Interval estimation for a binomial

proportion. Statistical Science, 16, 101-133.

Buckhaults et al., (Cancer Research 61:6996-7001 (2002) describes certain secreted and cell surface genes expressed in colorectal tumours.

Byrd, R. H., Lu, P., Nocedal, J. and Zhu, C. (1995) A limited memory algorithm for bound constrained optimization. *SIAM J. Scientific Computing*, 16, 1190–1208.

Dalgaard (2008). "Introductory Statistics with R, 2nd edition" ; Peter Dalgaard, Chapter 13 (2008), Springer, ISBN 978-0-387-79053-4.

DeLong, E. R., D. M. DeLong, and D. L. Clarke-Pearson. 1988. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. Biometrics 44: 837-845.

Gottschalk, P.G. and Dunn, J.R. (2005) The five-parameter logistic: A characterization and comparison with the four-parameter logistic. *Analytic Biochemistry*, 343, 54-65. Doi:10.1016/j.ab.2005.04.035

Grossfield G, Wolf J, Litwan M, Hricak H, Shuler C, Agerter D, et al. Asymptomatic microscopic hematuria in adults: summary of AUA best practice policy recommendations. AFP 2001:63:1145-54.

Hall et al (Laryngoscope 113(1):77-81 (2003) (PMID: 12679418) (Abstract) described predictive value of serum thyroglobulin in thyroid cancer.

Hoerl, A.E. (1962) Application of ridge analysis to regression problems. *Chemical Engineering Progress*, 58, 54-59.

Holyoake A, O'Sullivan P, Pollock R, Best T, Watanabe J, Kajita Y, et al. Development of a multiplex RNA urine test for the detection and stratification of transitional cell carcinoma of the bladder. Clin Cancer Res. 2008 Feb 1;14(3):742-9.

Hotte et al., (AJ. American Cancer Society 95(3):507-512 (2002) describes plasma osteopontin as a protein detectable in human body fluids and is associated with certain malignancies.

Kim et al., (JAMA 287(13):1671-1679 (2002) describes osteopontin as a potential diagnostic biomarker for ovarian cancer.

Koopman et al., (Cancer Epidemiol. Biomarkers Pre^{y} 13(3):487-491 (2004) (Abstract) describes osteopontin as a biomarker for pancreatic adenocarcinoma.

Kuo et al (Clin. Chim. Acta. 294(1-2):157-168 (2000) (Abstract) describes serum matrix metalloproteinase-2 and –9 in HCF- and HBV-infected patients.

Leman et al., (Urology, 69(4) 714-20 (2007) (Abstract) describes EPCA-2 as a serum marker for prostate cancer.

Loo R, Lieberman S, Slezak J, Landa H, Mariani A, Nicolaisen G, Aspera A and Jaconsen S: Stratifying risk of urinary tract malignant tumors in patients with asymptomatic microscopic hematuria. Mayo Clin Proc. 2013, 88(2); 129-138.

Lunn, D., Spiegelhalter, D., Thomas, A. and Best, N. (2009) The BUGS project: Evolution, critique and future directions (with discussion), *Statistics in Medicine* 28: 3049--3082.

Marchi et al., (Cancer 112, 1313-1324 (2008) (Abstract) describes ProApolipoprotein Al as a serum marker of brain metastases in lung cancer patients.

Martin et al., (Prostate Cancer Prostatic Dis. March 9, 2004 (PMID: 15007379) (Abstract) described use of human kallikrein 2, prostate-specific antigen (PSA) and free PSA as markers for detection of prostate cancer.

Mazzaferri et al., (J. Clin. Endocrinol. Metab. 88(4):1433-1441 (2003) (Abstract) describes thyroglobulin as a potential monitoring method for patients with thyroid carcinoma.

McDonald M, Swagerty D, Wetzel L: Assessment of Microscopic hematuria in adults. AFP 2006 73:10.

Melle et al., (Clin. Chem. 53(4), 629-635 (2007) (Abstract) describes HSP27 as a serum marker for pancreatic adenocarcinoma.

Moré, J.J. (1978) The Levenberg-Marquardt algorithm: implementation and theory, in *Lecture Notes in Mathematics* 630: Numerical Analysis, G.A. Watson (Ed.), Springer-Verlag: Berlin, pp. 105-116.

Nelder, J. A. and Mead, R. (1965) A simplex algorithm for function minimization. *Computer Journal* 7, 308–31.

O'Sullivan P, Sharples K, Dalphin M et al: A Multigene Urine Test for the Detection and Stratification of Bladder Cancer in Patients Presenting with Hematuria. J Urol 2012, Vol 188 No 3; 746

Pellegrini et al., (Cancer Immunol. Immunother. 49(7):388-394 (2000) (Abstract) describes measurement of soluble carcinoembryonic antigen and TIMP 1 as markers for pre-invasive colorectal cancer.

(R Development Core Team (2009). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0,

URL http://www.R-project.org/)

Richards, F.J. (1959) A flexible growth function for empirical use. *Journal of Experimental Botany*. 10, 290-300.

Rudland et al., (Cancer Rese arch 62: 3417-3427 (2002) describes osteopontin as a metastasis associated protein in human breast cancer.

Sing et al (2009). (Tobias Sing, Oliver Sander, Niko Beerenwinkel and Thomas Lengauer (2009). ROCR: Visualizing the performance of scoring classifiers.. R package version 1.0-4. http://CRAN.R-project.org/package=ROCR).

Speiss, A.-N., Feig, C. and Ritz, C. (2008) Highly accurate sigmoidal fitting of real-time RCR data by introducing a parameter for asymmetry. BMC Bioinformatics, 9, 221. Doi:10.1186/1471-2105-9-211.

Sprenger H, Lloyd AR, Lautens LL, Bonner TI, Kelvin DJ. Structure, genomic organization, and expression of the human interleukin-8 receptor B gene. J Biol Chem. 1994 Apr 15;269(15):11065-72.

StataCorp. 2009. *Stata Statistical Software: Release 11*. College Station, TX: StataCorp LP.

Schwartz G: Proper evaluation of asymptomatic microscopic hematuria in the era of evidence-based medicine-progress is being made. Mayo Clin Proc. 2013, 88(2); 123-125.

Trivedi C and Messing EM: Commentary: the role of cytologic analysis of voided urine in the work-up of asymptomatic microhematuria. *BMC Urology* 2009, **9**:13

Tsigkou et al., (I Clin Endocrinol Metab, 92(7) 2526-31 (2007) (Abstract) describes total inhibin as a potential serum marker for ovarian cancer.

Tsigkou et al., (I Clin Endocrinol Metab, 92(7) 2526-31 (2007) (Abstract) describes total inhibin as a potential serum marker for ovarian cancer.

Venables, W. N. & Ripley, B. D. (2002). Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0).

Whitley et al, (Dim Lab. Med. 24(1):29-47 (2004) (Abstract) describes thyroglobulin as a serum marker for thyroid carcinoma.

Yoshikawa et al., (Cancer Letters, 151: 81-86 (2000) describes tissue inhibitor of matrix metalloproteinase-1 in plasma of patients with gastric cancer.

Example 4: Triage of Patients Presenting with Hematuria Using G+P Index II

The G+P INDEX indicates a Positive when it takes values in the range of 11 to 15.

Definitions

As used herein, the following definitions are used in this and the following examples.

"AMH" means asymptomatic microhematuria;

"AUA" means American Urological Association;

"AUC" means area under the curve;

"CI" means confidence interval;

"CT" means computed tomography;

"ELISA" means enzyme-linked immunosorbent assay;

"FISH" means fluorescence *in situ* hybridization;

"HPF" means high-powered field;

"logOR" means log odds ratio;

"Hfreq" means average daily frequency of hematuria during the most recent hematuria episode;

"ISUP" means International Society of Urological Pathology;

"MRI" means magnetic resonance imaging;

"NPV" means negative predictive value;

"OR" means odds ratio;

"QC" means quality control;

"QoL" means quality of life;

"Phenotypic' is used to define clinical prognostic characteristics and to distinguish them from gene expression-based biomarkers that have been broadly defined as 'genotypic' variables.

"RBC" means red blood cell;

"ROC" means receiver operating curve;

"RT-qPCR" means quantitative reverse transcription polymerase chain reaction;

"STARD" means Standards for Reporting of Diagnostic Accuracy;

"UC" means urothelial carcinoma;

"WHO" means World Health Organization.

Introduction

Hematuria, which is most often associated with causes such as benign prostatic enlargement, infection or urinary calculi, but is also symptomatic of urothelial carcinoma (UC), is estimated to occur in between 1 and 22% of patients in a general population [1,2]. Macroscopic (macro-) hematuria is characterized by a visible colour change in the urine of patients, while microscopic (micro-) hematuria is defined more precisely as the presence of ≥3 red blood cells per high-powered field (RBCs/HPF) in three concurrently collected urine samples [2]. The overall prevalence of UC in patients with microhematuria has been reported to be approximately 4%, whereas several studies have consistently shown that the prevalence of UC is much higher in patients with macrohematuria, ranging from approximately 12–23% [2-6], yet up to four times as many patients with micro- versus macrohematuria present for urological evaluation [7]. Notably, given that recent changes to the American Urological Association (AUA) guidelines [2] have seen the threshold for asymptomatic microhematuria (AMH) lowered to ≥3 RBCs/HPF in a single sample, and even lower thresholds (≥1 RBC/HPF) have been proposed [8], a consequential increase in the number of patients with hematuria who will undergo a urological work-up to investigate potential UC and a corresponding increase in the overall clinical and financial burden of these patients on healthcare systems is expected.

Such hematuria-related referrals place a significant clinical burden on urologists, as all patients must undergo a full work-up to provide an often inconclusive diagnosis. Furthermore, the existing diagnostic tests – many of which are invasive or have high radiation loadings – can have a detrimental effect on patient quality of life (QoL), especially if the patient receives repeated cystoscopies as mandated in the current guidelines [2]. It has been reported that for cystoscopies performed without prophylactic antibiotics, 22% of patients had asymptomatic bacteriuria and 1.9% of patients developed a febrile urinary tract infection (UTI) within 30 days [9]. Other studies have also reported a high prevalence of macrohematuria, pain on voiding and transient erectile dysfunction in men following cystoscopy [10,11].

Healthcare systems also incur a significant financial burden as a result of patients with hematuria undergoing a full urological work up [12,13] and it has been concluded that urine cytology adds costs without offering any significant diagnostic benefit [14-16]. Consequently, integrating an accurate, non-invasive test into the primary clinical work-up of patients presenting with hematuria allows physicians to effectively triage patients with hematuria, thereby reducing the number of patients undergoing a full urological work-up and

investigative cystoscopy for UC, and offers significant benefits to both patients and healthcare systems [15-19].

Several clinical prognostic characteristics, including age, gender, smoking history and degree of hematuria, are well-established as risk factors for UC in patients with hematuria [3,20-22]. Recently, several groups have attempted to develop models based on clinical prognostic characteristics to predict the risk of UC in patients with hematuria [20-22], but critically, these models offer limited accuracy and have largely been focused on detecting patients with UC rather than ruling out patients who do not have disease. These detection-focused models have therefore been insufficient to reliably identify patients with disease during a primary evaluation, even if used in combination with urine cytology [20-22].

Despite the higher incidence of UC in patients presenting with macrohematuria, a number of studies show there is no significant difference in the distribution of UC by grade and stage in patients presenting with micro- compared with those presenting with macrohematuria [5,23-25]. Therefore, the AUA recommends that all patients with macrohematuria or AMH be referred to a urologist for a full urological work-up, as severity of hematuria is not sufficiently predictive for the presence of UC [2]. However, as patients with hematuria may only undergo limited urinalysis in a primary evaluation, consisting of cytology and in some cases imaging studies, such as ultrasound, a full urological work-up is often necessary to conclusively detect or rule out UC. While urine cytology is specified in current guidelines and routinely used in patients with suspected UC, cytology results are often inconclusive with atypical or suspicious findings and also suffer from a low diagnostic yield driven by a relatively high risk of false negative results for patients with UC-related hematuria [2,26,27]. Consequently, it can be difficult to rule out benign causes of hematuria, whether macrohematuria or AMH, during a primary evaluation, especially if UC-related hematuria is intermittent and appears to resolve following treatment for a benign cause [12].

A number of gene-based studies have set out to profile urinary biomarkers in patients with UC, and these biomarkers may be useful in their own right for detecting disease [28,29]. An opportunity also exists to triage out patients on the basis of their clinical characteristics and gene expression profile. Combining NMP22 enzyme-linked immunosorbent assay (ELISA) tests or a panel of gene markers with clinical characteristics has been shown to improve diagnostic accuracy compared with clinical characteristics alone, but these combined models have not yet delivered significant advances in overall diagnostic accuracy, especially

when attempting to identify low-risk patients [30,31]. Nevertheless it is considered that incorporating clinical factors and specific gene expression into a combined algorithm is likely to provide the best guidance for diagnosing and managing patients with hematuria or UC [32].

Cxbladder™ Detect (Pacific Edge Ltd., Dunedin, New Zealand), a multigene test performed on unfractionated urine has previously been shown to be more sensitive than urine cytology and NMP22 for detecting UC in patients with macrohematuria [33] and more accurate than urine cytology, NMP22 and fluorescence *in situ* hybridization (FISH) in a comparative analysis (Kasabov, Darling, Breen, et al., unpublished observations). Cxbladder Detect uses quantitative reverse transcription polymerase chain reaction (RT-qPCR) technology to quantify five mRNA markers, four markers that are overexpressed in UC alongside a fifth marker that is elevated in non-malignant inflammatory conditions, and offers a high level of specificity and sensitivity when used to detect UC in patients presenting with hematuria [33]. It was hypothesized that an integrated model combining high-performance genetic biomarkers with phenotypic variables collected from the same patients will provide superior clinical resolution using high sensitivity (i.e. a low probability of a patient with UC receiving a false negative result), high negative predictive value (i.e. a high proportion of all negative results being true) and a high test-negative rate to enable the accurate triage of patients who have a low probability of UC. These genotypic and phenotypic variables when combined into a novel segregation model enable patients with hematuria who have a low probability of UC to be identified and triaged, as opposed to undergoing a full urological work-up.

Methods

Patient selection

A prospective sample of 695 patients have been analysed, where true clinical outcome was determined using a conventional clinical evaluation. The study sample consists of an initial cohort of patients with hematuria was consented and sampled as previously described [33], where a consecutive series of 517 patients with a recent history of macrohematuria, aged ≥45 years and without a prior history of UC, were recruited prospectively from nine urology clinics in Australia and New Zealand. These patients were followed for three months for determination of UC status or alternative diagnosis following multigene analysis of urine samples, with a positive UC diagnosis being based on cystoscopical appearance and

histopathological examination. The stage of disease was classified according to the TNM staging criteria determined by pathology and diagnostic imaging investigations and tumor grade was classified according to local pathology practice, using the 1998 World Health Organization (WHO)/International Society of Urological Pathology (ISUP) consensus classification [34].

Additional cohorts of 94 and 84 patients undergoing urological investigations following a macrohematuria event were subsequently recruited from two centers in New Zealand between March 2012 and April 2013 and included in the development of models. Centers were selected to participate on the basis of their previous experience participating in the initial study and their willingness to evaluate the Cxbladder Detect product within individual clinical settings.

An additional representative test set of 45 patients presenting with microhematuria were prospectively collected and used for further validation of the G+P INDEX, as set out below.

Eligibility criteria were similar to those of [33], except that patients aged ≥18 years and those who had previously undergone a cystoscopy to investigate UC that proved to be negative were eligible for enrolment. Furthermore, as in [33], patients exhibiting symptoms indicative of a UTI, or bladder or renal calculi, were excluded.

Ethical approval for this study was granted by all participating centers and informed consent obtained from all patients providing samples.

Urine sample collection and assessment

To provide gene expression data, a single mid-stream urine sample was collected from participants using the Urine Sampling System from Pacific Edge. Multigene analysis of samples from all studies was carried out in accordance with the standard operating procedure, as is used for the commercially available Cxbladder Detect multigenic test. All urine samples (4.5 mL) from the initial cohort were collected at a clinic prior to cystoscopy and transferred to a stabilization liquid via vacuum driven aspiration and sent to Pacific Edge within 48 hours. The samples were then stored at -80°C until required for batch analysis. Samples from the subsequent cohorts were collected in the same manner, but shipped to Pacific Edge at ambient temperature and processed within 7 days of sample collection in accordance with revised quality control (QC) limits and tolerance testing performed at the Pacific Edge diagnostic

laboratory.

Statistical analysis

Univariate logistic regression was used to estimate the unadjusted (raw) log odds ratio (logOR) co-efficients for four binary phenotypic variables associated with UC: age, gender, smoking history and average daily frequency of hematuria during the patient's most recent hematuria episode (Hfreq; see **Table 9**).

Table 9: Definitions of binary phenotypic variables associated with UC and their corresponding scores

Phenotypic parameter	**Score**	
	0	**1**
Gender	Female	Male
Age	<60 years	≥60 years
Smoking history	Never smoked	Current or past smoker
Hfreq	≤1 episode/day	>1 episode/day

Multivariate logistic regression on all four phenotypic variables was used to generate adjusted logOR co-efficients in the phenotypic model (P INDEX).

G INDEX was developed using logistic regression to determine the association between UC and mRNA concentrations for the five Cxbladder® Detect genes (*IGFBP5*, *HOXA13*, *MDK*, *CDK1* and *CXCR2*) in urine samples. A multivariate genotypic-phenotypic model (G+P INDEX) was generated using a combination of all nine variables from the G INDEX and P INDEX. These linear models determined the logOR from which the probability of a patient having UC was derived.

The relative performance of each of model was illustrated in receiver operating curves (ROCs) plotting the false positive rate versus the true positive rate when testing for UC, as determined by each model. Area under the curve (AUC) was used to compare the relative efficiency of each model with an AUC approaching 1 deemed to be optimal.

To reduce potential bias when model estimation and prediction are performed on the same data set, a bias-corrected AUC was calculated for each of the three logistic regression models using bootstrap resampling [35]. The difference between the nominal AUC from the original sample and the average AUC from the bootstrap samples is an estimate of the sample

bias and the nominal AUCs were adjusted accordingly. Bootstrap estimates of bias-corrected confidence intervals (CIs) were also obtained [36].

Furthermore, it was a design criteria for this clinical test that the performance characteristics of each model must exceed a threshold NPV of 0.97, with as high a sensitivity as possible with the further caveat of having a high test-negative rate. The test negative rate is selected to provide a high clinical resolution when triaging out patients presenting with hematuria who have a low probability of having UC. Comparisons were made between the G INDEX, P INDEX and G+P INDEX and the performance of each model was determined in terms of sensitivity and NPV with a sufficiently high test-negative rate to provide an effective tool for triaging out patients with haematuria who have a low probability of UC.

Results

Sample demographics

Of the 695 patients with macrohematuria registered across the three cohorts, 23 were deemed to be ineligible and samples from a further 85 patients were excluded after enrolment due to the absence of sufficient data or samples failing to meet QC standards (see **FIG. 20A**). In total, samples from 587 patients were available for modelling comprising 72 UC-positive and 515 UC-negative samples.

Of the 45 samples from patients with microhematuria provided, 40 were suitable for analysis with 5 patients deemed ineligible and excluded from the analysis (see **FIG. 20B**). All 45 patients had received a full urological evaluation and clinical truth was confirmed as UC-negative. Full demographic data from both sample populations is presented in **Table 10**.

Table 10: Sample population demographics for patients with macro- and microhematuria with complete data

Parameter		**Patients with macrohematuria (N=587), n (%)**	**Patients with microhematuria (N=40), n (%)**
Age, years	0–49	65 (11.1)	21 (52.5)
	50–59	111 (18.9)	

	60–69	145 (24.7)	19 (47.5)
	70–79	175 (29.8)	
	80–100	91 (15.5)	
Gender	Female	113 (19.3)	25 (62.5)
	Male	474 (80.7)	15 (37.5)
Smoking history	Never smoked	246 (41.9)	25 (62.5)
	Current or past smoker	341 (58.1)	15 (37.5)
Hfreq (episodes/day)	≤1	332 (56.6)	40 (100)
	>1	255 (43.4)	–
Tumor stage	Normal	515 (87.7)	40 (100)
	T1	16 (2.7)	–
	T2	11 (1.9)	–
	T3	2 (0.3)	–
	Ta	40 (6.8)	–
	Tis	3 (0.5)	–

Relationship between phenotypic variables and risk of UC in patients with macrohematuria

Unadjusted univariate logistic regression analyses of each of the four binary phenotypic variables indicated that age ≥60 years, male gender, a history of smoking and a high frequency of macrohematuria were all associated with an increased risk of UC (**Table 11**).

Table 11: Unadjusted and adjusted ORs for UC by phenotypic and genotypic factors for patients with hematuria

			Unadjusted OR (95% CI)	**Adjusted P variable OR (95% CI)**	**Adjusted G+P variable OR (95% CI)**
Phenotypic variables	Control	UC			

Age, years	<60	151	11	2.30 (1.22–4.73)	2.24 (1.18–4.65)	1.89 (0.85–4.64)
	≥60	364	61			
Gender	Female	105	8	2.05 (1.01–4.75)	1.58 (0.76–3.72)	3.03 (1.12–9.36)
	Male	410	64			
Smoking history	Never smoked	227	19	2.20 (1.29–3.91)	2.19 (1.27–3.92)	2.67 (1.34–5.64)
	Current or past smoker	288	53			
Hfreq (average episodes/day)	≤1	300	32	1.74 (1.06–2.88)	1.80 (1.08–3.00)	1.76 (0.93–3.35)
	>1	215	40			
Genotypic variables				**Unadjusted OR (95% CI)**	**Adjusted G variable OR (95% CI)**	**Adjusted G+P variable OR (95% CI)**
IGFBP5				7.34 (4.59–12.33)	2.15 (1.03–4.58)	2.21 (1.03–4.83)
HOXA13				6.27 (3.92–10.34)	0.33 (0.13–0.83)	0.20 (0.07–0.56)
MDK				7.10 (4.73–11.10)	4.76 (1.74–13.62)	8.14 (2.64–26.60)
CDK1				7.80 (5.11–12.39)	3.47 (1.39–9.13)	2.59 (0.98–7.18)
CXCR2				1.69 (1.36–2.10)	0.65 (0.45–0.92)	0.69 (0.47–0.98)

Adjusted P INDEX, G INDEX and G+P INDEX variable ORs are the exponentiated co-efficients in the P INDEX, G INDEX and G+P INDEX, respectively.

Adjusted logOR co-efficients were calculated in the multivariate logistic regression model.

P INDEX= -3.78 + 0.81 × Age + 0.46 × Gender + 0.78 × Smoking history + 0.59 × Hfreq

where each phenotypic variable is assigned a binary score of 0 or 1, as designated in **Table 9**, and the confidence intervals for the co-efficients are presented in **Table 11**. The bias-corrected estimate for AUC for the P INDEX is 0.66 (95% CI: 0.55–0.67; **FIG. 21**).

Relationship between genotypic variables and risk of UC in patients with macrohematuria

The G INDEX was estimated by logistic regression using the log mRNA concentrations of the five genes *IGFBP5*, *HOXA13*, *MDK*, *CDK1* and *CXCR2* in urine samples to predict UC occurrence.

G INDEX = -6.22 + 0.77 × *IGFBP5* – 1.11 × *HOXA13* +1.56 × *MDK* + 1.24 × *CDK1* – 0.43 × *CXCR2*

The G INDEX gives a bias-corrected AUC of 0.83 (95% CI: 0.74–0.89; **FIG. 21**).

Relationship between genotypic and phenotypic variables and risk of UC in patients with macrohematuria

The five continuous genotypic variables were then combined with the four binary phenotypic variables to estimate the G+P INDEX using mulitvariate logistic regression.

G+P INDEX = -8.46 + (0.79 × *IGF* – 1.60 × *HOXA* + 2.10 × *MDK* + 0.95 × *CDC* – 0.38 × *IL8R*) +
(0.64 × Age + 1.11 × Gender + 0.98 × Smoking history + 0.56 × Hfreq)

The G+P INDEX gives a bias-corrected AUC of 0.86 (95% CI: 0.80–0.91).

Comparison between G INDEX and G+P INDEX

There is overlap between the confidence intervals for the G INDEX and G+P INDEX, so a bootstrap version of a paired test was constructed by determining the difference in AUC for the G INDEX and G+P INDEX for each bootstrap sample. Ten thousand bootstrap samples with a sample size of n=587 were generated by random sampling with replacement from the original 587 samples available for analysis. The resulting 95% CI for the difference between models was 0.01–0.08. Thus the probability that the true difference between the two AUCs is less than 0.01 is <0.025, indicating that there is a high likelihood of the AUC for the

G+P INDEX being significantly greater than the AUC for the G INDEX.

NPV and sensitivity of models

The G+P INDEX generated an NPV >0.97 over the range of test-negative rates from 0.2 to 0.7 and was almost always higher than the NPV for the G INDEX model (**FIG. 22**). The G+P INDEX offered performance characteristics of sensitivity of 0.95 and NPV 0.98 when the test-negative rate was 0.4 (**Table 12**; **FIG. 22**). In contrast, the G INDEX only achieved sensitivity of 0.86 and an NPV of 0.96 when the test-negative rate was 0.4 (**Table 12**).

Table 12: Performance Characteristics of Each Model When Thresholds Are Set For Varying Test Negative Rates

Threshold (logOR)	Test-negative rate (95% CI)	NPV (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
P INDEX				
-2.54	0.25 (0.21–0.28)	0.97 (0.92–0.99)	0.93 (0.85–0.98)	0.27 (0.23–0.31)
-2.52	0.38 (0.34–0.42)	0.95 (0.91–0.97)	0.83 (0.74–0.91)	0.41 (0.37–0.45)
-2.39	0.42 (0.37–0.45)	0.95 (0.91–0.97)	0.82 (0.72–0.90)	0.45 (0.40–0.49)
-1.95	0.51 (0.47–0.54)	0.92 (0.89–0.95)	0.68 (0.56–0.78)	0.53 (0.49–0.57)
-1.93	0.51 (0.46–0.55)	0.92 (0.89–0.95)	0.68 (0.56–0.78)	0.53 (0.49–0.58)
-1.73	0.82 (0.79–0.85)	0.90 (0.87–0.92)	0.32 (0.22–0.43)	0.84 (0.81–0.87)
G INDEX				
-3.46	0.20 (0.17–0.23)	0.94 (0.88–0.97)	0.90 (0.80–0.95)	0.22 (0.18–0.25)
-3.23	0.30 (0.26–0.34)	0.95 (0.91–0.98)	0.89 (0.80–0.95)	0.33 (0.28–0.37)
-3.04	0.40 (0.36–0.44)	0.96 (0.92–0.98)	0.86 (0.77–0.93)	0.44 (0.40–0.48)
-2.86	0.50 (0.46–0.54)	0.97 (0.94–0.98)	0.86 (0.77–0.93)	0.55 (0.51–0.59)
-2.63	0.60 (0.56–0.63)	0.96 (0.94–0.98)	0.82 (0.71–0.90)	0.66 (0.62–0.69)
-2.41	0.70 (0.66–0.73)	0.96 (0.94–0.98)	0.78 (0.65–0.86)	0.77 (0.73–0.80)
G+P INDEX				
-4.02	0.20 (0.17–0.23)	0.97 (0.93–0.99)	0.96 (0.88–0.99)	0.22 (0.19–0.26)
-3.67	0.30 (0.26–0.33)	0.98 (0.94–0.99)	0.94 (0.87–0.99)	0.33 (0.29–0.37)
-3.33	0.40 (0.36–0.44)	0.98 (0.95–1.00)	0.95 (0.86–0.98)	0.45 (0.40–0.49)
-2.99	0.50 (0.46–0.54)	0.98 (0.96–0.99)	0.92 (0.83–0.97)	0.56 (0.52–0.60)
-2.71	0.60 (0.56–0.64)	0.97 (0.95–0.99)	0.86 (0.76–0.93)	0.67 (0.63–0.71)
-2.37	0.70 (0.66–0.73)	0.97 (0.94–0.98)	0.80 (0.70–0.88)	0.77 (0.73–0.80)

Application of the G+P INDEX in patients with microhematuria

While the G+P INDEX was developed using data from patients with macrohematuria, its robustness was tested in a further 40 samples from patients with microhematuria (Hfreq=0). A higher test-negative rate was expected in a microhaematuria population as the incidence of UC is lower in this population, and using a test negative rate of 0.4, 32 (80%) patients tested negative and would be correctly triaged out, therefore not requiring a full urological work-up for the determination of UC.

Discussion

This study defines a clinical tool that offers clinicians and physicians the ability to effectively triage-out patients presenting with hematuria from the need to have a full urological work-up for the detection of UC. The study presents an internally validated genotypic-phenotypic model, G+P INDEX, with bootstrap-based CI estimates, that offers a combination of high sensitivity and high NPV (i.e. a low probability of an individual patient with UC providing a false-negative result and a high proportion of all negative results being true) that is not offered by models derived exclusively from genotypic or phenotypic data alone. This provides clinicians and physicians with a unique opportunity to triage out patients with both micro- and macrohematuria, in particular by identifying patients with a low risk of having UC who do not require a full urological work up.

A high test-negative rate in the context of high sensitivity is an important consideration for an effective triage-out test that aims to direct patients with a low probability of UC away from a full clinical work-up [37]. Accordingly, at a test-negative rate of 0.4 the sensitivity of the G+P INDEX presented here maximizes both the sensitivity and NPV (0.95 and 0.98, respectively). This can be compared with the best fit selected from the genotypic model published in [33] (sensitivity=0.82; NPV=0.97) and is also comparable with the sensitivity and NPV of both cystoscopy (sensitivity=0.89–0.98; NPV=0.99) and virtual cystoscopy using computed tomography (CT) scans or magnetic resonance imaging (MRI) (sensitivity=0.94 and 0.91, respectively) [38-40].

It is acknowledged that the sample population used to derive the G INDEX, P INDEX and G+P INDEX in this instance consisted of patients with macrohematuria. However, the high sensitivity of this test at a test negative rate of 0.4 in patients with macrohematuria allows the G+P INDEX to be applied across both macro- and microhematuria populations.

Presuming that patients with and without UC are similarly distributed amongst the micro- and macrohematuria patient populations, but with an expected UC prevalence of 4% in the microhematuria population, a high NPV can also be expected in the microhematuria patient population.

By applying the G+P INDEX to the sample population of patients with microhematuria who do not have UC it was shown that 80% of the patients would have been triaged out on the basis of the result. Only 20% would be referred for a full urological work-up. This compares with conventional guidelines that would currently see all of the patients (100%) with microhematuria that cannot be attributed to a benign cause undergoing a full urological work-up, incurring significant unnecessary costs and negatively impacting patient QoL.

Severity of hematuria is correlated with the probability of a patient having UC, but not the stage or grade of any tumour, and an estimated 96% and 77–88% of patients with micro- and macrohematuria, respectively, referred to a urologist will not have UC [2-6]. Therefore, avoiding potentially unnecessary urological work-ups for patients with hematuria has several benefits. Cystoscopy may be associated with adverse effects, such as pain on voiding, bleeding, UTIs, male sexual dysfunction and the anxiety that accompanies an inconclusive or unconfirmed UC diagnosis [9-11]. Most notably, this novel approach has the potential to reduce the burden on resources and the financial cost associated with a full urological work-up on UC-negative patients. For example, in the UK, avoiding cystoscopy in patients with hematuria with an initial negative cytology and/or tumor biomarker test has been estimated to save approximately US$770 per patient (£483 per patient) evaluated [13]. The G+P INDEX described here provides an effective alternative to the use of urine cytology when used in a primary evaluation setting. This is particularly relevant in settings where primary evaluation is carried out by primary care physicians.

On this basis, if we assign an arbitrary 'nominal cost' of US$4,500 for each full urological work up, the total cost for working up 1,000 patients with microhematuria would approach US$4.5 million. In contrast, if 80% of patients with microhematuria are triaged out using the G+P INDEX at an arbitrary nominal cost of US$2,500, the total direct cost of testing and full urological work-ups for the remaining 20% of patients would total US$3.4 million. This provides a notional net saving in direct costs of approximately US$1.1 million per 1,000 patients with microhematuria.

While the genotypic algorithm developed by O'Sullivan et al. [33] comprised the same genotypic constituents as the G+P INDEX presented here, the balance between sensitivity and specificity was calibrated for the optimal primary detection of UC in symptomatic patients (i.e. presenting with hematuria) who were undergoing a full urological work-up. In contrast, the G+P INDEX in this study also incorporated phenotypic variables and has been optimized for high sensitivity and high NPV, to segregate out those patients with hematuria who do not require a full urological work-up for suspected UC. No attempt is made to define or select patients with UC. Instead the aim is to confidently rule out those who do not have UC, and as such, all patients not segregated out would progress for a full urological work-up.

While several studies have previously sought to develop predictive models that consider phenotype when assessing the risk of UC in patients presenting with hematuria, the accuracy of phenotype-dependent models alone appears to be limited. For example, Loo et al. [21] prospectively investigated whether phenotypic parameters could be used to identify patients with microhematuria who may not have required a urological referral and full work-up and concluded that age, male gender and a recent diagnosis of macrohematuria were significant predictors of UC. A history of smoking and >25 RBCs/HPF in a recent urinalysis were not statistically significant predictors of UC, in isolation, but even when included in their 'Hematuria Risk Index' to improve predictive accuracy, this index resulted in an AUC of 0.809 [21]. Interestingly, the phenotypic ORs in this study and those identified by Loo et al. are comparable, with overlapping 95% CIs for smoking history and gender, and while age, gender and smoking history have similar weightings in each model, the influence of the genotypic component of the G+P INDEX presented here is likely to account for the higher AUC [21].

Likewise, Cha et al. [20] reported that age, smoking history and degree of hematuria, but not gender, were significantly correlated with the presence of UC in patients with asymptomatic hematuria and used a multivariate model to develop a nomogram comprised of phenotypic and urine cytology data for predicting UC. As with Loo et al. [21], the reported phenotypic ORs are comparable to those reported here, but even after incorporating urine cytology into the nomogram, the AUC of 0.831 reported in [20] was lower than that of the G+P INDEX.

In another study, Tan et al. [22] retrospectively stratified patients with hematuria who had been referred to a specialist urology clinic into high- and low-risk groups using a

nomogram derived from patient age, gender, smoking history and the degree of hematuria. While comparisons with this study must be made with caution given the high proportion of patients who were excluded due to an absence of data (80 out of 405 patients), the AUC of 0.804, sensitivity of 0.900 and NPV of 0.953 were all lower than the G+P INDEX described here.

Several attempts have also been made to improve the accuracy of phenotypic models by supplementing them with the results of urinary biomarker tests. When the nuclear matrix protein NMP22 point of care proteomic assay is used in isolation to detect UC it has a sensitivity of 0.557 and NPV of 0.968 [17]. Lotan et al. [41] published a multivariable algorithm comprising phenotypic factors, NMP22 and urine cytology with an AUC for predicting UC of 0.826 that was then prospectively validated with an AUC of 0.802 [31]. However, it is important to note that this model attempted to discriminate between high-risk patients who did and did not have UC, as opposed to maximizing sensitivity and NPV to triage-out patients with a low probability of UC.

The improved accuracy obtained with algorithms comprising both genotypic and phenotypic data have previously been demonstrated in breast cancer, in particular [42-45]. Likewise, Mitra et al. [30] used a combination of molecular markers and smoking intensity to calculate a multivariate model that was superior to routine clinicopathological parameters in predicting survival in patients with UC. However, the present study is the first to demonstrate that phenotypic risk factors can be combined with genotypic data to increase the accuracy of a model for separating patients with haematuria into categories requiring differential levels of urological follow up and clinical care rather than survivorship prediction.

When phenotypic data are combined with genotypic data in a model, the resolution of data is likely to impact the accuracy of the model. For example, smoking is a well understood risk factor for UC and is included in most phenotypic models for detecting UC. In Cha et al. [20], Tan et al. [22], Lotan et al. [31,41] and the current study, the binary discriminants never smoked and current/ex-smoker were used, whereas Mitra et al. [30] calculated smoking intensity on the basis of years of smoking and number of cigarettes smoked each day and Loo et al. [21] categorized smokers into never smoked, passive smokers, smokers who had ceased and current smokers. While it is known that the risk of UC increases substantially with exposure to smoking [46], arbitrarily defining phenotypic variables may limit the overall accuracy and utility of phenotypic models. In contrast, an interaction between a patient's

genotypic and phenotypic variables would not be unexpected. However, combining the impact of phenotypic factors and genetic variables in a single tool improved the accuracy of the model described in this study. A similar principle also applies to describing hematuria phenotype. Patients presenting with micro- or macrohematuria are essentially on a biological continuum and have different likelihoods of having UC [2-6,21]. Accordingly, despite all patients with microhematuria in this study having a Hfreq score of 0, the severity of their hematuria, in combination with other phenotypic factors, is likely to be indirectly accounted for in the genotypic component of the G+P INDEX.

References

The articles recited immediately below refer to Example 4, and are all incorporated herein fully by reference.

1. Kelly JD, Fawcett DP, Goldberg LC: Assessment and management of non-visible haematuria in primary care. *BMJ* 2009, 338:a3021.
2. Davis R, Jones JS, Barocas DA, Castle EP, Lang EK, Leveillee RJ, Messing EM, Miller SD, Peterson AC, Turk TMT, Weitzel W: Diagnosis, evaluation and follow-up of asymptomatic microhematuria (AMH) in adults: AUA guideline. *J Urol* 2012, 188(6 Suppl):2473-2481.
3. Sutton JM. Evaluation of hematuria in adults. *JAMA* 1990, 263:2475-2480.
4. Khadra MH, Pickard RS, Charlton M, Powell PH, Neal DE: A prospective analysis of 1,930 patients with hematuria to evaluate clinical practice. *J Urol* 2000, 163:524-527.
5. Davidson P: Re-design of a haematuria clinic: Assessment of 2346 haematuria patients. *J Urol* 2011, 185(4S):e495.
6. Price SJ, Shephard EA, Stapley SA, Barraclough K, Hamilton WT: Non-visible versus visible haematuria and bladder cancer risk: A study of electronic records in primary care. *Br J Gen Pract* 2014, 64:e584-e589.
7. Buteau A, Seideman CA, Svatek RS, Youssef RF, Chakrabati G, Reed G, Bhat D, Lotan Y: What is evaluation of hematuria by primary care physicians? Use of electronic medical records to assess practice patterns with intermediate follow-up. *Urol Oncol* 2014, 32:128-134.
8. Jimbo M: Evaluation and management of hematuria. *Prim Care* 2010, 373:461-472.

9. Herr HW: The risk of urinary tract infection after flexible cystoscopy in bladder tumor patients who did not receive prophylactic antibiotics. *J Urol* 2014 Jul 18. pii: S0022-5347(14)03963-9.

10. Burke DM, Shackley DC, O'Reilly PH: The community-based morbidity of flexible cystoscopy. *BJU Int* 2002, 89:347-349.

11. Stav K, Leibovici D, Goren E, Livshitz A, Siegel YI, Lindner A, Zisman A: Adverse effects of cystoscopy and its impact on patients' quality of life and sexual performance. *Isr Med Assoc J* 2004, 6:474-478.

12. Rao PK, Jones JS: How to evaluate 'dipstick hematuria': What to do before you refer. *Cleve Clin J Med* 2008, 75:227-233.

13. Rodgers M, Nixon J, Hempel S, Aho T, Kelly J, Neal D, Duffy S, Ritchie G, Kleijnen J, Westwood M: Diagnostic tests and algorithms used in the investigation of haematuria: systematic reviews and economic evaluation. *Health Technol Assess* 2006, 10:iii-iv, xi-259.

14. Falebita OA, Lee G, Sweeney P: Urine cytology in the evaluation of urological malignancy revisited: is it still necessary? *Urol Int* 2010, 84:45-49.

15. Feifer AH, Steinberg J, Tanguay S, Aprikian AG, Brimo F, Kassouf W: Utility of urine cytology in the workup of asymptomatic microscopic hematuria in low-risk patients. *Urology* 2010, 75:1278-1282.

16. Svatek RS, Hollenbeck BK, Holmäng S, Lee R, Kim S, Stenzl A, Lotan Y. The economics of bladder cancer: Costs and considerations of caring for this disease. *Eur Urol* 2014, 66:253-262.

17. Grossman HB, Messing E, Soloway M, Tomera K, Katz G, Berger Y, Shen Y: Detection of bladder cancer using a point-of-care proteomic assay. *JAMA* 2005, 293:810-816.

18. Friedlander DF, Resnick MJ, You C, Bassett J , Yarlagadda V , Penson DF , Barocas DA: Variation in the intensity of hematuria evaluation: a target for primary care quality improvement. *Am J Med* 2014, 127:633-640.

19. Shinagare AB, Silverman SG, Gershanik EF, Chang SL, Khorasani R: Evaluating hematuria: impact of guideline adherence on urologic cancer diagnosis. *Am J Med* 2014, 127:625-632.

20. Cha EK, Tirsar LA, Schwentner C, Hennenlotter J, Christos PJ, Stenzl A, Mian C, Martini T, Pycha A, Shariat SF, Schmitz-Dräger BJ: Accurate risk assessment of patients with asymptomatic hematuria for the presence of bladder cancer. *World J Urol* 2012, 30:847-852.

21. Loo RK, Lieberman SF, Slezak JM, Landa LM, Mariani AJ, Nicolaisen G, Aspera AM, Jacobsen SJ: Stratifying risk of urinary tract malignant tumors in patients with asymptomatic microscopic hematuria. *Mayo Clin Proc* 2013, 88:129-138.

22. Tan GH, Shah SA, Ann HS, Hemdan SN, Shen LC, Galib NAFA, Singam P, Kong CHC, Hong GE, Bahadzor B, Zainuddin ZM: Stratifying patients with haematuria into high or low risk groups for bladder cancer: a novel clinical scoring system. *Asian Pac J Cancer Prev* 2013, 14:6327-6330.

23. Sultana S, Goodman C, Bryne D, Baxby K: Microscopic haematuria: urological investigation using a standard protocol. *Br J Urol* 1996, 78:691-698.

24. Sugimura K, Ikemoto S-I, Kawashima H, Nishisaka N, Kishimoto T: Microscopic hematuria as a screening marker for urinary tract malignancies. *Int J Urol* 2001, 8:1-5.

25. Viswanath S, Zelhof B, Ho E, Sethia K, Mills R: Is routine urine cytology useful in the haematuria clinic? *Ann R Coll Surg Engl* 2008, 90:153-155.

26. Steiner H, Bergmeister M, Verdorfer I, Granig T, Mikuz G, Bartsch G, Stoehr B, Brunner A: Early results of bladder-cancer screening in a high-risk population of heavy smokers. *BJU Int* 2008, 102:291-296.

27. Yeung C, Dinh T, Lee J: The health economics of bladder cancer: An updated review of the published literature. *Pharmacoeconomics* 2014, 32:1093-1104.

28. Holyoake A, O'Sullivan P, Pollock R, Best T, Watanabe J, Kajita Y, Matsui Y, Ito M, Nishiyama H, Kerr N, da Silva Tatley F, Cambridge L, Toro T, Ogawa O, Guilford P: Development of a multiplex RNA urine test for the detection and stratification of transitional cell carcinoma of the bladder. *Clin Cancer Res* 2008, 14:742-749.

29. Sapre N, Anderson PD, Costello AJ, Hovens CM, Corcoran NM: Gene-based urinary biomarkers for bladder cancer: an unfulfilled promise? *Urol Oncol* 2014, 32:48.e9-e17.

30. Mitra AP, Castelao JE, Hawes D, Tsao-Wei DD, Jiang X, Shi SR, Datar RH, Skinner EC, Stein JP, Groshen S, Yu MC, Ross RK, Skinner DG, Cortessis VK, Cote RJ:

Combination of molecular alterations and smoking intensity predicts bladder cancer outcome. *Cancer* 2013, 119:756-765.

31. Lotan Y, Svatek RS, Krabbe LM, Xylinas E, Klatte T, Shariat SF: Prospective external validation of model for bladder cancer detection. *J Urol* 2014, 192:1343-1348.
32. Abogunrin F, O'Kane HF, Ruddock MW, Stevenson M, Reid CN, O'Sullivan JM, Anderson NH, O'Rourke D, Duggan B, Lamont JV, Boyd RE, Hamilton P, Nambirajan T, Williamson KE: The impact of biomarkers in multi- variate algorithms for bladder cancer diagnosis in patients with hematuria. *Cancer* 2012, 118:2641-2650.
33. O'Sullivan P, Sharples K, Dalphin M, Davidson P, Gilling P, Cambridge L, Harvey J, Toro T, Giles N, Luxmanan C, Felipe Alves C, Yoon HS, Hinder V, Masters J, Kennedy-Smith A, Beaven T, Guilford P: A multigene urine test for the detection and stratification of bladder cancer in patients presenting with hematuria. *J Urol* 2012, 188:741-747.
34. Epstein JI, Amin MB, Reuter VR, Mostofi FK: The World Health Organization/International Society of Urological Pathology consensus classification of urothelial (transitional cell) neoplasms of the urinary bladder. Bladder Consensus Conference Committee. *Am J Surg Pathol* 1998, 22:1435-1448.
35. Efron B, Tibshirani RJ: *An introduction to the bootstrap*. New York: Chapman Hill; 1993.
36. DiCiccio TJ, Tibshirani RJ: Bootstrap confidence intervals. *Statist Sci* 1996, 11:189-228.
37. Van't Hoog AH, Cobelens F, Vassall A, van Kampen S, Dorman SE, Alland D, Ellner J: Optimal triage test characteristics to improve the cost-effectiveness of the Xpert MTB/RIF assay for TB diagnosis: A decision analysis. *PLoS ONE* 2013, 8:e82786.
38. Qu X, Huang X, Wu L, Huang G, Ping X, Yan W: Comparison of virtual cystoscopy and ultrasonography for bladder cancer detection: A meta-analysis. *Eur J Radiol* 2010, 80:188-197.
39. Mowatt G, N'Dow J, Vale L, Nabi G, Boachie C, Cook JA, Fraser C, Griffiths TR; Aberdeen Technology Assessment Review (TAR) Group: Photodynamic diagnosis of bladder cancer compared with white light cystoscopy: Systematic review and meta-analysis. *Int J Technol Assess Health Care* 2011, 27:3-10.

40. Blick CG, Nazir SA, Mallett S, Turney BW, Onwu NN, Roberts IS, Crew JP, Cowan NC: Evaluation of diagnostic strategies for bladder cancer using computed tomography (CT) urography, flexible cystoscopy and voided urine cytology: results for 778 patients from a hospital haematuria clinic. *BJU Int* 2012, 110:84-94.

41. Lotan Y, Capitanio U, Shariat SF, Hutterer GC, Karakiewicz PI: Impact of clinical factors, including a point-of-care nuclear matrix protein-22 assay and cytology, on bladder cancer detection. *BJU Int* 2009, 103:1368-1374.

42. Tyrer J, Duffy SW, Cuzick J: A breast cancer prediction model incorporating familial and personal risk factors. *Stat Med* 2004, 23:1111-1130.

43. Dubsky P, Brase JC, Jakesz R, Rudas M, Singer CF, Greil R, Dietze O, Luisser I, Klug E, Sedivy R, Bachner M, Mayr D, Schmidt M, Gehrmann MC, Petry C, Weber KE, Fisch K, Kronenwett R, Gnant M, Filipits M, on behalf the Austrian Breast and Colorectal Cancer Study Group (ABCSG): The EndoPredict score provides prognostic information on late distant metastases in ER+/HER2− breast cancer patients. *Br J Cancer* 2013, 109:2959-2964.

44. Brentnall AR, Evans DG, Cuzick J: Distribution of breast cancer risk from SNPs and classical risk factors in women of routine screening age in the UK. *Br J Cancer* 2014, 110:827-828.

45. Filipits M, Nielsen TO, Rudas M, Greil R, Stöger H, Jakesz R, Bago-Horvath Z, Dietze O, Regitnig P, Gruber-Rossipal C, Müller-Holzner E, Singer CF, Mlineritsch B, Dubsky P, Bauernhofer T, Hubalek M, Knauer M, Trapl H, Fesl C, Schaper C, Ferree S, Liu S, Cowens JW, Gnant M, Austrian Breast and Colorectal Cancer Study Group: The PAM50 risk-of-recurrence score predicts risk for late distant recurrence after endocrine therapy in postmenopausal women with endocrine-responsive early breast cancer. *Clin Cancer Res* 2014, 20:1298-1305.

46. Krabbe LM, Svatek RS, Shariat SF, Messing E, Lotan Y: Bladder Cancer Risk: Use of the PLCO and NLST to Identify as Suitable Screening Cohort. Urol Oncol. 20149 July 16): pii: S1078-1439 (14)00215-4.

Conclusions

In conclusion, the G+P INDEX reported here shows a significant opportunity to change clinical utility. G+P INDEX is able to accurately triage out patients who present to

their clinician or physician with hematuria, who have a low probability of UC with a high overall test-negative rate, high level of sensitivity and high NPV. This model is suitable for use during primary evaluation of patients with hematuria to triage out patients who do not require a full urological work-up, thereby potentially reducing the number of patients with hematuria requiring referral for specialist urological evaluation for UC, helping to maintain patient QoL and helping to reduce diagnosis-related costs.

Example 5: Triage of Patents with Hematuria Using the G+P Index III ("G2")

A further use of the G+P INDEX is to categorize patients according to Risk of UC. The risk category is then used to prioritize patients for follow up investigation.

This example provides an alternative method for triaging patients using an index, herein termed the "G2 + P Index." It is similar to the G+P Index described above in Example 4.

Definition of G_2

This classifier calculates the following formula:

M1 = [IGF] –[HOXA] + [MDK] + [CDC];

then,

$$R = -6.9802 + 0.2007 * M1 + 1.7893 * [IL8R] + 0.1552 * M1^2 - 0.2882 * [IL8R]^2 - 0.0720 * M1 * [IL8R]$$

to obtain the score:

$SCORE = e^R/(1+ e^R)$, where [IGF], [HOXA], [MDK], [CDC] and [IL8R] are the logarithms of the sample concentrations for the genes IGF, HOXA, MDK, CDC and IL8R, respectively; '*' is ordinary multiplication, and e = 2.718282... is the base of the natural (Napierian) logarithm.

A high SCORE indicates a higher likelihood of UC being present. As an example, we may set a threshold of 0.12 and declare a SCORE >= 0.12 as having a high likelihood of having UC, and those scores below 0.12 has having a low likelihood of having UC.

Example 6: Concurrent Use of [G1 = P] –Index and G2 to Triage Out Patients Who Present with Hematuria Who Have a Low Probability of Urothelial Carcinoma

A. Patents With Micro-Haematuria

Patient data-set

- 45 patient samples were assayed
- 5 patients lack Smoking Status and were excluded from this analysis
- All patients have completed a full urological work-up and none have urothelial carcinoma.
- Phenotypic variables: Gender, Age, SmokingStatus
- Patient demographics are shown in **Table 13** below.

Table 13

Gender	Age	NonSmoker	ExSmoker	Smoker
Female	Age < 60	10	1	2
	Age >= 60	9	1	2
Male	Age < 60	3	2	3
	Age >= 60	3	2	2

Observed Test Negative rate:

Using a Test Negative Rate of 40% as the threshold, data is shown below in **Table 14.**

Table 14

TNR 40%	LR < 0.12	0.12 <= LR < 0.23	0.23 <= LR
Triage Negative	28	3	1
Triage Positive	4	2	2

Using a Negative Test Rate of 50% as the threshold, data is shown below in **Table 15.**

Table 15

TNR 50%	LR < 0.12	0.12 <= LR < 0.23	0.23 <= LR
Triage Negative	28	3	3
Triage Positive	4	2	0

Using a Negative Test Rate of 60% as the threshold gives an identical results/

The highest risk group (Male, Current or Ex-Smoker, Age >= 60) were all positive for this Triage classification.

Summary and Conclusions

1. By deed of the clinical guidelines all 41 patients would normally receive a full urological work-up.
2. All 41 did receive a full-work-up and all 41 were determined to have no urothelial carcinoma.
3. At a Test Negative Rate (TNR of 50 % 85% of the micro-hematuria patients would be screened out and therefore consequentially not receive a full urological work-up.
4. If Cxbladder-triage was used at a TNR of 50% (Triage Index -3.33) 85.4% of the patients would be triaged out and consequently, correctly would not receive a full urological work-up.
5. If Cxbladder-triage was used at a TNR of 40% (Triage Index -3.0) 80.5% of the patients would be triaged out and consequently, correctly would not receive a full urological work-up.

B. Patients with Macrohematuria

Patient Data Set

587 samples from Clinical Trial data and North Shore and CURT product trials were used. This data set was a subset consisting of complete data containing Age, Gender, Smoking Status and Haematuria frequency, as well as gene concentrations for IGF, HOXA, MDK, CDC, IL8R.

We used the same data used to develop the Cxbladder-triage model below;

$$\text{G+P INDEX} = -8.46 + 0.79\,\text{IGF} - 1.60\,\text{HOXA} + 2.10\,\text{MDK} + 0.95\,\text{CDC} - 0.38\,\text{IL8R} + 0.98\,\text{SNS} + 0.56\,\text{Hfreq} + 1.11\,\text{Gender} + 0.64\,\text{Age}$$

We plotted the triage score against G2 diagnostic score in **FIG 24**. The Triage thresholds -3.33, -2.99 and -2.71 correspond to Test Negative rates of 40%, 50% and 60% respectively. The verticals are the Cxbladder thresholds of 0.12 and 0.23. Filled circles

correspond to tumours; green are Ta, and red are all other higher stage tumours. **Table 16** shows the clinical findings.

Table 16

Stage	No-cancer	T1	T2	T2a	T3	Ta	Tis
Count	515	16	10	1	2	40	3

We considered the counts in the 4 quadrants of **FIG. 24** and determined by various cutoffs for Triage Index and G2.

The threshold for the Cxbladder-triage (Triage Index) of -3 corresponds to a Test Negative rate of 50%. **Table 17** shows these results.

Table 17

Quadrant	Thresholds	Control	TCC
Bottom Left	Triage < -3 AND ~~LR~~ G2 < 0.12	272	6
Top Left	Triage > -3 AND ~~LR~~ G2 < 0.12	155	7
Bottom Left	Triage < -3 AND ~~LR~~ G2 >= 0.12	16	0
Top Right	Triage > -3 AND ~~LR~~ G2 >= 0.12	72	59

Using a Cxbladder-triage threshold of -3.33 (Test Negative Rate of 40%) with the same G2 threshold of 0.12, we observed the data shown in **Table 18**.

Table 18

Quadrant	Thresholds	Control	TCC
Bottom Left	Triage < -3.33 AND ~~LR~~ G2 < 0.12	222	4
Top Left	Triage > -3.33 AND ~~LR~~ G2 < 0.12	205	9
Bottom Right	Triage < -3.33 AND ~~LR~~ G2 >= 0.12	9	0
Top Right	Triage > -3.33 AND ~~LR~~ G2 >= 0.12	79	59

Summary and Conclusions

1. Use of a serial combination of Cxbladder Triage (G+P) and Cxbladder Detect (G2) on the same patients in the same time interval provided a comprehensive segregation of patients into four key clinical groupings.

2. For a combined population of patients presenting with micro and macro hematuria and using a Test Negative Rate of 40% there was a total of 235 of the 587 patients (40.0%) of the patients triaged out. We conclude that these patients will not need a full work-up for UC.

3. For the same population the corresponding residual group of 352 (60%) patients will receive a full urological work-up.

4. This residual group contained all high grade and late stage tumours. These patients were not triaged out and would therefore consequently, correctly receive a full urological work-up.

5. A total of 4 low-grade Ta's (5.6% of the total number of tumours) will be triaged out and will not receive the full work-up.

6. If the triage rules allow triage out of all patients below the 40% Test Negative Rate and with G2 scores < 0.12 then there were a total of 13 low grade Ta's triaged out that will not receive a full work-up and all high grade late stage tumours will be caught and receive a full urological work-up.

7. These modified triage rules also resulted in a total of 440 patients triaged out of a total of 587 (75%).

Advantages and **General Conclusions**

In conclusion, the G+P INDEX reported here shows a significant opportunity to change clinical utility. G+P INDEX is able to accurately triage out patients who present to their clinician or physician with haematuria, who have a low probability of UC with a high overall test-negative rate, high level of sensitivity and high NPV. This model is suitable for use by primary care physicians to triage out patients who do not require a full urological work up, thereby reducing the number of patients with haematuria requiring referral for specialist urological evaluation for UC, helping to maintain patient quality of life and reducing diagnosis-related costs. The disclosed methods provided unexpectedly accurate assessment of the lack of need for follow-up investigation for those with hematuria. These represent excellent effects that could not have been achieved without use of the disclosures contained

herein.

Incorporation by Refrence

All patents, patent applications and non-patent literature citations are herein incorporated fully by reference as if separately so incorporated.

Industrial Applicability

Embodiments of this invention are useful in the fields of healthcare and medicine.

Technical Arts

Embodiments of this invention provide highly accurate, sensitive, and specific computer-impolemented methods for triaging patients to determine which patients do not require substantial short-term procedures or follow up. The methods improve computer operations by providing new and non-obvious computer operations based on analysis of specific genetic and phenotypic information from patients with hematuria that yield tangible, useful, and concrete results to improve the quality of health care and reduce cost.

What is Claimed is:

1. A method for determining, in a patient presenting with hematuria,or the level of risk for having urothelial cancer, comprising:

a. providing a sample of urine from said patient;

b. quantifying a value, MI, comprising quantifying the levels of expression of human MDK, CDC2, HOXA13, and IGFBP5 in said sample;

c. assessing the phenotypic variables HFREQ, AgeGT, sex, SMK, and RBC of said patient;

d. calculating G+P INDEX according to either:

formula (i), G+P INDEX = (1*HFREQ+3*Gender+4*SMK) + (5*M1+2*IL-8), or

formula (ii), G+P INDEX=

(w1*HFREQ+w2*AgeGT50+w3*Gender+w4*SMK+w5*RBC) + (w6*M1+w7*IL-8), or

G+P INDEX = -8.46 + 0.79 IGF -1.60 HOXA + 2.10 MDK + 0.95 CDC – 0.38 IL8R + 0.98 SNS + 0.56 Hfreq +1.11 Gender + 0.64 Age; and

e. determining whether the G+P INDEX is greater than a threshold indicating the level of risk that the patient has urothelial cancer.

2. The method of Claim 1 where said threshold is selected from the group of G+P INDEX values of from 0 to 5, from 6 to 10, or from 11-15, where said value of from 0 to 5 indicates Low Risk, 6 to 10 indicates Moderate Risk, and 11-15 indicates High Risk.

3. The method of Claim 2, where if said threshold is a G+P INDEX value of from 6-10, said patient undergoes additional clinical or laboratory tests.

4. The method of Claim 2, where if said threshold is a G+P INDEX value of from 11-15, said patient undergoes additional clinical or laboratory tests.

5. The method of Claim 2, where if said threshold is a G+P INDEX value of from 0-5, the patient is placed on a watch list for further clinical or laboratory tests.

6. The method of any of claims 1 to 5, where the threshold is established using a statistical method.

7. The method of any of claims 1 to 6 wherein the statistical method is any one of Linear Discriminant Analysis (LDA), Logistic Regression (LogReg), Support Vector Machine (SVM), K-nearest 5 neighbors (KN5N), and Partition Tree Classifier (TREE).

8. The method of any of claims 1 to 7, further comprising quantifying expression of one additional genotypic marker selected from Figure 6 or Figure 7.

9. The method of any of claims 1 to 8, where said step of quantifying genetic expression is carried out by detecting the levels of mRNA.

10. The method of any of claims 1 to 8, wherein said step of quantifying genetic expression is carried out by detecting the levels of cDNA.

11. The method of any of claims 1 to 8, where said step of quantifying genetic expression is carried out using an oligonucleotide complementary to at least a portion of said cDNA.

12. The method of any of claims 1 to 9, where said step of quantifying genetic expression is carried out using qRT-PCR method using a forward primer and a reverse primer.

13. The method of any of claims 1 to 8, where said step of quantifying genetic expression is carried out by detecting the levels of a protein.

14. The method of any of claims 1 to 8, where said step of quantifying genetic expression is carried out by detecting the levels of a peptide.

15. The method of any of claims 1 to 8 or 13-14, where said step of quantifying genetic expression is carried out using an antibody directed against said marker.

16. The method of any of claims 1 to 8 or 13-15, where said step of quantifying genetic expression is carried out using a sandwich-type immunoassay method, or using an antibody chip.

17. The method of any of claims 1 to 8 or 13-16, where said quantifying genetic expression is carried out using a monoclonal antibody.

18. The method any of claims 1 to 8 or 13-17, where said quantifying genetic expression is carried out using a polyclonal antiserum.