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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
29 December 2016 (29.12.2016)

WIPO | PCT

(10) International Publication Number
WO 2016/207656 A1

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/GB2016/051903

(22) International Filing Date:
24 June 2016 (24.06.2016)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
1511152.9 24 June 2015 (24.06.2015) GB

(71) Applicant: **UCL BUSINESS PLC** [GB/GB]; The Network Building, 97 Tottenham Court Road, London Greater London W1T 4TP (GB).

(72) Inventors: **FEBER, Andrew**; UCL Cancer Institute, c/o UCL Business Plc, The Network Building, 97 Tottenham Court Road, London Greater London W1T 4TP (GB). **KELLY, John**; Division of Surgery and Interventional Science, c/o UCL Business Plc, The Network Building, 97 Tottenham Court Road, London Greater London W1T 4TP (GB).

(74) Agent: **WILKINSON, Marc George**; J A Kemp, 14 South Square, Gray's Inn, London Greater London WC1R 5JJ (GB).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

WO 2016/207656 A1

(54) Title: METHODS OF DIAGNOSING BLADDER CANCER

(57) Abstract: The present invention relates to methods of diagnosing bladder cancer in a patient, involving determining the methylation status of Methylation Variable Positions (MVPs) in DNA from the patient and providing a diagnosis based on methylation status data. The invention also relates to methods of treating bladder cancer comprising providing a diagnosis of bladder cancer by the diagnostic methods defined herein followed by administering one or more anti-cancer agents to a patient. The invention also relates to methylation-discriminatory arrays comprising probes directed to the MVPs defined herein and kits comprising the arrays.

METHODS OF DIAGNOSING BLADDER CANCER

Field of the Invention

The present invention relates to methods of diagnosing bladder cancer in a patient, involving determining the methylation status of Methylation Variable Positions (MVPs) in DNA from the patient and providing a diagnosis based on methylation status data. The invention also relates to methods of treating bladder cancer comprising providing a diagnosis of bladder cancer by the diagnostic methods defined herein followed by administering one or more anti-cancer agents to a patient. The invention also relates to methylation-discriminatory arrays comprising probes directed to the MVPs defined herein and kits comprising the arrays.

Background to the Invention

Bladder cancer represents one of the most common malignancies in the western world, ranking as the 5th most common cancer in the United States and causing approximately 3% of all cancer-related deaths [1, 2]. The foremost clinical sign at presentation is hematuria, and bladder cancer is detected in about 10% of all such cases investigated [3]. Bladder cancer is more likely in older male patients, current or past smokers and patients exposed to industrial carcinogens [4]. Younger females with nonvisible hematuria are less likely to harbor bladder cancer and for these patients delay in detection of bladder cancer, following misdiagnosis of haematuria, is a frequent event [5]. Cystoscopy is the current gold standard for detecting bladder cancer and is an invasive, uncomfortable procedure requiring clinic or hospital attendance and posing a small but significant risk of infection [6-9].

Each year in the UK, approximately 10,300 people are diagnosed with bladder cancer and 5,000 die from the disease. However, more than 100,000 cases per year are referred from primary care to urology haematuria clinics for cystoscopy and imaging. Bladder cancer is detected in only 10% of patients referred.

Of those with confirmed disease, two thirds are non-muscle invasive bladder cancer (NMIBC) and of these, 70% will recur and 15% will progress to muscle invasive bladder cancer (MIBC). Surveillance by cystoscopy is necessary to detect recurrence

and is performed as frequently as 3 monthly for 2 years then 6 monthly and annually thereafter for cases at high risk of recurrence. The investigation of haematuria and the subsequent surveillance for recurrence poses a significant health economic cost estimated as £55.39 million, ranking bladder as one of the most expensive cancers to manage [10, 11]. There is therefore a great need for improved assays which can better identify patients harbouring disease and reduce the need for unnecessary cystoscopy.

No urinary-based biomarker has FDA approval as a standalone test for the detection of bladder cancer, and consequently guidelines recommend cystoscopy of all patients with visible haematuria and persistent nonvisible haematuria [10, 11]. Urine cytology is frequently used as a diagnostic aid in conjunction with cystoscopy but has low sensitivity to detect cancer other than high grade disease and carcinoma *in situ,* and cannot replace cystoscopy [12, 13]. Similarly commercially available assays based on single targets or small panels of targets fail to detect bladder cancer with sufficient sensitivity and are approved for use only in conjunction with cystoscopy [14].

Several studies have now shown the potential utility of DNA methylation biomarkers in body fluids, including urine [15-23], plasma/serum [24-26], and sputum [27, 28], for the non-invasive detection of cancer. Changes in DNA methylation play a key role in malignant transformation, leading to the silencing of tumor-suppressor genes and overexpression of oncogenes [29]. The ontogenic plasticity and relative stability of DNA methylation makes epigenetic changes ideal biomarkers for diagnosis.

Detection assays involving the presence of specific proteins in voided urine have been developed and commercialised. In these cases the number of proteins detected per assay is low and specificities and sensitivities remain unsatisfactory [14]. Detected protein biomarkers include human complement factor H-related protein, carcinoembryonic antigen (CEA), bladder tumor cell–associated mucins and nuclear mitotic apparatus protein 22 (NMP22).

In terms of assays assessing the expression of certain proteins, WO2014042763 describes a nine-biomarker panel consisting of IL-8, MMP9, SDC1, CCL18, SERPINE1, CD44, VEGF-A, CA9, and ANG for detection of protein in urine samples; a further nine-biomarker panel consisting of CA9, CCL18, MMP12, TMEM45A,

MMP9, SEMA3D, ERBB2, CRH, and MXRA8; as well as a three-biomarker panel consisting of CCL18, CD44, and VEGF-A.

To date, DNA methylation biomarker assays for the detection of bladder cancer have been centered on the analysis of only a small number of loci, in part due to technological limitations and derivation of targets with cancer specificity [11-19]. In general, reported sensitivities and specificities are high relative to established assays, but would fail to attain performance characteristics achieved by cystoscopy. Methylation markers for bladder cancer previously studied include *DAPK*, *BCL2*, *TERT*, *TWIST1*, *NID2*, *RARbeta, E-cadherin* and *p16*. International patent application publication WO2013/144362 describes a diagnostic assay for bladder cancer involving detecting methylation of the promoter of the *ECRG4* and/or the *ITIH5* gene. US patent application publication US2013224738 describes a diagnostic assay for bladder cancer involving assessing the methylation status of genes consisting of *BCL2*, *CDKN2A* and *NID2*.

Improved assays for the accurate diagnosis of bladder cancer are sought and would be of significant clinical and economic benefit, particularly assays which are non-invasive.

Summary of the Invention

Diagnostic methods are provided which can detect bladder cancer from a biological sample, particularly a voided urine sample, with robust and high sensitivity and specificity, and which have the potential to reduce the need for cystoscopy in patients referred with haematuria and in patients undergoing surveillance for disease recurrence. Avoiding cystoscopy will reduce the cost of bladder cancer management and positively impact on patient wellbeing, reducing both the number of hospital visits and the inconvenience of an invasive investigation. Thus the invention provides the following:

The invention provides a method of diagnosing bladder cancer in an individual comprising:

(a) providing DNA from a sample from the individual;

(b) determining whether each one of a group of MVPs selected from a panel comprising the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG] is methylated, wherein the group comprises at least 25 of the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG]; and

(c) diagnosing bladder cancer in the individual when at least 25 of the MVPs of the group of (b) are methylated.

In any such method the group of MVPs may comprises at least 40 of the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG], and wherein bladder cancer is diagnosed when at least 25 of the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG] are methylated.

The group of MVPs may comprise at least 50 of the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG], or may comprise at least 100 of the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG].

The group of MVPs may comprise all 150 of the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG].

In methods described above, cancer may be diagnosed in the individual when at least 40 of the MVPs selected from the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG] are methylated, or when at least 50 of the MVPs selected from the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG] are methylated, or when at least 100 of the MVPs are methylated, or when all 150 MVPs are methylated.

In methods described above, the MVPs determined to be methylated may include the MVPs identified in SEQ ID NOS 1 to 3 and denoted by [CG], or may include the MVPs identified in SEQ ID NOS 1 to 5 and denoted by [CG], or may include the MVPs identified in SEQ ID NOS 1 to 10 and denoted by [CG], or may include the MVPs identified in SEQ ID NOS 1 to 40 and denoted by [CG].

In methods described above, the group of MVPs may comprise all 150 of the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG], wherein bladder cancer is diagnosed in the individual when at least 40 of the MVPs selected from the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG] are methylated, and wherein the MVPs determined to be methylated include the MVPs identified in SEQ ID NOS 1 to 10 and denoted by [CG].

In any of the methods described above, the step of determining whether each one the MVPs is methylated may comprise bisulphite converting the DNA.

In any of the methods described above, the step of determining whether each one the MVPs is methylated may comprise:

1) performing a sequencing step to determine the sequence of MVPs;
2) hybridising DNA to an array comprising probes capable of discriminating between methylated and non-methylated forms of MVPs and applying a detection system to the array to discriminate methylated and non-methylated forms of the MVPs; or
3) performing an amplification step using methylation-specific primers, wherein the status of an MVP as methylated or non-methylated is determined by the presence or absence of an amplified product.

Before the sequencing or hybridization steps, an amplification step may be performed, wherein loci comprising each MVP are amplified. Amplification may be performed by PCR.

A capturing step may be performed before the sequencing or hybridization steps. The capturing step may involve binding polynucleotides comprising the MVP loci to binding molecules specific to the MVP loci and collecting complexes comprising MVP loci and binding molecules; and wherein:

i. the capturing step occurs before the step of bisulphite converting the DNA;
ii. the capturing step occurs after the step of bisulphite converting the DNA but before the amplification or hybridization steps; or
iii. the capturing step occurs after the step of bisulphite converting the DNA and after the amplification step.

The binding molecules may be oligonucleotides specific for each MVP, preferably DNA or RNA molecules each comprising a sequence which is complementary to the corresponding MVP.

The binding molecule may be coupled to a purification moiety.

The purification moiety may comprise a first purification moiety and the step of collecting complexes comprising MVP loci and binding molecules may comprise

binding the first purification moiety to substrates comprising a second purification moiety, wherein first and second purification moieties form an interaction complex.

The first purification moiety may be biotin and the second purification moiety may be streptavidin; or the first purification moiety may be streptavidin and the second purification moiety may be biotin.

The step of amplifying loci comprising MVPs may comprise the use of primers which are independent of the methylation status of the MVP.

The step of amplifying loci comprising MVPs may be performed by microdroplet PCR amplification.

In any of the methods described above, the biological sample obtained from the individual may be a sample of urine, blood, serum, plasma or cell-free DNA.

In any of the methods described above, the method may achieve a ROC sensitivity of 95% or greater and a ROC specificity of 90% or greater; preferably a ROC sensitivity of 96% and a ROC specificity of 97%, preferably a ROC AUC of 95% or greater, preferably 98%.

In any of the methods described above, the method may achieve a negative predictive value (NPV) of 95% or greater, preferably 97%.

In any of the methods described above, the step of diagnosing bladder cancer in the individual may further comprise:

I. stratifying the grade of the tumor; and/or

II. determining the risk of recurrence of the tumor; and/or

III. determining the risk of progression of non-muscular invasive disease; and/or

determining the likely response to treatment therapy.

The invention additionally provides a method of treating bladder cancer in an individual comprising:

(a) obtaining DNA from a sample from the individual and determining whether each one of a group of MVPs selected from a panel comprising the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG] is methylated, wherein the group comprises at least 25 of the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG];

(b) diagnosing bladder cancer in the individual when at least 25 MVPs of the group of (a) are methylated; and

(c) administering one or more bladder cancer treatments to the individual.

The invention additionally provides a method of treating bladder cancer in an individual comprising:

(a) providing DNA from a sample from the individual and determining whether each one of a group of MVPs selected from a panel comprising the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG] is methylated, wherein the group comprises at least 25 of the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG];

(b) diagnosing bladder cancer in the individual when at least 25 MVPs of the group of (a) are methylated; and

(c) administering one or more bladder cancer treatments to the individual.

The invention additionally provides a method of treating bladder cancer in an individual comprising:

(a) determining whether each one of a group of MVPs selected from a panel comprising the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG] is methylated in DNA from a sample from the individual, wherein the group comprises at least 25 of the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG];

(b) diagnosing bladder cancer in the individual when at least 25 MVPs of the group of (a) are methylated; and

(c) administering one or more bladder cancer treatments to the individual.

The invention additionally provides a method of treating bladder cancer in an individual comprising administering one or more bladder cancer treatments to the individual, wherein the individual has been diagnosed with bladder cancer by steps comprising:

(a) providing DNA from a sample from the individual and determining whether each one of a group of MVPs selected from a panel comprising the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG] is methylated, wherein

the group comprises at least 25 of the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG]; and

(b) diagnosing bladder cancer in the individual when at least 25 MVPs of the group of (a) are methylated.

The invention additionally provides a method of diagnosing bladder cancer in an individual comprising:

(a) obtaining data which identify whether each one of a group of MVPs selected from a panel comprising the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG] is methylated, wherein the group comprises at least 25 of the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG]; and

(b) diagnosing bladder cancer in the individual when at least 25 MVPs of the group of (a) are methylated;

wherein the data were obtained by a method comprising:

i. obtaining DNA from the sample; and

ii. determining whether MVPs are methylated in the DNA.

In any of the methods described above the cancer may be a non-muscle invasive bladder cancer (NMIBC). The cancer may be a muscle invasive bladder cancer (MIBC).

The invention additionally provides an array capable of discriminating between methylated and non-methylated forms of MVPs; the array comprising oligonucleotide probes specific for a methylated form of each MVP in a MVP panel and oligonucleotide probes specific for a non-methylated form of each MVP in the panel; wherein the panel consists of at least 25 MVPs selected from the MVPs identified in SEQ ID NOS 1 to 150.

In certain embodiments the array is not an Infinium HumanMethylation450 BeadChip array. In certain embodiments the number of MVP-specific oligonucleotide probes of the array is less than 482,421, preferably 482,000 or less, 480,000 or less, 450,000 or less, 440,000 or less, 430,000 or less, 420,000 or less, 410,000 or less, or 400,000 or less.

In an array as described above, the panel may consist of at least 40 MVPs selected from the MVPs identified in SEQ ID NOS 1 to 150; preferably at least 50

MVPs, at least 60 MVPs, at least 70 MVPs, at least 80 MVPs, at least 90 MVPs, at least 100 MVPs, at least 110 MVPs, at least 120 MVPs, at least 130 MVPs, at least 140 MVPs, at least 145 MVPs, or all 150 MVPs identified in SEQ ID NOS 1 to 150.

In an array as described above, the panel may include the MVPs defined by SEQ ID NOS 1 to 3, or the MVPs defined by SEQ ID NOS 1 to 5, or the MVPs defined by SEQ ID NOS 1 to 10, or the MVPs defined by SEQ ID NOS 1 to 20, or the MVPs defined by SEQ ID NOS 1 to 30, or the MVPs defined by SEQ ID NOS 1 to 40, or the MVPs defined by SEQ ID NOS 1 to 50, or the MVPs defined by SEQ ID NOS 1 to 60, or the MVPs defined by SEQ ID NOS 1 to 70, or the MVPs defined by SEQ ID NOS 1 to 80, or the MVPs defined by SEQ ID NOS 1 to 90, or the MVPs defined by SEQ ID NOS 1 to 100, or the MVPs defined by SEQ ID NOS 1 to 100, or the MVPs defined by SEQ ID NOS 1 to 120, or the MVPs defined by SEQ ID NOS 1 to 130, or the MVPs defined by SEQ ID NOS 1 to 140, or the MVPs defined by SEQ ID NOS 1 to 150.

The panel may include all MVPs defined by SEQ ID NOS 1 to 150.

In array as described above, the array may further comprise one or more oligonucleotides comprising a MVP selected from any of the MVPs defined in SEQ ID NOS 1 to 150, wherein the one or more oligonucleotides are hybridized to corresponding oligonucleotide probes of the array.

The one or more oligonucleotides may comprise at least 20 MVPs selected from the MVPs identified in SEQ ID NOS 1 to 150; preferably at least 50 MVPs, at least 60 MVPs, at least 70 MVPs, at least 80 MVPs, at least 90 MVPs, at least 100 MVPs, at least 110 MVPs, at least 120 MVPs, at least 130 MVPs, at least 140 MVPs, at least 145 MVPs, or all 150 MVPs identified in SEQ ID NOS 1 to 150.

The one or more oligonucleotides may comprise the MVPs defined by SEQ ID NOS 1 to 10, or the MVPs defined by SEQ ID NOS 1 to 20, or the MVPs defined by SEQ ID NOS 1 to 30, or the MVPs defined by SEQ ID NOS 1 to 40, or the MVPs defined by SEQ ID NOS 1 to 50, or the MVPs defined by SEQ ID NOS 1 to 60, or the MVPs defined by SEQ ID NOS 1 to 70, or the MVPs defined by SEQ ID NOS 1 to 80, or the MVPs defined by SEQ ID NOS 1 to 90, or the MVPs defined by SEQ ID NOS 1 to 100, or the MVPs defined by SEQ ID NOS 1 to 110, or the MVPs defined by SEQ ID

NOS 1 to 120, or the MVPs defined by SEQ ID NOS 1 to 130, or the MVPs defined by SEQ ID NOS 1 to 140, or the MVPs defined by SEQ ID NOS 1 to 150.

The one or more oligonucleotides may comprise all MVPs defined by SEQ ID NOS 1 to 150.

Arrays as described above may be obtainable by hybridizing to an array as described above a group of oligonucleotides each comprising a different MVP selected from any of the MVPs defined in SEQ ID NOS 1 to 150, and wherein the group comprises at least 40 oligonucleotides.

In such a hybridized array, the group may comprise at least 50 oligonucleotides. The group may comprise at least 60, at least 70, at least 80, at least 90, at least 100, at least 110, at least 120, at least 130, at least 140, at least 145, or at least 150 oligonucleotides.

In a hybridized array, the group may comprise at least 40 oligonucleotides comprising the MVPs defined by SEQ ID NOS 1 to 20, or wherein the group may comprise at least 50 oligonucleotides comprising the MVPs defined by SEQ ID NOS 1 to 50, or wherein the group may comprise at least 60 oligonucleotides comprising the MVPs defined by SEQ ID NOS 1 to 60, or wherein the group may comprise at least 70 oligonucleotides comprising the MVPs defined by SEQ ID NOS 1 to 70, or wherein the group may comprise at least 80 oligonucleotides comprising the MVPs defined by SEQ ID NOS 1 to 80, or wherein the group may comprise at least 90 oligonucleotides comprising the MVPs defined by SEQ ID NOS 1 to 90, or wherein the group may comprise at least 100 oligonucleotides comprising the MVPs defined by SEQ ID NOS 1 to 100, or wherein the group may comprise at least 110 oligonucleotides comprising the MVPs defined by SEQ ID NOS 1 to 110, or wherein the group may comprise at least 120 oligonucleotides comprising the MVPs defined by SEQ ID NOS 1 to 120, or wherein the group may comprise at least 130 oligonucleotides comprising the MVPs defined by SEQ ID NOS 1 to 130, or wherein the group may comprise at least 140 oligonucleotides comprising the MVPs defined by SEQ ID NOS 1 to 140, or wherein the group may comprise at least 145 oligonucleotides comprising the MVPs defined by SEQ ID NOS 1 to 145, or wherein the group may comprise at least 150 oligonucleotides

comprising the MVPs defined by SEQ ID NOS 1 to 150. The group may comprise at least the 150 oligonucleotides comprising the MVPs defined by SEQ ID NOS 1 to 150.

The invention also provides a process for making the hybridized array as defined above, comprising contacting an array as defined above with a group of oligonucleotides each comprising a different MVP selected from any of the MVPs defined in SEQ ID NOS 1 to 150, and wherein the group comprises at least 25 oligonucleotides.

The invention also provides a process for making a hybridized array as defined above, comprising contacting an array as defined above with a group of oligonucleotides as defined above.

The invention also provides a kit comprising any of the arrays described above.

The kit may further comprise a DNA modifying regent that is capable of modifying a non-methylated cytosine in a MVP dinucleotide but is not capable of modifying a methylated cytosine in a MVP dinucleotide, optionally wherein the dinucleotide is CpG. The DNA modifying regent may be a bisulphite reagent.

Brief description of the Figures

Figure 1. Heatmap of 9786 MVPs (1746 hypermethylated MVPs, 8040 hypomethylated MVPs) between bladder cancer (red) and 30 normal urothelium (blue).

Figure 2. Heatmap of the 150 UroMark loci for non-cancer urine (n=10), normal urothelium (n=30), bladder cancer (n=81) and blood (n=489).

Figure 3. Heat map of the 150 loci involved in the UroMark assay for normal urothelium (Blue) bladder cancer (Red) compared with the predicted (Light blue/light red) and actual (Blue/Red) status of bladder cancer.

Figure 4. MDS plots for bladder tumour and normal urothelium based on the methylation state of 150 loci in bladder cancer samples from UCL and TCGA Bladder Cancer. The MDS (Multidimensional scaling) plot represents the dissimilarly of

phenotypes based on the methylation state of the 150 loci with the panel, and clearly shows that the 150 marker can accurately separate tumour from normal bladder. Axis represent the Euclidean distance between samples.

Figure 5. ROC plot for UroMark model for the detection of bladder cancer in urine.

Figure 6. ROC plots for top performing (A) 3, (B) 5 and (C) 10 marker panels. Top 3 MVPs are listed as SEQ ID NOs: 1 to 3, Top 5 MVPs are listed as SEQ ID NOs: 1 to 5, Top 10 MVPs are listed as SEQ ID NOs: 1 to 10, all in rank order (see Table 1).

Figure 7. ROC plot for UroMark model for the detection of bladder cancer from 176 unique urine samples (98 non-cancer urines and 78 cancer urines).

Figure 8. Comparison of DNA quality (concentration, purity and integrity) from patients' urine. A) Concentration and purity of samples collected in clinic (n=123) compared with samples collected at home (n=41) where red indicates low yield or purity, amber indicates moderate yield and intermediate purity and green is high yield and purity. B) Percentage of samples exhibiting low, intermediate or high yield and purity. C) Representative Bioanalyzer electropherogram demonstrating recovery of high molecular weight of DNA in samples with concentrations quantifiable by spectrophotometry. D) Increased median yield and improved purity of urinary DNA using an extended digest step at 21°C compared with the manufacturer's standard protocol of 56°C for 1 hour. E) Comparison of two urine preservation methods: UCL established standard operating procedure versus Norgen urine preservation tubes. The number of bacteria were quantified by qPCR for the rpoB gene and expressed as copies rpoB per copy of human YWHAZ. Data are the mean +/- standard deviation.

Figure 9. ROC plot for UroMark model for the detection of bladder cancer from Validation Cohort 2 - 96 unique urine samples (64 non-cancer urines and 32 cancer urines).

Figure 10. ROC plot for UroMark model for the detection of bladder cancer from Validation Cohort 3 - 92 urine samples (65 non-cancer urines and 27 cancer urines).

Detailed Description of the Invention

Bladder cancer

As discussed above, bladder cancers represent one of the most prevalent groups of cancers in the western world. Transitional cell carcinoma is the most common type, and accounts for approximately 90% of bladder cancers. Transitional cell carcinomas arise from the transitional epithelium, which is a tissue lining the inner surface of the bladder. The remaining 10% of bladder cancers are mainly comprised of squamous cell carcinoma, adenocarcinoma, sarcoma, and small cell carcinoma. Squamous cell carcinoma also arises from epithelial tissue, from squamous cells. These are thin, flat cells found in the most superficial epithelial layer. Adenocarcinomas form from epithelial cells having glandular characteristics and/or origin. Sarcomas derive from cells of mesenchymal origin, such as the cells of the fat and muscle layers of the bladder. Small cell carcinomas have a rapid doubling time and are capable of earlier metastases, making them particularly aggressive.

Bladder cancers may also be classified as non-muscle invasive bladder cancer (NMIBC) and muscle invasive bladder cancer (MIBC).

The diagnostic and treatment methods described herein are capable of positively identifying malignant cells of all classifications of bladder cancer. Thus, any of the methods described herein may be used to diagnose transitional cell carcinoma of the bladder, squamous cell carcinoma of the bladder, adenocarcinoma of the bladder, sarcoma of the bladder, small cell carcinoma of the bladder, metastatic bladder cancer, leiomyosarcoma (a tumor arising from smooth muscle), lymphoma (a tumor that usually arises in the lymph nodes), malignant melanoma (a tumor that usually arises from the skin) and large cell neuroendocrine carcinoma. Primary forms and recurrent forms of bladder cancer are included. The cancer to be diagnosed or treated as described herein may be a urothelial cell cancer. Thus, the cancer may be cancer of the ureter, urethra or renal pelvis.

The most preferred patient type to which the diagnostic assays described herein are applicable are humans. The diagnostic assays described herein may also be used to identify bladder cancer in a non-human animal. For example, non-human animals may contain tissue derived from humans, e.g. xenografts. Thus, diagnostic assays may be used to diagnose human bladder cancer in an animal model of human bladder cancer. Typical non-human animals to which the diagnostic assays described herein are applicable are rodents such as rats or mice.

Methylation Variable Positions (MVPs)

Methylation of DNA is a recognised form of epigenetic modification which has the capability of altering the expression of genes and other elements such as microRNAs [51]. In cancer development and progression, methylation may have the effect of *e.g.* silencing tumor suppressor genes and/or increasing the expression of oncogenes. Other forms of dysregulation may occur as a result of methylation. Methylation of DNA occurs at discrete loci which are predominately dinucleotide consisting of a CpG motif, but may also occur at CHH motifs (where H is A, C, or T). During methylation, a methyl group is added to the fifth carbon of cytosine bases to create methylcytosine.

Methylation can occur throughout the genome and is not limited to regions with respect to an expressed sequence such as a gene. Methylation typically, but not always, occurs in a promoter or other regulatory region of an expressed sequence.

A Methylation Variable Position (MVP) as defined herein is any dinucleotide locus which may show a variation in its methylation status between phenotypes, *i.e.* between tumour and normal tissue. An MVP is preferably a CpG or a CHH dinucleotide motif. An MPV as defined herein is not limited to the position of the locus with respect to a corresponding expressed sequence.

Typically, an assessment of DNA methylation status involves analysing the presence or absence of methyl groups in DNA, for example methyl groups on the 5^{th} position of one or more cytosine nucleotides. Preferably, the methylation status of one or more cytosine nucleotides present as a CpG dinucleotide (where C stands for Cytosine, G for Guanine and p for the phosphate group linking the two) is assessed.

By assessing the methylation status of an MVP or determining whether an MVP is methylated it is meant that a determination is made as to whether an MVP was methylated or unmethylated in the starting sample of DNA obtained from the individual prior to subsequent processing.

An MVP is herein defined as methylated if one or more alleles of that MVP in a sample of genomic DNA from the patient is determined to possess one or more methylated CpG dinucleotide loci.

In any of the methods described herein, the MVPs determined to be methylated are methylated relative to normal urothelium control and/or whole blood control.

Specific MVPs useful for diagnostic purposes are set forth in Table 1 and are identified by SEQ ID number, as well as Illumina ID number (Ilmn ID). Exemplary primers for amplifying the defined MVPs are set forth in Table 2 and are also identified by SEQ ID number.

Identification and assessment of Methylation Variable Position (MVP) status

A variety of techniques are available for the identification and assessment of Methylation Variable Positions (MVPs), as will be outlined briefly below. The diagnostic methods described herein encompass any suitable technique for the determination of MVP status.

Methyl groups are lost from a starting DNA molecule during conventional *in vitro* handling steps such as PCR. To avoid this, techniques for the detection of methyl groups commonly involve the preliminary treatment of DNA prior to subsequent processing, in a way that preserves the methylation status information of the original DNA molecule. Such preliminary techniques involve three main categories of processing, *i.e.* bisulphite modification, restriction enzyme digestion and affinity-based analysis. Products of these techniques can then be coupled with sequencing or array-based platforms for subsequent identification or qualitative assessment of MVP methylation status.

Techniques involving bisulphite modification of DNA have become the most common methods for detection and assessment of methylation status of CpG dinucleotide. Treatment of DNA with bisulphite, *e.g.* sodium bisulphite, converts

cytosine bases to uracil bases, but has no effect on 5-methylcytosines. Thus, the presence of a cytosine in bisulphite-treated DNA is indicative of the presence of a cytosine base which was previously methylated in the starting DNA molecule. Such cytosine bases can be detected by a variety of techniques. For example, primers specific for unmethylated versus methylated DNA can be generated and used for PCR-based identification of methylated CpG dinucleotides. A separation/capture step may be performed, *e.g.* using binding molecules such as complementary oligonucleotide sequences. Standard and next-generation DNA sequencing protocols can also be used.

In other approaches, methylation-sensitive enzymes can be employed which digest or cut only in the presence of methylated DNA. Analysis of resulting fragments is commonly carried out using microarrays.

Affinity-based techniques exploit binding interactions to capture fragments of methylated DNA for the purposes of enrichment. Binding molecules such as anti-5-methylcytosine antibodies are commonly employed prior to subsequent processing steps such as PCR and sequencing.

Olkhov-Mitsel and Bapat (2012) [51] provide a comprehensive review of techniques available for the identification and assessment of MVP-based biomarkers involving methylcytosine.

For the purposes of assessing the methylation status of the MVP-based biomarkers characterised and described herein, any suitable method can be employed.

Preferred methods involve bisulphite treatment of DNA, including amplification of the identified MVP loci for methylation specific PCR and/or sequencing and/or assessment of the methylation status of target loci using methylation-discriminatory microarrays.

Amplification of MVP loci can be achieved by a variety of approaches. Preferably, MVP loci are amplified using PCR. MVP may also be amplified by other techniques such as multiplex ligation-dependent probe amplification (MLPA). A variety of PCR-based approaches may be used. For example, methylation-specific primers may be hybridized to DNA containing the MVP sequence of interest. Such primers may be designed to anneal to a sequence derived from either a methylated or non-methylated MVP locus. Following annealing, a PCR reaction is performed and the

presence of a subsequent PCR product indicates the presence of an annealed MVP of identifiable sequence. In such methods, DNA is bisulphite converted prior to amplification. Such techniques are commonly referred to as methylation specific PCR (MSP) [53].

In other techniques, PCR primers may anneal to the MVP sequence of interest independently of the methylation status, and further processing steps may be used to determine the status of the MVP. Assays are designed so that the MVP site(s) are located between primer annealing sites. This method scheme is used in techniques such as bisulphite genomic sequencing [54], COBRA [55], Ms-SNuPE [56]. In such methods, DNA can be bisulphite converted before or after amplification.

Preferably, small-scale PCR approaches are used. Such approaches commonly involve mass partitioning of samples (*e.g.* digital PCR). These techniques offer robust accuracy and sensitivity in the context of a highly miniaturised system (pico-liter sized droplets), ideal for the subsequent handling of small quantities of DNA obtainable from the potentially small volume of cellular material present in biological samples, particularly urine samples. A variety of such small-scale PCR techniques are widely available. For example, microdroplet-based PCR instruments are available from a variety of suppliers, including RainDance Technologies, Inc. (Billerica, MA; http://raindancetech.com/) and Bio-Rad, Inc. (http://www.bio-rad.com/). Microarray platforms may also be used to carry out small-scale PCR. Such platforms may include microfluidic network-based arrays *e.g.* available from Fluidigm Corp. (www.fluidigm.com).

Following amplification of MVP loci, amplified PCR products may be coupled to subsequent analytical platforms in order to determine the methylation status of the MVPs of interest. For example, the PCR products may be directly sequenced to determine the presence or absence of a methylcytosine at the target MVP or analysed by array-based techniques.

Any suitable sequencing techniques may be employed to determine the sequence of target DNA. In the methods of the present invention the use of high-throughput, so-called "second generation", "third generation" and "next generation" techniques to sequence bisulphite-treated DNA are preferred.

In second generation techniques, large numbers of DNA molecules are sequenced in parallel. Typically, tens of thousands of molecules are anchored to a given location at high density and sequences are determined in a process dependent upon DNA synthesis. Reactions generally consist of successive reagent delivery and washing steps, *e.g.* to allow the incorporation of reversible labelled terminator bases, and scanning steps to determine the order of base incorporation. Array-based systems of this type are available commercially *e.g.* from Illumina, Inc. (San Diego, CA; http://www.illumina.com/).

Third generation techniques are typically defined by the absence of a requirement to halt the sequencing process between detection steps and can therefore be viewed as real-time systems. For example, the base-specific release of hydrogen ions, which occurs during the incorporation process, can be detected in the context of microwell systems (*e.g.* see the Ion Torrent system available from Life Technologies; http://www.lifetechnologies.com/). Similarly, in pyrosequencing the base-specific release of pyrophosphate (PPi) is detected and analysed. In nanopore technologies, DNA molecules are passed through or positioned next to nanopores, and the identities of individual bases are determined following movement of the DNA molecule relative to the nanopore. Systems of this type are available commercially *e.g.* from Oxford Nanopore (https://www.nanoporetech.com/). In an alternative method, a DNA polymerase enzyme is confined in a "zero-mode waveguide" and the identity of incorporated bases are determined with florescence detection of gamma-labeled phosphonucleotides (see *e.g.* Pacific Biosciences; http://www.pacificbiosciences.com/).

In other methods in accordance with the invention sequencing steps may be omitted. For example, amplified PCR products may be applied directly to hybridization arrays based on the principle of the annealing of two complementary nucleic acid strands to form a double-stranded molecule. Hybridization arrays may be designed to include probes which are able to hybridize to amplification products of an MVP and allow discrimination between methylated and non-methylated loci. For example, probes may be designed which are able to selectively hybridize to an MVP locus containing thymine, indicating the generation of uracil following bisulphite conversion of an unmethylated cytosine in the starting template DNA. Conversely, probes may be

designed which are able to selectively hybridize to an MVP locus containing cytosine, indicating the absence of uracil conversion following bisulphite treatment. This corresponds with a methylated MVP locus in the starting template DNA.

Following the application of a suitable detection system to the array, computer-based analytical techniques can be used to determine the methylation status of an MVP. Detection systems may include, *e.g.* the addition of fluorescent molecules following a methylation status-specific probe extension reaction. Such techniques allow MVP status determination without the specific need for the sequencing of MVP amplification products. Such array-based discriminatory probes may be termed methylation-specific probes.

Any suitable methylation-discriminatory microarrays may be employed to assess the methylation status of the MVPs described herein. A preferred methylation-discriminatory microarray system is provided by Illumina, Inc. (San Diego, CA; http://www.illumina.com/). In particular, the Infinium HumanMethylation450 BeadChip array system may be used to assess the methylation status of diagnostic MVPs for bladder cancer as described herein. Such a system exploits the chemical modifications made to DNA following bisulphite treatment of the starting DNA molecule. Briefly, the array comprises beads to which are coupled oligonucleotide probes specific for DNA sequences corresponding to the unmethylated form of an MVP, as well as separate beads to which are coupled oligonucleotide probes specific for DNA sequences corresponding to the methylated form of an MVP. Candidate DNA molecules are applied to the array and selectively hybridize, under appropriate conditions, to the oligonucleotide probe corresponding to the relevant epigenetic form. Thus, a DNA molecule derived from an MVP which was methylated in the corresponding genomic DNA will selectively attach to the bead comprising the methylation-specific oligonucleotide probe, but will fail to attach to the bead comprising the non-methylation-specific oligonucleotide probe. Single-base extension of only the hybridized probes incorporates a labeled ddNTP, which is subsequently stained with a fluorescence reagent and imaged. The methylation status of the MVP may be determined by calculating the ratio of the fluorescent signal derived from the methylated and unmethylated sites.

Because the bladder cancer-specific diagnostic MVP biomarkers defined herein were initially identified using the Illumina Infinium HumanMethylation450 BeadChip array system, the same chip system can be used to interrogate those same MVPs in the diagnostic assays described herein. Alternative or customised arrays could, however, be employed to interrogate the bladder cancer-specific diagnostic MVP biomarkers defined herein, provided that they comprise means for interrogating all MVPs for a given method, as defined herein.

Techniques involving combinations of the above-described methods may also be used. For example, DNA containing MVP sequences of interest may be hybridized to microarrays and then subjected to DNA sequencing to determine the status of the MVP as described above.

In the methods described above, sequences corresponding to MVP loci may also be subjected to an enrichment process. DNA containing MVP sequences of interest may be captured by binding molecules such as oligonucleotide probes complementary to the MVP target sequence of interest. Sequences corresponding to MVP loci may be captured before or after bisulphite conversion or before or after amplification. Probes may be designed to be complementary to bisulphite converted DNA. Captured DNA may then be subjected to further processing steps to determine the status of the MVP, such as DNA sequencing steps.

Capture/separation steps may be custom designed. Alternatively a variety of such techniques are available commercially, *e.g.* the SureSelect target enrichment system available from Agilent Technologies (http://www.agilent.com/home). In this system biotinylated "bait" or "probe" sequences (*e.g.* RNA) complementary to the DNA containing MVP sequences of interest are hybridized to sample nucleic acids. Streptavidin-coated magnetic beads are then used to capture sequences of interest hybridized to bait sequences. Unbound fractions are discarded. Bait sequences are then removed (*e.g.* by digestion of RNA) thus providing an enriched pool of MVP target sequences separated from non-MVP sequences. In a preferred method of the invention, template DNA is subjected to bisulphite conversion and target loci are then amplified by small-scale PCR such as microdroplet PCR using primers which are independent of the methylation status of the MVP. Following amplification, samples are subjected to a

capture step to enrich for PCR products containing the target MVP, *e.g.* captured and purified using magnetic beads, as described above. Following capture, a standard PCR reaction is carried out to incorporate DNA sequencing barcodes into MVP-containing amplicons. PCR products are again purified and then subjected to DNA sequencing and analysis to determine the presence or absence of a methylcytosine at the target genomic MVP [31].

The MVP biomarker loci defined herein are identified e.g. by Illumina® identifiers (IlmnID). These MVP loci identifiers refer to individual MVP sites used in the commercially available Illumina® Infinium Human Methylation450 BeadChip kit. The identity of each MVP site represented by each MVP loci identifier is publicly available from the Illumina, Inc. website under reference to the MVP sites used in the Infinium Human Methylation450 BeadChip kit.

Further information regarding MVP loci identification used in Illumina, Inc products is found in the technical note entitled "Technical Note: Epigenetics. CpG Loci Identification. A guide to Illumina's method for unambiguous CpG loci identification and tracking for the Golden Gate® and Infinium® Assay for Methylation" published in 2010 and found at:

http://www.illumina.com/documents/products/technotes/technote_cpg_loci_identification.pdf.

Further information regarding the Illumina® Infinium Human Methylation450 BeadChip system can be found at:

http://www.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/datasheet_humanmethylation450.pdf;

and at:

http://www.illumina.com/content/dam/illumina-marketing/documents/products/technotes/technote_hm450_data_analysis_optimization.pdf.

To complement evolving public databases to provide accurate MVP/CpG loci identifiers and strand orientation, Illumina® has developed a method to consistently designate MVP/CpG loci based on the actual or contextual sequence of each individual MVP/CpG locus. To unambiguously refer to MVP/CpG loci in any species, Illumina®

has developed a consistent and deterministic MVP loci database to ensure uniformity in the reporting of methylation data. The Illumina® method takes advantage of sequences flanking a MVP locus to generate a unique MVP locus cluster ID. This number is based on sequence information only and is unaffected by genome version. Illumina's standardized nomenclature also parallels the TOP/BOT strand nomenclature (which indicates the strand orientation) commonly used for single nucleotide polymorphism (SNP) designation.

Illumina® Identifiers for the Infinium Human Methylation450 BeadChip system are also available from public repositories such as Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/).

An MVP as defined herein thus refers to the CG dinucleotide motif identified in relation to each SEQ ID NO. and Illumina Identifier (Ilmn ID) as listed in Table 1, wherein the cytosine base of the dinucleotide (noted in bold and square brackets in the sequences listed at Table 1) may (or may not) be modified. Thus by determining the methylation status of a CpG defined by or identified in a given SEQ ID NO., or determining whether such a CpG is methylated, it is meant that a determination is made as to whether the cytosine of the CG dinucleotide motif identified in bold and in square brackets in a sequence shown in Table 1 is methylated or not at one or more loci in the sample of DNA from the individual, accepting that variation in the sequence upstream and downstream of any given CpG may exist due to sequencing errors or variation between individuals.

The invention provides a method of diagnosing bladder cancer in an individual comprising:

(a) providing DNA from a sample from the individual;

(b) determining whether each one of a group of MVPs selected from a panel comprising the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG] is methylated, wherein the group comprises at least 25 of the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG]; and

(c) diagnosing bladder cancer in the individual when at least 25 of the MVPs of the group of (b) are methylated.

In any such method described herein, the group of MVPs (i.e. those MVPs the methylation status of which are to be determined) may comprise 26 or more of the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG]; or the group may comprise 27 or more, 28 or more, 29 or more, 30 or more, 31 or more, 32 or more, 33 or more, 34 or more, 35 or more, 36 or more, 37 or more, 38 or more, 39 or more, 40 or more, 41 or more, 42 or more, 43 or more, 44 or more, 45 or more, 46 or more, 47 or more, 48 or more, 49 or more, 50 or more, 51 or more, 52 or more, 53 or more, 54 or more, 55 or more, 56 or more, 57 or more, 58 or more, 59 or more, 60 or more, 61 or more, 62 or more, 63 or more, 64 or more, 65 or more, 66 or more, 67 or more, 68 or more, 69 or more, 70 or more, 71 or more, 72 or more, 73 or more, 74 or more, 75 or more, 76 or more, 77 or more, 78 or more, 79 or more, 80 or more, 81 or more, 82 or more, 83 or more, 84 or more, 85 or more, 86 or more, 87 or more, 88 or more, 89 or more, 90 or more, 91 or more, 92 or more, 93 or more, 94 or more, 95 or more, 96 or more, 97 or more, 98 or more, 99 or more, 100 or more, 101 or more, 102 or more, 103 or more, 104 or more, 105 or more, 106 or more, 107 or more, 108 or more, 109 or more, 110 or more, 111 or more, 112 or more, 113 or more, 114 or more, 115 or more, 116 or more, 117 or more, 118 or more, 119 or more, 120 or more, 121 or more, 122 or more, 123 or more, 124 or more, 125 or more, 126 or more, 127 or more, 128 or more, 129 or more, 130 or more, 131 or more, 132 or more, 133 or more, 134 or more, 135 or more, 136 or more, 137 or more, 138 or more, 139 or more, 140 or more, 141 or more, 142 or more, 143 or more, 144 or more, 145 or more, 146 or more, 147 or more or 148 or more of the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG]. The group may comprise 149 or 150 of the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG].

In any of the methods described above, bladder cancer may be diagnosed when at least 25 of the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG] are methylated. Bladder cancer may be diagnosed when 26 or more of the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG] are methylated; or when 27 or more, 28 or more, 29 or more, 30 or more, 31 or more, 32 or more, 33 or more, 34 or more, 35 or more, 36 or more, 37 or more, 38 or more, 39 or more, 40 or more, 41 or more, 42 or more, 43 or more, 44 or more, 45 or more, 46 or more, 47 or more, 48 or more, 49 or

more, 50 or more, 51 or more, 52 or more, 53 or more, 54 or more, 55 or more, 56 or more, 57 or more, 58 or more, 59 or more, 60 or more, 61 or more, 62 or more, 63 or more, 64 or more, 65 or more, 66 or more, 67 or more, 68 or more, 69 or more, 70 or more, 71 or more, 72 or more, 73 or more, 74 or more, 75 or more, 76 or more, 77 or more, 78 or more, 79 or more, 80 or more, 81 or more, 82 or more, 83 or more, 84 or more, 85 or more, 86 or more, 87 or more, 88 or more, 89 or more, 90 or more, 91 or more, 92 or more, 93 or more, 94 or more, 95 or more, 96 or more, 97 or more, 98 or more, 99 or more, 100 or more, 101 or more, 102 or more, 103 or more, 104 or more, 105 or more, 106 or more, 107 or more, 108 or more, 109 or more, 110 or more, 111 or more, 112 or more, 113 or more, 114 or more, 115 or more, 116 or more, 117 or more, 118 or more, 119 or more, 120 or more, 121 or more, 122 or more, 123 or more, 124 or more, 125 or more, 126 or more, 127 or more, 128 or more, 129 or more, 130 or more, 131 or more, 132 or more, 133 or more, 134 or more, 135 or more, 136 or more, 137 or more, 138 or more, 139 or more, 140 or more, 141 or more, 142 or more, 143 or more, 144 or more, 145 or more, 146 or more, 147 or more, or 148 or more of the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG] are methylated. Bladder cancer may be diagnosed when 149 or 150 of the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG] are methylated.

Preferably, bladder cancer may be diagnosed when 40 or more of the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG] are methylated.

Bladder cancer may also be diagnosed when 50 or more, 60 or more, 70 or more or 80 or more, 90 or more or 100 or more of the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG] are methylated.

In any of the methods described above the MVPs determined to be methylated may include the MVPs identified in SEQ ID NOS 1 to 3 and denoted by [CG], or may include the MVPs identified in SEQ ID NOS 1 to 5 and denoted by [CG], or may include the MVPs identified in SEQ ID NOS 1 to 10 and denoted by [CG], or may include the MVPs identified in SEQ ID NOS 1 to 20 and denoted by [CG], or may include the MVPs identified in SEQ ID NOS 1 to 30 and denoted by [CG], or may include the MVPs identified in SEQ ID NOS 1 to 40 and denoted by [CG], or may include the MVPs identified in SEQ ID NOS 1 to 50 and denoted by [CG], or may

include the MVPs identified in SEQ ID NOS 1 to 60 and denoted by [CG], or may include the MVPs identified in SEQ ID NOS 1 to 70 and denoted by [CG], or may include the MVPs identified in SEQ ID NOS 1 to 80 and denoted by [CG], or may include the MVPs identified in SEQ ID NOS 1 to 90 and denoted by [CG], or may include the MVPs identified in SEQ ID NOS 1 to 100 and denoted by [CG].

In one embodiment, the group of MVPs (i.e. those MVPs the methylation status of which are to be determined) may comprises all 150 of the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG], and in this method bladder cancer is diagnosed in the individual when at least 40 of the MVPs selected from the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG] are methylated, and in this method the MVPs determined to be methylated include the MVPs identified in SEQ ID NOS 1 to 10 and denoted by [CG].

Bioinformatic tools and statistical metrics

Software programs which aid in the *in silico* analysis of bisulphite converted DNA sequences and in primer design for the purposes of methylation-specific analyses are generally available and have been described previously [57, 58, 59].

Sensitivity and specificity metrics for bladder cancer diagnosis based on the MVP methylation status assays described herein may be defined using standard receiver operating characteristic (ROC) statistical analysis [52]. In ROC analysis 100% sensitivity corresponds to a finding of no false negatives, and 100% specificity corresponds to a finding of no false positives.

Based on analyses conducted using a panel of 150 MVP biomarkers, a bladder cancer diagnostic assay in accordance with the invention described herein can achieve a ROC sensitivity of 90% or greater, 91% or greater, 92% or greater, 93% or greater, 94% or greater, 95% or greater, 96% or greater, 97% or greater, 98% or greater or 99%. The ROC sensitivity may be 100%.

Diagnostic assays in accordance with the invention can achieve a ROC specificity of 90% or greater, 91% or greater, 92% or greater, 93% or greater, 94% or greater, 95% or greater, 96% or greater, 97% or greater, 98% or greater or 99%. The ROC specificity may be 100%.

Diagnostic assays in accordance with the invention may have an associated combination of ROC sensitivity and ROC specificity values wherein the combination is any one of the above-listed sensitivity values and any one of the above-listed specificity values, provided that the sensitivity value is equal to or less than the specificity value.

Thus, the ROC specificity may be 100% and the ROC sensitivity may be 90% or greater, 91% or greater, 92% or greater, 93% or greater, 94% or greater, 95% or greater, 96% or greater, 97% or greater, 98% or greater, 99% or 100%.

The ROC specificity may be 99% and the ROC sensitivity may be 90% or greater, 91% or greater, 92% or greater, 93% or greater, 94% or greater, 95% or greater, 96% or greater, 97% or greater, 98% or 99%.

The ROC specificity may be 98% and the ROC sensitivity may be 90% or greater, 91% or greater, 92% or greater, 93% or greater, 94% or greater, 95% or greater, 96% or greater, 97% or 98%.

The ROC specificity may be 97% and the ROC sensitivity may be 90% or greater, 91% or greater, 92% or greater, 93% or greater, 94% or greater, 95% or greater, 96% or 97%.

The ROC specificity may be 96% and the ROC sensitivity may be 90% or greater, 91% or greater, 92% or greater, 93% or greater, 94% or greater, 95% or 96%.

The ROC specificity may be 95% and the ROC sensitivity may be 90% or greater, 91% or greater, 92% or greater, 93% or greater, 94% or 95%.

The ROC specificity may be 94% and the ROC sensitivity may be 90% or greater, 91% or greater, 92% or greater, 93% or 94%.

The ROC specificity may be 93% and the ROC sensitivity may be 90% or greater, 91% or greater, 92% or 93%.

The ROC specificity may be 92% and the ROC sensitivity may be 90% or greater, 91% or 92%.

The ROC specificity may be 91% and the ROC sensitivity may be 90% or 91%.

The ROC specificity may be 90% and the ROC sensitivity may be 90%.

Preferably, the assay may achieve a ROC sensitivity of 95% or greater and a ROC specificity of 90% or greater; preferably a ROC sensitivity of 96% and a ROC specificity of 97%.

ROC plots corresponding to example methods representative of the diagnostic methods defined herein are presented at Figures 5 and 7, demonstrating the exquisite sensitivity and selectivity of the MVP-based assays. This contrasts with assays conducted with smaller panels of biomarkers, which have been described previously. Thus, comparative data demonstrate the superior predictive power of the assays defined herein.

A further metric which can be employed to classify the accuracy of the MVP-based assays is ROC AUC. In ROC analysis, the area under the curve of a ROC plot (AUC) is a metric for binary classification. In a random binary classifier the number of true positives and false positives will be approximately equal. In this situation the AUC score for the ROC plot will be 0.5. In a perfect binary classifier the number of true positives will be 100% and the number of false positives will be 0%. In this situation the AUC score for the ROC plot will be 1.

Based on analyses conducted using biomarkers described herein, a bladder cancer diagnostic assay in accordance with the invention can achieve a ROC AUC of 0.90 or greater, 0.91 or greater, 0.92 or greater, 0.93 or greater, 0.94 or greater, 0.95 or greater, 0.96 or greater, 0.97 or greater, 0.98 or greater, 0.99 or 1. Preferably the diagnostic assay can achieve a ROC AUC of 0.98 or greater.

Bladder cancer diagnostic tests based on the MVP methylation status assays described herein may also be characterised using a Negative Predictive Value (NPV) metric. The NPV is a measure of the proportion of negative results that are true negative results.

Based on analyses conducted using a panel of 150 MVP biomarkers, a bladder cancer diagnostic assay in accordance with the invention described herein can achieve an NPV of 90% or greater, 91% or greater, 92% or greater, 93% or greater, 94% or greater, 95% or greater, 96% or greater, 97% or greater, 98% or greater or 99% or 100%.

Biological samples

The bladder cancer diagnostic assays described herein may be performed on any suitable biological material obtained from the patient. Preferred biological

material is urine. However, samples of bladder tissue, *e.g.* obtained via biopsy or aspirates, or obtained from preserved samples (*e.g.* cryopreserved material, tissue sections etc.) may be used. Samples of biological material may also include solid tissue samples, aspirates, samples of biological fluids, blood, serum, plasma, ascitic fluid, lymph, peripheral blood, cerebrospinal fluid, fine needle aspirate, saliva, sputum, bone marrow, skin, epithelial samples (including buccal, cervical or vaginal epithelia) or other tissue derived from the ectoderm, vaginal fluid, semen etc. Tissue scrapes may include biological material from e.g. buccal, oesophageal, bladder, vaginal, urethral or cervical scrapes. The cells of the sample may comprise inflammatory cells, such as lymphocytes.

Any of the assays and methods described herein may involve providing a biological sample from the patient as the source of patient DNA for methylation analysis.

Any of the assays and methods described herein may involve obtaining patient DNA from a biological sample which has previously been obtained from the patient.

Any of the assays and methods described herein may involve obtaining a biological sample from the patient as the source of patient DNA for methylation analysis. Procedures for obtaining a biological sample from the patient may be non-invasive, such as collecting cells from urine. Alternatively, invasive procedures such as biopsies may be used.

In the methods described herein the level of detection is such that 2 tumor cells may be detected in a sample comprising 150,000 cells or more. In such methods the sample may comprise 160,000 cells or more, 170,000 cells or more, 180,000 cells or more, 190,000 cells or more, 200,000 cells or more, 210,000 cells or more, 220,000 cells or more, 230,000 cells or more, 240,000 cells or more, 250,000 cells or more, 260,000 cells or more, 270,000 cells or more, 280,000 cells or more, 280,000 cells or more, or 300,000 cells or more.

In any such method, the number of tumor cells that can be detected is 10 or more, 20 or more, 30 or more, 40 or more, 50 or more, 60 or more, 70 or more, 80 or more, 90 or more, 100 or more, 200 or more, 300 or more, 400 or more, 500 or more, 600 or more, 700 or more, 800 or more, 900 or more, 1000 or more, 2000 or more, 3000

or more, 4000 or more, 5000 or more, 6000 or more, 7000 or more, 8000 or more, 9000 or more, 10000 or more, 20000 or more, 30000 or more, 40000 or more, 50000 or more, 60000 or more, 70000 or more, 80000 or more, 90000 or more or 100000 or more.

Methods of treatment

The invention also encompasses the performance of one or more treatment steps following a positive diagnosis of bladder cancer by the diagnostic methods described herein.

Thus the invention also encompasses a method of treating bladder cancer in an individual comprising:

(a) obtaining DNA from a sample from the individual and determining whether each one of a group of MVPs selected from a panel comprising the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG] is methylated, wherein the group comprises at least 25 of the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG];

(b) diagnosing bladder cancer in the individual when at least 25 MVPs of the group of (a) are methylated; and

(c) administering one or more bladder cancer treatments to the individual.

The invention also encompasses a method of treating bladder cancer in an individual comprising:

(a) providing DNA from a sample from the individual and determining whether each one of a group of MVPs selected from a panel comprising the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG] is methylated, wherein the group comprises at least 25 of the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG];

(b) diagnosing bladder cancer in the individual when at least 25 MVPs of the group of (a) are methylated; and

(c) administering one or more bladder cancer treatments to the individual.

The invention also encompasses a method of treating bladder cancer in an individual comprising:

(a) determining whether each one of a group of MVPs selected from a panel comprising the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG] is methylated in DNA from a sample from the individual, wherein the group comprises at least 25 of the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG];

(b) diagnosing bladder cancer in the individual when at least 25 MVPs of the group of (a) are methylated; and

(c) administering one or more bladder cancer treatments to the individual.

The invention also encompasses a method of treating bladder cancer in an individual comprising administering one or more bladder cancer treatments to the individual, wherein the individual has been diagnosed with bladder cancer by steps comprising:

(a) providing DNA from a sample from the individual and determining whether each one of a group of MVPs selected from a panel comprising the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG] is methylated, wherein the group comprises at least 25 of the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG]; and

(b) diagnosing bladder cancer in the individual when at least 25 MVPs of the group of (a) are methylated.

In any of the above-described methods of treating bladder cancer, the group of MVPs which are selected from a panel comprising the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG] (i.e. the group of MVPs whose methylation status is to be determined) may comprise any number of MVPs as described and defined herein, provided that the group comprises at least 25 of the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG]. In any of these methods, bladder cancer may be diagnosed in the individual when the number of MVPs of the group which are determined to be methylated is any number of MVPs as described and defined herein, provided that at least 25 of the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG] are determined to be methylated.

Thus, the invention encompasses administration of one or more surgical procedures, one or more chemotherapeutic agents, one or more immunotherapeutic

agents, one or more radiotherapeutic agents, one or more hormonal therapeutic agents or any combination of the above following a positive diagnosis of bladder cancer.

Surgical procedures include transurethral resection of bladder tumor (TURBT), cystectomy, open radical cystectomy (ORC), laparoscopic radical cystectomy (LRC) and robot-assisted radical cystectomy (RARC).

Chemotherapeutic agents include the following. Alkylating agents, which include the nitrogen mustards, nitrosoureas, tetrazines, aziridines, cisplatin and platinum based derivatives, as well as the non-classical alkylating agents. Antimetabolites, which include the anti-folates, fluoropyrimidines, deoxynucleoside analogues and thiopurines. Microtubule disrupting agents, which include the vinca alkaloids and taxanes, as well as dolastatin 10 and derivatives thereof. Topoisomerase inhibitors, which include camptothecin, irinotecan and topotecan. Topoisomerase II poisons, which include etoposide, doxorubicin, mitoxantrone and teniposide. Topoisomerase II catalytic inhibitors, which include novobiocin, merbarone, and aclarubicin. Cytotoxic antibiotics, which include anthracyclines, actinomycin, bleomycin, plicamycin, and mitomycin.

Combinations of agents include but are not limited to MVAC (Methotrexate, Vinblastine, Vinblastine and Vinblastine), Gem-Cis (GC) (Gemcitabine and Cisplatin), Lapatinib and gemcitabine.

Immunotherapeutics include bacilli Calmette-Guérin (BCG) immunotherapy as well as monoclonal antibodies and antibody-drug conjugates. Antibody-drug conjugates include antibodies conjugated to microtubule disrupting agents and DNA modifying agents as described above.

Combination therapies include intravesical, sequential BCG, followed by electromotive administration (EMDA) of MMC (EMDA-MMC) as well as microwave-induced bladder wall hyperthermia (HT) and intravesical MMC.

Cancer therapeutic agents are administered to a subject already suffering from a disorder or condition, in an amount sufficient to cure, alleviate or partially arrest the condition or one or more of its symptoms. Such therapeutic treatment may result in a decrease in severity of disease symptoms, or an increase in frequency or duration of symptom-free periods. An amount adequate to accomplish this is defined as

"therapeutically effective amount". Effective amounts for a given purpose will depend on the severity of the disease as well as the weight and general state of the subject. As used herein, the term "subject" includes any human.

Arrays

The invention also encompasses arrays capable of discriminating between methylated and non-methylated forms of MVPs as defined herein; the arrays may comprise oligonucleotide probes specific for methylated forms of MVPs as defined herein and oligonucleotide probes specific for non-methylated forms of MVPs as defined herein.

By “specific” it is meant that the probes comprise sequences which are complementary to those of the oligonucleotides comprising the MVP such they may hybridize, particularly under stringent conditions.

In some embodiments the array is not an Illumina Infinium HumanMethylation450 BeadChip array (Infinium HumanMethylation450 BeadChip array).

Separately or additionally, in some embodiments the number of MVP-specific oligonucleotide probes of the array is less than 482,421, preferably 482,000 or less, 480,000 or less, 450,000 or less, 440,000 or less, 430,000 or less, 420,000 or less, 410,000 or less, or 400,000 or less, 375,000 or less, 350,000 or less, 325,000 or less, 300,000 or less, 275,000 or less, 250,000 or less, 225,000 or less, 200,000 or less, 175,000 or less, 150,000 or less, 125,000 or less, 100,000 or less, 75,000 or less, 50,000 or less, 45,000 or less, 40,000 or less, 35,000 or less, 30,000 or less, 25,000 or less, 20,000 or less, 15,000 or less, 10,000 or less, 5,000 or less, 4,000 or less, 3,000 or less or 2,000 or less.

The invention further encompasses the use of any of the arrays as defined herein in any of the methods which require determining the methylation status of MVPs for the purposes of diagnosing bladder cancer cells in an individual.

Kits

Any of the arrays as defined herein may be comprised in a kit.

The kit may comprise any array as defined herein.

The kit may comprise any array as defined herein together with instructions for use.

The kit may additionally comprise a DNA modifying regent, such as a bisulphite reagent.

The kit may additionally comprise reagents for amplifying DNA, such as primers directed to any of the MVPs as defined herein as identified in SEQ ID NOS 1 to 150 (see Table 2).

Methods of determining a methylation profile of a sample

The invention further encompasses a method of determining a methylation profile of a sample from an individual, the method comprising:

i. providing DNA from a sample from the individual;

ii. determining whether each one of a group of MVPs selected from a panel comprising the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG] is methylated, wherein the group comprises at least 25 of the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG]; and

iii. based on the methylation status of the MVPs of the group, determining a methylation profile of the sample.

In any of the above-described methods of determining a methylation profile of a sample, the group of MVPs which are selected from a panel comprising the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG] (i.e. the group of MVPs whose methylation status is to be determined) may comprise any number of MVPs as described and defined herein, provided that the group comprises at least 25 of the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG].

Furthermore, in any such methods, the methylation status of MVPs may be determined using any of the arrays described herein.

Further methods

In any of the diagnostic methods described herein, the step of diagnosing bladder cancer in the individual may further comprise:

I. stratifying the grade of the tumor; and/or

II. determining the risk of recurrence of the tumor; and/or

III. determining the risk of progression of non-muscular invasive disease; and/or

IV. determining the likely response to treatment therapy.

The invention also encompasses a method of determining the risk of the development of bladder cancer in an individual, the method comprising:

(a) providing DNA from a sample from the individual;

(b) determining whether each one of a group of MVPs selected from a panel comprising the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG] is methylated, wherein the group comprises at least 25 of the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG]; and

(c) based on the methylation status of the MVPs of the group, determining the risk of the development of bladder cancer in the individual.

In any such method of determining the risk of the development of bladder cancer in an individual, the group of MVPs which are selected from a panel comprising the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG] (i.e. the group of MVPs whose methylation status is to be determined) may comprise any number of MVPs as described and defined herein, provided that the group comprises at least 25 of the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG].

Furthermore, in any such methods, the methylation status of MVPs may be determined using any of the arrays described herein.

Further uses

The invention also encompasses the use of a group of MVPs in the diagnosis of bladder cancer in an individual or in determining the risk of the development of bladder cancer in an individual.

In any such use, the group of MVPs are selected from a panel comprising the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG].

In any such use, the group of MVPs may comprise any number of MVPs as described and defined herein, provided that the group comprises at least 25 of the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG].

In any of the above-described uses, the diagnosis of bladder cancer in an individual or the determination of the risk of the development of bladder cancer in an individual may be performed by any of the respective methods described and defined herein. Furthermore, in any such methods, the methylation status of MVPs may be determined using any of the arrays described herein.

The invention is illustrated by the following Examples:

Examples

Materials and methods

Introduction

Emerging techniques that utilize next-generation DNA sequencing platforms hold particular promise for the development of highly sensitive epigenetic biomarker panels. For example, the microdroplet-based PCR amplification of bisulphite converted DNA, followed by next-generation sequencing of the amplified target loci developed by RainDance Technologies [30] enables the sensitive, specific and simultaneous amplification of up to 20,000 bisulfite-converted target loci. Highly parallel microdroplet-based PCR amplification of bisulphite-converted DNA has shown utility in the validation of epigenetic alterations in a range of tissues [31-33]. However, as yet it has not been applied to the development of non-invasive diagnostic biomarkers for the detection of bladder cancer.

To derive a sensitive assay for the detection of bladder cancer, the inventors have performed one of the largest independent studies of genome-wide methylation in bladder cancer to date. From this, a panel of bladder-specific epigenetic biomarkers have been defined and the sensitivity and specificity of a 150 loci panel using RainDrop-BS [31] on urinary DNA have been validated for the detection of bladder cancer with a high degree of diagnostic precision.

Study Population

Genome-wide DNA methylation profiling was performed on DNA from 81 bladder cancers and 30 age-matched normal urothelium samples collected from UCLH (London) Department of Urology and CIEMAT (Madrid). The cohort included 35 low-grade non-muscle-invasive cancers. Pathological review of representative H&E sections was conducted to include specimens with tumor cellularity > 80%. Blood methylome data was download from the MARMAL-aid database (http:/marmal-aid.org, [34]).

Independent validation data were obtained from the Cancer Genome Atlas Project (https://tcga-data.nci.nih.gov/tcga/tcgaCancerDetails.jsp?diseaseType=BLCA&diseaseName=Bladder%20Urothelial%20Carcinoma), consisting of MIBC bladder cancer and 20 normal urothelium samples.

For validation studies sequential urine samples were collected from patients attending UCLH one-stop haematuria and surveillance cystoscopy clinics (n= 86 bladder cancer, n= 96 non-cancer controls).

Urine collection

Urine samples were obtained from patients attending haematuria clinics or undergoing cystoscopy for recurrent bladder cancer at University College Hospital. For comparison of clinic versus home urine samples, patients were asked to supply three samples: a clinic sample and two home samples, one of which should be a first void. 40-100ml were obtained per sample. The home urine kit for one sample comprised up to four 25ml sterile tubes, mailing tubes with absorbent pads and a pre-addressed padded envelope, designed to fit through a Royal Mail post box.

DNA extraction and quantification

Urinary DNA was extracted using a DNeasy blood and tissue kit (Qiagen, UK) in accordance with the manufacturer's instructions. DNA was quantified by spectrophotometry (Nanodrop 1000) and fluorometry (Qubit ds DNA HS assay, Invitrogen, UK).

RainDance Microdroplet PCR

For microdroplet PCR, 7.20 µl of bisulfite-treated (and optionally, whole-genome amplified) DNA were added to 4.70 µl of 10× High-Fidelity Buffer (Invitrogen), 1.80 µl of 50 mM $MgSO_4$ (Invitrogen), 1.62 µl of 10 mM dNTP solution mix (NEB), 3.60 µl of 4 M betaine solution (Sigma-Aldrich), 3.60 µl of droplet stabilizer (RainDance Technologies), 1.80 µl of 100% dimethyl sulfoxide (Sigma-Aldrich) and 0.72 µl of 5 U/µl Platinum Taq Polymerase High-Fidelity (Invitrogen), to a total volume of 25 µl. The sample plate was sealed using an ALPS 50V microplate heat sealer (Thermo Scientific).

The bisulfite-treated genomic DNA template mix was then applied to a fully automated ThunderStorm system (RainDance Technologies) following the manufacturer's instructions. In brief, primer panel droplets (MethylSeq Solution, RainDance Technologies) were dispensed to a microfluidic chip. The DNA template mix was converted into droplets within the microfluidic chip. The primer pair droplets and template droplets were then paired together in a 1:1 ratio. Paired droplets passed through an electric field inducing the discrete droplets to coalesce into a single PCR droplet (26 pl); approximately 1 million PCR droplets are collected per sample.

PCR droplets were processed in a PTC-225 thermocycler (MJ Research) as follows: 94°C for 2 min; 55 cycles of 94°C for 30 s, 54°C for 45 s, 68°C for 80 s; followed by 68°C for 10 min; 4°C until further processing. The ramp rate was set to 1°C per second. Following PCR amplification, 70 µl of droplet destabilizer (RainDance Technologies) were added to each sample to break the PCR droplet emulsion and release the amplicons contained within the droplets. The solution was mixed well and incubated for 15 min at RT. Samples were purified using Agencourt AMPure XP magnetic beads (Beckman Coulter) following the manufacturer's protocol. For each sample, 234 µl of beads were used. Samples were eluted from magnetic beads in 40 µl of Buffer. The integrity and concentration (fragment range: 120–300 bp) of purified amplicon DNA were assessed using a High Sensitivity DNA Kit (Agilent Technologies) on a 2100 Bioanalyzer (Agilent Technologies).

Methylation Array

500 ng of DNA was bisulphite converted and hybridised to the Infinium 450K Human Methylation array and processed in accordance with the manufacturer's recommendations. DNA bisulphite conversion was carried out using the EZ DNA Methylation kit (Zymo Research) as per the manufacturer's instructions. Samples were processed in a single batch. R statistical software (version 2.14.0 [35]) was used for the subsequent data analysis. The ChAMP pipeline was used to extract and analyze data from iDat files, samples were normalised using BMIQ [36, 37]. Raw β values (methylation value) were subjected to a stringent quality control analysis as follows: samples showing reduced coverage were removed and only probes with detection levels above background across all samples were retained (detection $P < 0.01$). DMRs (Differentially Methylated Regions) were determined using *Lasso* [38, 39].

High-throughput DNA sequencing

The pooled sequencing library (12 pM) and custom sequencing primers (0.5 μM) were applied to a MiSeq-cycle PE consumable cartridge (Illumina) according to the manufacturer's protocol. The DNA sequences of the custom sequencing primers are provided in Table 2 below. Sequencing was performed on a MiSeq DNA sequencer (Illumina) using 75-bp paired-end reads.

The RainDance ThunderStorm® System was also used for the sequencing of nucleic acids (http://raindancetech.com/targeted-dna-sequencing/thunderstorm/).

Data and statistical analyses

Sequencing adapters were trimmed from the raw sequencing reads using the fastq-mcf tool of ea-utils v1.1.2-537 [60]. Trimmed sequencing data were mapped to an in silico bisulfite-converted human reference genome (GRCh37) using Bismark v0.7.12 [40, 61]. Methylation information was extracted using the methylation_extractor tool of Bismark v0.7.12 [61]. Targeted DNA sequencing analyses were performed using the R package TEQC v3.2.0.25.

Example 1: Methylation profiling of low and high grade bladder cancer

The epigenetic alterations associated with bladder cancer were initially defined by performing genome-wide DNA methylation profiling on DNA from 81 high-grade and 30 normal urothelium.

Supervised analysis, using a Wilcoxon rank-sum test to assign directionality, was used to identify MVPs (Methylation Variable Positions) between bladder cancer and normal tissue. MVPs were selected on the basis of statistical significance (Wilcoxon P-value>0.001). An additional filter of $\Delta\beta$>0.30(+/-) was applied to compensate for not taking into account the absolute difference in methylation between the groups. The cut-off is empirically defined to result in a false discovery rate (FDR) of <2% and reduced candidate loci to those with largest methylation differences and therefore greatest potential for a discriminatory effect. A total of 9786 MVPs met these requirements (1746 hypermethylated MVPs, 8040 hypomethylated MVPs) (Figure 1).

Example 2: Bladder cancer specific urinary biomarkers

To define a DNA methylation biomarker panel, those loci which were determined to be methylated (β >50%) in at least 50% of cancers and unmethylated in normal urothelium (β <10%) were identified. By "methylated (β >50%)", as discussed herein in relation to the development of this initial DNA methylation biomarker panel, it is meant that for any given locus, >50% of cells in a patient sample are determined to be methylated with respect to that MVP. In order to remove potential false positive biomarkers and better define alterations which are bladder cancer specific, whole blood and urine from 10 patients attending hematuria clinics and who had a confirmed non-cancer diagnosis was also profiled. Subsequently, any loci which showed any methylation (β > 10%) in DNA from these non-cancer control urine and bloods were removed. A maximum of 432 loci were identified which are unmethylated in non-cancers and methylated (β >50%) in the majority of cancer tissue.

To derive a bladder cancer DNA methylation signature for detection we used the Random Forest framework, which resulted in a classification signature consisting of

150 CpG loci (Figure 2, see Table 1 below). Using this core set of 150 markers we performed an internal cross validation of the classifier with predicted likelihood values *i.e.* likelihood of a sample being cancer or not for each sample independent of its relationship to the group of samples. This resulted in a cross validated sensitivity of 100% and specificity of 100% for the detection of cancer, showing that 150 epigenetic loci can clearly stratify normal urothelium from bladder cancer (Figure 3, Figure 4).

To determine sensitivity of the 150 CpG (MVP) marker panel for detection of bladder cancer using the classification algorithm, we assessed the methylation profiles of a further 179 bladder cancers (144 muscle invasive and 35 non-muscle invasive) and 20 normal cases. The panel correctly classified all bladder cancers, with a resulting sensitivity and specificity of 1.

Example 3: Validation of detection panel

To test the 150 loci panel (epi-signature) for the detection of bladder cancer in urine samples, DNA from urinary sediment cells was obtained from a cohort (n=86) including 52 patients with bladder cancer (low-grade = 27, intermediate/high-grade = 25) and 34 non-cancer patient controls. Applying the epi-signature to this validation cohort of give a sensitivity of 95% and specificity of 96% and an AUC of 97% for the detection of bladder cancer in this independent validation cohort (Figure 5).

The large marker panel was also compared to the best performing single markers from the training cohort, this includes CpG loci from regions previously published as potential urinary biomarkers in genes including *OTX1*, *COD1* and *MEIS1*. For each CpG a methylation threshold was defined, based on the highest β-value obtained from non-bladder cancer controls in the training cohort. This value was then used to predict the likely presence of bladder cancer in the validation cohort. The best performing single markers were combined and a predictive classifier developed to explore the potential for an "oligo panel" based on 3, 5 or 10 markers. Although sensitivities improve over single markers (best single marker 72%, best combined marker 70%), they still do not reach the required level to replace cystoscopy.

These data show that although single markers perform reasonably well alone or in combination (Figures 4-6, Table 1), the sensitivity to detect cancer using single loci or an oligo panel approach is limited and below a detection level desirable for clinical utility. Use of a large panel of markers out performs that of a single marker panel and Table 3 shows the sensitivity, specificity, AUC, PPV and NPV for the top 10 best performing markers and the combined 150 loci signature.

Example 4: Validation of detection panel using high throughput technology

RainDrop BS-seq [31], allows large scale targeted bisulphite sequencing of a large number of regions (up to 20,00 unique amplicons) in parallel. This technology has been validated previously and shown to be highly correlated with the 450K methylation array, and its utility with low template input has also been validated [31, 32]. A bisulphite converted sequencing primer panel was designed to measure the methylation state of the 150 selected genomic loci (see Table 2 below). Primers were designed to interrogate both Watson and Crick strands independently where possible. Validation of the urinary epi-signature was conducted using RainDrop BS-seq in a second independent cohort of 96 cases. DNA from urinary sediment cells was obtained from 26 patients with bladder cancer and 64 non-cancer patient controls. Methylation score for each of the 150 loci was generated using the Bismark algorithm, using this data the urinary epi-signature predicted the likely presence of bladder cancer with a sensitivity of 96%, specificity of 97% and an AUC of 0.96.

Combining the methylation data from all validation samples allows an increase in the number of samples tested. Combining samples allowed assessment of 176 unique urine samples, 98 non-cancer urines and 78 cancer urines. The urinary epi-signature predicted the presence of bladder cancer with an AUC = 0.98, independent of profiling technology (Figure 7).

Conclusion

Biomarker-driven early non-invasive detection of bladder cancer has the potential to radically improve the management of this disease. Highly sensitive and specific assays have the potential utility in both the detection of *de novo* disease in patients attending haematuria clinic and also in the screening for recurrence in existing bladder cancer populations.

Several non-invasive tests are commercially available, and are based on cytology, FISH 311 analysis, and detection of mutations. Despite being FDA approved, the tests have reported sensitivities of 54% to 86% and specificities of 61% to 90% [12, 13, 41]. Performance characteristics are not sufficient to replace cystoscopy and therefore have not been taken up into clinical practice. There is therefore significant room for improvement and development of novel biomarkers, combinations of biomarker panels and the use of novel technologies may be most helpful for this purpose.

DNA methylation patterns are highly cell-specific and the ontogenic stability of these epigenetic events make DNA methylation an ideal biomarker for the detection and diagnosis of disease. Changes in global DNA methylation patterns are a common feature of neoplastic transformation and is a frequent event in bladder cancer. Previous studies have shown that methylation changes between bladder cancer, both non-muscle invasive bladder cancer (NMIBC) and muscle invasive bladder cancer (MIBC) and normal urothelium are reflected in urinary sediment cells from bladder cancer patients, and as such could be a useful diagnostic marker. Several studies now have shown the utility of urinary epigenetic markers in the diagnosis of bladder cancer. However, although these studies have shown good sensitivities and specificities, they have failed to be taken up in clinical practice, predominately because they still do not command the performance characteristics to replace cystoscopy. DNA methylation biomarker assays (along with other DNA based markers, *e.g.* mutations) have been limited by the low resolution of primary analysis in identifying putative biomarkers, using either a candidate approach or a low-resolution microarray based platforms, and also by the limitations in the technology available for analysing candidate markers in urine. This has resulted in single/small biomarker panels being interrogated in the final biomarker

panel [15-23]. However, in order for small biomarker panels to show the sensitivity and specificity to match those of cystoscopy they rely heavily on a low intra and inter tumour heterogeneity across a wide spectrum of disease states [42].

Novel technologies, such as next generation bisulphite sequencing and large scale multiplex PCR, now allow for larger panels of epigenetic biomarkers to be utilised [31, 32]. In order to define the epigenetic alterations involved in bladder cancer development and define a biomarker panel one of the largest unbiased genome wide DNA methylation screens of bladder cancer to date was carried out. Besides allowing an insight into the epigenetic alterations driving bladder cancer development, from these data a panel of the epigenetic biomarkers which has high sensitivity and specificity for detection of bladder cancer was also identified.

A biomarker panel of 150 single CpG loci was defined, which are predictive of bladder cancer. Although only a relatively small cohort, these data show the utility of a using a large panel of epigenetic markers, compared to single marker and small panel biomarker panels. With a sensitivity of 96% and a specificity of 97%, the negative predictive value (NPV) of this test is 97%. This is 1.2 to 8.7 times superior to what can be achieved by PSA (prostate-specific antigen) testing (PPV = 30%–43%), mammography (PPV = 9%–19%), or fecal occult blood screening (PPV = 6%–11%; [43-48]) and similarly to that of cystoscopy (PPV = 66.7%– 98%) [49].

The inventors have shown that by applying a large scale highly multiplexed next generation assay, which is both highly sensitive and quantitative, the presence of bladder cancer in urine can be detected with a higher sensitivity and specificity than previously published methylation assays, and has a PPV comparable to cystoscopy.

The invention demonstrates that the combination of novel technologies, which allow the interrogation of larger panels, and bladder cancer-specific epigenetic biomarkers can be utilized to detect bladder cancer, allowing a reduction in the number of cystoscopies and consequently improve the quality of life for the patients as well as decrease health care expenditure. Furthermore, the utility of large panel assays allows for the potential of multiple clinical parameters to be evaluated from within the same data. For example, the stratification of tumour grade, recurrence or progression of non-

muscular invasive disease, the likely response to therapy for muscular invasive disease or the differential diagnosis of multiple conditions.

Example 5: Optimisation of DNA sample handling and processing

Optimisation studies were performed in order to maximise DNA yield from samples of patient urine. DNA was incubated with proteinase K for either 1 hr at 56 °C or for 48 hours at 21 °C. Incubation was performed in the presence of RNAse A, 100 mg/mL. An increase in both amount and purity of the DNA was observed with the extended incubation protocol (Figure 8D).

The UCL home urine collection kit was compared with a commercially available kit (Norgen, Cat# 18124, https://norgenbiotek.com/display-product.php?ID=424). The standard UCL urine collection tubes contain 70mg/ml of Stabilur™ urinary preservative. This study was performed on urine from healthy volunteers; half of each sample was treated with the UCL standard method and the other half with Norgen preservative.

Little difference was observed in integrity of DNA from urine preserved with the two methods. No difference in DNA purity or yield was observed between first void samples and samples voided at other times (for concentration vs. time $c^1_{(1)}=0.255$, $p=0.614$ and for purity vs. time $c^1_{(1)}=1.046$, $p=0.306$).

However, an apparent increase was noted in DNA yield over 8 days in urine treated with Norgen preservative, whereas DNA from urine treated with the UCL standard protocol was of similar amount irrespective of time.

An apparent increase in urine DNA yield with the Norgen system was found to be attributable to bacterial growth over time whereas growth was effectively inhibited using the UCL established protocol (Figure 8E).

Example 6: additional validation of detection panel using high throughput technology

96 urine samples from 32 confirmed bladder cancer cases and 64 non-cancer cases (Validation Cohort 2) were analysed as described above. Sequencing of the 150 UroMark loci described above was performed following bisulphite conversion using the RainDance ThunderStorm® System. Statistical analysis was performed as described above, and the resulting ROC plot is shown in Figure 9 (AUC=0.96, Sensitivity= 0.97, Specificity=0.97, NVP= 0.98).

In a yet further study, 92 urine samples from a cohort of haematuria and known cancer samples (Validation Cohort 2) were analysed as described above. This cohort consisted of 27 confirmed cancer cases and 65 non-cancer cases. Again, sequencing of the 150 UroMark loci described above was performed following bisulphite conversion using the RainDance ThunderStorm® System. Statistical analysis was performed as described above, and the resulting ROC plot is shown in Figure 10 (AUC= 0.955, Sensitivity= 0.98, Specificity=0.97, NPV= 0.97).

Table 1

Table 1 below provides a list of 150 MVPs (CpGs) as used in the methods described herein. Provided for each CpG is the Illumina Identifier (Ilmn ID), the chromosome number (CHR) and chromosome position (MAPINFO) specifying the location of the CpG in the human genome, the gene name (if available), a forward sequence encompassing each CpG and a corresponding SEQ ID number. The cytosine of the CG dinucleotide motif subject to modification is identified in bold and in square brackets in each sequence. The CpGs are listed in rank order from 1 to 150. The rank order is in respect to the number of tumours in which a given locus is methylated. Thus, the MVP corresponding to SEQ ID NO: 1 is methylated in the largest number of tumours, whilst the MVP corresponding to SEQ ID NO: 150 is methylated in the lowest number of tumours.

MVP No.	Ilmn ID	CHR	MAPINFO	UCSC_REFGENE_NAME	FORWARD SEQUENCE	SEQ ID NO.
1	cg25522366	2	63281139	OTX1	TTGCGAAGGCCGAGATCTGGGCCTGCCAGGGGCCTGCCCGAGTCCTCTATCGCGGGTCCA**[CG]**TGGCCACCAATGACCCGCGGCGCCCCCGCGTGTCCCCGCAGCCACTCCGCGGAAGCAGCG	1
2	cg20302133	1	111217194	KCNA3	CCTGCTCGCCCGCGGCCGGCAGTGAGGGCGGCAGCGGCTCGTAGCGGTCGCAGCCGCCGC**[CG]**CCACAGCCGCCTTGAGGCGGGGCCCCTCCACCATCGGCCACCTCCGGCTCCAGCAGGTGG	2
3	cg13046832	14	29254680	C14orf23	CCGGGACCCCGTCCCTCTTTCCCCTTCAGTCTTCAGGGAGGGGGAGGCGCTCCGCATTAG**[CG]**GGGCAGTTCAGCAACCCCGACCCCACCCGCGTGGCTCCAGGCCCAGGGGTCCGTTCACTT	3
4	cg17945976	2	66667433	MEIS1	GTGGAGAGCTCACTCTGCAGAATAAAATCAAGAACACCACGGTTGTAGCTGCTGTGGACT**[CG]**GCGGTCTGAGCAGCCTCCCCGAGAGCCGTAGTTGCTAGTAGAAGTAAGATTGAGCTCTGG	4
5	cg21472506	2	63283967	OTX1	CTCCAAATCAAAACCACTAAGAGTTCCTCCCGCGCAGACTGCTGCCCCTTCAGCTGCCCT**[CG]**ATTTTGCTCCACGCCTGCCGGCCAGAGCCTCCCGGCGTTTCTTCCGCCCCAGCGGAGTGC	5
6	cg26492446	20	61638574	BHLHE23	GGAGCGCTAGTACTTGCTTCTCGACTCCCCGGCCGCCGCCTCCGGCCGCCCCGGGGATTC**[CG]**CCACCAAACGCACGCGTCCCAGGTGGGCACCCGCCTCGGTCCGTCTTTGAGTCTGACCCT	6
7	cg11142705	3	42306974	CCK	GCCTCTCTTGACGCAGCTGTAAAATGCGGATGACACCATCTGGTTTTGCTCAGAGGAATC**[CG]**GTTTGGGAAAGGGATGTGTTTTCTTCCCGGGCCAAGTTACCACCACCCGCGGCGCCCACT	7
8	cg26970841	16	85932666	IRF8	TGGCTGCCGGGGGCGGGAAAGTGATTTCTCGGAAAGCAGAGCACTTCGAAGAAGGCGGGC**[CG]**CGCGAGCCAAGCTGACGCTATTGGTCGGTGTGGCCGTCGCTCTGCGCACCGCCCGTCCCC	8
9	cg23497016	2	66666684	MEIS1	GCAGTGGAGGGGACGAGGGCTTGTCGGGTGGGAAACTTAATTCAAAATGGCTGCTGGAAA**[CG]**CTTGGGTTTTATTCGTAGCAAATGTTGCCAATTTCTCCGGCCAGATACGCTAAACCGATC	9
10	cg19178853	4	174450408	NBLA00301; HAND2	GGCGGCGGCGGCAGCTGCGGCGGCGGCGGCGGCAAACGGGTAGCCCTCGTGGTGCACCAC**[CG]**GGTGGTGGGGAAAACCACCTACCAGACTCATTTCGCCCTCCGCGCCCCTCCACGCGCCCC	10
11	cg14428146	8	23563925	NKX2-6	TCTCCAAGACTGCCTCACAGGGACCCCCAGGAGGCTCCGAACCATCCAGCTTTCTGTCAC**[CG]**CCGCCGCCACCAGCGTTGTGAACCTCTGACCCTCGCGGCTCTGCGTCCATTCTCAGGTAC	11

12	cg15887264	4	147561775	POU4F2	TCCCACCGCGGAGAAGCTCACCAAGCCTGAACTCTTCAATGGCGCGGAGAAGAAGCGCAAG[CG]CACGTCCATCGCTGCGCCAGAAGCGCTCGCTCGAAGCCTACTTTGCCATTCAGCCTCG	12
13	cg06392169	6	391936	IRF4	CCAGAGGTTCGACCTCCAGGGCAGCGCAGGGTACCCCGGCTTCGGAGCGGGAAAGGGAGCGG[CG]CCCCGTCCTGGAGCTCCGACTCCCACCCCATCTCGCTGAGCCGGAGGCGCTGGTTTGGG	13
14	cg11433522	2	66667549	MEIS1	CTCTGGGCTTGTTTTCAGCTTATTTAATTCACACTGAAATGTTTCTGCGCGGGACGAACT[CG]GTGTCACCGGGTCCCTCCCGGAGGGTTACTTCCTGCCCCCGACAGTGTAATGAGGCAAGA	14
15	cg01791874	5	16180055	Mar-11	CGGGGGTGACGGGGGCGCGGGCCGGGGGTGGGCTGGGCGCGGATCAGTGGGACGGAGTT[CG]GGGTTCGGCTCCGAGCGGGCGGCTGGAAGTGGGGGATCCCTCAGCCGCCTCCACGGGCC	15
16	cg23180938	5	115152485	CDO1	GGAGGACGAGGCGGAGAGCCACCCAAGAAAGGTGGCGGAGCGCGGGAGACCTGCGGGCA[CG]GCTCACGCGACATCCCCGGCTTCCCCGGGCTCCGCGCCTTCCCAAGAGCCCCGTTGTCT	16
17	cg26990102	2	119606746	EN1	ACTTGTCTCCGGATTCGTTTTGGAGGACCAGTATCCGACTATGTCTGAAGCCTGCGGTGGC[CG]AGAAAGGCGTCAGCAAACGAGGCCTTCGGGGGTGTCATCTTAGAGCTCCAGCGGACCTCC	17
18	cg11595545	1	111217497	KCNA3;KCNA3	GGGGAAGAGGCCGGCAGCGGTGAGGCCAGGTCGCTCCTCCTCGCGCTCCCCGCCCTTTCGC[CG]CCTCCGCCCCGAGCCGACGCCTGTTGCAGCCAAAGCCGCGATGCTCTGTCTGG	18
19	cg00986824	12	75501465	KCNC2;KCNC2	GTCGCGCTAGTAATTGAGCACATAGGCGAAGACGCCCGGGTGTCCGGTCGAAGAAGAAACTCG[CG]GCCGCCACCGGGATGGTCGCTGGCCCTGCCGCCGCGGGAACTGCACGTTGCCCGCGCCGCC	19
20	cg09462924	2	66666470	MEIS1	GGCTCTCCGAGGGCCTTGGGGTTGGGATCCCTAGGTGCAGCCCGTTGACAGTCGGCCCCA[CG]GCCATGGACGTCCTTTCCCCAAGTTAGCTGAGCGCCTGCCACCGAGATCCCCCGAGCCTG	20
21	cg07790515	1	237205174	RYR2	AGATTAGCCACCCCGTGCGCCCAGGTGAAAGATATCATTCTTCCGTGCGATCCGAAGTGC[CG]TGGAAGTTAGTGCCCTAGCCCAGTCCAGGAGGAAGGGCGTCGTGCCGGCGGTTTTAAGC	21
22	cg06785999	14	61097564	SIX6;SIX6	GCCGAGCCCGAACCCCAAGCCGCCGGAGCCAGCAACTCTCTCCAGTCGGGGTCGTCCGCTCC[CG]GCCGTTGAGCCACCGCCGCCACCCGGTAGTGTGTCCCGCTGCCCCAATCCGCCTCATCAA	22
23	cg04550737	1	119530600	TBX15	GAGACTGCGGCTCGGCGGGTCTCTCCACCCTCCCCCTGCGTCCTCCTCGCCCTCCTCTGC[CG]GATCCGACCTGCGCCCCTACGCTGGCCCAGCTGCTAGGAACTAGCGCCCCGAGCGCCGCC	23
24	cg19978181	5	76923988		TTACTTGGGAGGTCCCGGGTATCTGAAGCGGATCCCGGGTCTGGGGACATGAAGGGGCGC[CG]TGGCCTTAGGGAAGGCCCCAAAGAGGCTAGGCCCCGGAGGAGGCAAGAGCCGCGGCCTA	24
25	cg23065934	5	16180266	Mar-11	GGGGAGAGGCCGCGGGTCCAGAAACCGTTACTGGATGGCCGGTGGGATGTGGCGCGGGC[CG]GGTGGGGCCGCGACAGTCTGAGCCCAGACCCGCGTGGGCTTAAGGGTGCGCGAGCCGGGTG	25
26	cg14470895	5	115152431	CDO1	CGGGCCCCTTTTAAGCGCTTGGAGTCACTAGCGAATGTACCAACGGCCCTCGGAGGGAGGA[CG]AGGCGGAGAGCCACCCAAGAAAGGTGGCGGGAGGCGGGGAGACCCTGCGGGCACGGCTCAC	26
27	cg11036833	5	115152494	CDO1	GGCGGAGAGCCACCCAAGAAAGGTGGCGGAGGCGGGGAGACCCTGCGGGCACGGCTCACG[CG]CACATCCCCGGCTTCCCCGGGCTCCGCGCCTTCCCAAGAGCCCCGTTGTCTCCGGCGTCC	27
28	cg23727983	11	125774082	DDX25;PUS3	ACACGGGGCGTAAAGCGCGGCGGGGAGTCCGGGGGGCTCCCGCCTGGAGGGCTGTGTGTGAG[CG]GCGGGCCGCGGGGGCGGCGCGGGGGCGCTCTCCACTCTGCGGAAGCTGCCCCCTCTGCCC	28
29	cg16732516	1	50886782	DMRTA2	CGCCGCACACTGAACCGAAGACCTCGTAGGCGGGCCTCGGGGGGATGATGCCGTTGGCGG5[CG]GCCAGCGCCAGCCCCTCGCAGTGCCGTAGAGCAGCTGCAGCTCGAGCGCCTCGTTCTCC	29
30	cg15422098	10	15761881	ITGA8	CCCTGCGGGGCAAGGGGGGCTGGTGGAATCTGGCGGTCCCCAGCTGCCCGTGTCCCGAGTT[CG]GTTGCGCTCGGGCGCACCCGTGGTGACAGTGCCCGGGCTCTGCTCCCACCGGCCCCGCCC	30
31	cg13520034	8	72755052	MSC	AAAAAGAACGTGAGATATTAGAGAGAAACGATTGTCTCAAACCGAAACAGCTCTCCTACG[CG]AACCCCAGATATTCCTGACTTGGAGTAGCTAAGATTTTATCAGCATTCTGGGAAATTTGCA	31
32	cg09424526	2	223163809	PAX3;PAX3;PAX3;CCDC140;PAX3;PAX3;PAX3;PAX3	CTTTTACCCAAAGCTTGGTCAGGAGCCCTGAGCTGCGATTGGCCGACGGGTAGACCGTCC[CG]GGTGGCGGAGACACGCGCTGATTGGGCAACAGCGACCACTTCTCCTTCCCATCTCTGGTG	32
33	cg09813525	13	53420386	PCDH8;PCDH8	GGCGTGTCCCATTGCAGCACCGACCCGGAGCACACCCGAGCCGAGAGCCAGGCGGGGCGGGAA[CG]CTCCGGGCTTCCTGCACTGGCAGGCGCAGCCGGCCCACGCCCCTGCTGTTACCAC	33
34	cg18082337	3	147113726	ZIC4;ZIC4;ZIC4	GATTTTCTGATCTAGCAAAGACCTTCCCACACCCCGGGAAAGGACAAGGGAAGGGCTTCT[CG]CCCGTGTGCACGCGGATGTGATTTACAAGTTTGTATTTGGCTTTGAAGGGCTTTCCCTGG	34
35	cg27655158	4	1396593	C4	GTGAAGTTCCGCAAAGTGCTTCGCAGCGGGCCGCGGGAGCCACCTGCCCGGCCCC[CG]ACGCGCATGGTCATTTATAAATTTAAAACTCTTCGTAGCAACCGGTTATGTACAGAGTC	35

36	cg26802289	7	20817859		GTGGCGCTCGCTGTCCGTGCAGCGGGCGGAGGCGGCCGCGGTGCCTTTGTGTGCGGTGGG[CG]CGGCGATGGGCTGCTCCGGCCCGCAGCCCGAGGGGAGCGGCGGGCCCGGGCCGGGCCAG	36
37	cg18074954	18	44787492		CCTGTCCGCGGCCGCCCGGCGGTCTCCGGCCCCCTGCGCGAGGTCCTCCCTCTCGTCGCCTCGC[CG]CTCCACGGCTCCTCCGCGGCTCTGCCTCCCGCGGCCGCTCCCCTGCGCGGGATCA	37
38	cg25947619	15	86233220	AKAP13;AKAP13;AKAP13	TTACGATTTCACTAATGGCAGCTCAAAGCTGCTGAGCTCTGCCCTGACGGCATGCGCCCCC[CG]CTTCTGTGAGAGCACGCTTCCTGGCATGGGGGTCGAATCATGTCCATAAAATGGGCTCGG	38
39	cg15384598	14	24045549	JPH4;JPH4	TGCCTCCCTCGTCGCCCGGCAAGGGCAGGGGCGGGGGGTGGCGTCGGGGGGTCGCTGTGGCG[CG]GAATCCAGGGAGGTGCGGCGGGGCGAGCGCAGCAGCGCCGCCTGATGGTAGGGCACACTC	39
40	cg16178503	2	66667101	MEIS1	CGAGCCGGGGGGTGGCGGGCGGGGACGTCTGCTCGTCAGTCATTCAATGAAGATATAGC[CG]TGTTCGCCAAACAGGTCAGCAAAATAGATGTTAAAAAGTAAAAGAAACAAAAAGAGAGCC	40
41	cg23241781	14	70653964	SLC8A3;SLC8A3;SLC8A3;SLC8A3	TTTAATTCTTCTGCCCGGGAGACCTAACCGCCGAATGCGTTCGGAGGTATGTTTTTTAAT[CG]CCAGGAAAGGGGGAGAGAGAGAGAGAGAGAGAGAGGAAAGGAGGAGAGA	41
42	cg01419831	2	162283705		AGGAGGAGCACGTGCGGTCGGTGGGGCGCAGGGCCGGGGAGCCAGGGGGGTGCCGGGGCT[CG]AGGCCGCGGGGCCGCGGGGGCCGCTCTGCTCTCCGCCCAGGCTGGGCCACTGGAGCGC	42
43	cg17039236	18	77548049		AGGCGAAGGCCGCCCCGGGAGAGCGGGGTCCCGGGAGAGCGGGGTCCCGGCTGTGGGGGA[CG]CGGGCCGAGGCTGTCGCGAAGCCGCTGACGGCCGAGGCGCTCCCGGTTTTCGCGGCGCAC	43
44	cg21901718	5	16180076	Mar-11	GCGGGGTGGGCTGGGGCCGCGGATCAGTGGGACGGAGTTCGGGGTTCGGCTCCGAGCGGG[CG]GGCTGGAAGTGGGGGATCCCTCAGCCGCCTCCACGGGCCGGCCCCGCGCTCACGTCGGTT	44
45	cg25191628	1	50886949	DMRTA2	CGCTGGCGCTCCCGCGATGAAGCGTGCACTTGGCGCACAGGCAGTCCTTCCAGCGACAGTAG[CG]TTTGTGGCCCTTGAGGGCCGACACCACGCCATGGTTGCGACAGCGGCGCGCACTTGGGGGT	45
46	cg27277463	12	62585031	FAM19A2	CCGCGGTCCTGCAGTTGCCGCTCCGGTCCCCAGCGCTGGCCGGCGAGCCGCGGCT[CG]CACCTACCTGCAGCCCCGCTTCCCGGTGGCGGCAACACCTAGCGATGCTCCTGCAGCTTT	46
47	cg05311410	7	27225523	HOXA11AS;HOXA11	CTGGTGGCTTGTCGATTTGCACGGTGACTTGATTACACTCTCTCATTCATGGTCACTTC[CG]AAGCGCTTTAGTGCCTTCGTCCCTAAACCGCCAACAGCCAGAACGGCTTCTCCCCGCGG	47
48	cg23229261	2	63284066	OTX1	TTTCTTCCGCCCCAGCGGAGTGCGCTGGGGCGCGCCAGGGCTAGGCCCCGCCGGAGGAGCGG[CG]TCCCCAGCCTTCCGCGCACAGACGCGCATCCCGCCCCCGCCCTGCGCTGGACTGGTTCAAG	48
49	cg10397440	8	57359258	PENK;PENK;PENK	CCCGGGGCTTAACGGCTGCTGGAGCCACTTTATAATTAGCCCCCAAACCGAAGGAGGCGCGG[CG]CGCCCCAATCGCCGGCGGGCTGCAGCTGACGCAGGCCCTACGCCAGCCCCGCGCCGACGC	49
50	cg18326021	10	106401479	SORCS3	GGCGAGCCGCGCGGGGCCGGGGCATCCCAGCTCCTGCCAAGCTTGGCGCGCGCGAGGAGGAGT[CG]CCGGGCGACGCCCCCAATCACCCAGGAACGCGGGGACGCCTGGCGCCACTGCTCCGGCCGA	50
51	cg06533244	6	27258460		CACAGGTTTCCGTGGTGTAGTGGTTATCACATTCGCCTTACACGCGAAAGGTCCTCGGGT[CG]AAACCGAGCGGAAACAACTTGCAATTTTTCGGGGTGTTCTGTTTTCCAAGATTCCCTTA	51
52	cg03078363	12	54408664		GTAAAAACCCGTTTTATGGGGGAACGTAATTGTGAGCGGGATGCGCTCTCTTTAGAATCG[CG]TCCTCCCAAATGCTCCCGCCGTTCCCATTACCGGAATGGGGGACCATTCGGCTGCTGCAGAT	52
53	cg01796166	14	95239586		GGCCCCGCATCGCCTGGCGCAATTGGAATAACAAATGCGACGCACGCAAATTGTCCACCT[CG]TGTTGCTAAGCGATTGTTTGTCGGCCCCGCACGCACAGCTCAGCATGGGAGGACCGCGAG	53
54	cg00339556	5	16180048	Mar-11	GGCGCGGCGCGGGGGTGACGGGGCGCGGGCGCGGGTGGGCTGGGGGCGCGATCAGTGGGGA[CG]GAGTTCGGGGTTCGGCTCCGAGCGGGCGGGCTGGAAGTGGGGGATCCCTCAGCCGCCTCC	54
55	cg03978375	16	85932668	IRF8	GCTGCCGGGGGGCGGGAAAGTGATTTCTCGGAAAGCAGACGACTTCGAAGAAGGCGGGCCG[CG]CGAGCCAAGCTGACGCTATTGGTCGGTGTGGCCGTCGCTCTGCGCACCGGCCCGTCCCCCC	55
56	cg26013553	1	111217406	KCNA3	CTGCTCGCTGGGCGCTGAGGAGGGTGGGCCGCGGTGGCGGGCGTGAGGGCGGCGCGCGGCGAG[CG]CAGAAGGCTGAGGCGCCTCGTCCATGCGCGCGGGGAAGAGGCGGCGCAGCGGTGAGGCCAGGTC	56
57	cg09734791	8	72756155	MSC	ACTGCTTGCACTCTGCGGCTGAGCCCTTGGCCGGGAGGGGCTTCTTGCCACCACCGCCCCG[CG]CTACCACCTGCGCCGCCGCCCCCAGCCACACGGGGCCGCTTCCTCTTGCAGCCTTCCGCG	57

58	cg12040830	11	112833773	NCAM1;NCAM1;NCAM1	GGCGGGCCAGGGAGCACCCAGTGCGCCCCCTCCGCGGGCGGCACAAGAGCAGCGCTCGGC[CG]CCGCCTCCAGCCAACTCGGGTCCCTCCCACGGCGACCAATCAGTGCGAAGCTGGCTGGGC	58
59	cg18617005	5	140787504	PCDHGA4;PCDHGA6;PCDHGA9;PCDHGA1;PCDHGA8;PCDHGA5;PCDHGB1;PCDHGB6;PCDHGB4;PCDHGA3;PCDHGA2;PCDHGA7;PCDHGB2;PCDHGB6;PCDHGB5;PCDHGB3	ATTTCGGAGACCGAATTCAAAATGAAAAACCGGGCTGCTGTCCCGCACGGAGCCTCTGGG[CG]CCGCTGTCGGCCAGTGCAGAGCAAGGCTGAGCCGGGGATCCGTCAGCCTCTGGCCTGG	59
60	cg19597382	4	147559423	POU4F2	ACCTCCGCTCGTATTGGGCTGGGAGTTCAGAGCCGCGCGCAGAACCCGGTTGGCCGCAA[CG]TCTGTGTTCTCAGCGGTGGCCGGGAACCTGGGATCAGGGTCACCTGAGCTGACGGGGTGG	60
61	cg05134015	14	36973365	SFTA3	CCGGAGGTTCGATGGCCGCCGGGCCAGTGCGGGCTCAGAGGAAGACCCTGCAAAAAAGAG[CG]CTCGCCCCCACCCCTGGAGCCGACCCTGCGCAGTAGGGCCGCAGCCGGTCCCCGCGGCA	61
62	cg02700891	3	129693586	TRH;TRH	GGCGCAGATATAAGCGGCGGCCCATCTGAAGAGGGCTCGGCAGGCGCCCGGGGTCCTCAG[CG]CTGCAGACTCCTGACCTGCCGACTGCGGATCCCGAGTCCCGGATCCCGGACCCATCCTG	62
63	cg13476854	1	119549263		GTCGGATCTCTAAATTATCTAATCTGGCGGCTGCGTACGACTCAGGGAAAGCCCTGGCCGG[CG]AGCTTTTTCACCAGGCTTGAGCTCAGCAGCCGGGCCCGCAGTGTTGCCGCCAGTGGGGAG	63
64	cg00017221	1	149719536		GTCTCTCCACGCGCTGCCGCCTAGCAAAGGCGCATCTTTAGGTCGGTAGTGAGGTGCGGC[CG]GGACGCTGCAACTCGCTCCGGGACTTGTAAACCTGGCAGGTGTTCGAAGAGGGCCACTGG	64
65	cg05310764	6	27513479		CACTCAACCATTATAAGTTCACCCAGCGTCAGCGATGGCGTAGGTAGGTAGTCGTGGC[CG]AGTGGTTAAGGCGATGGACTTGAAATCCATTGGGGTTTCCCCGCGCAGGTTCGAATCCTG	65
66	cg04495995	7	19147157		GTGTAGAGAACAACAGTCGCTCCTTAGATATTACTCCAGGACGGAAAACCTGATTGCAAAC[CG]CTGTTCCTTCGAAACTTGGAAAACCCGGAACAGAAAACTCCCGCCCCAGCCAATTTTAGCT	66
67	cg01642521	11	20618250		ATTCAGACCGAATGGCTGCGCGGTGATGGATGCGGATTTACGGCCTCCTTGGCTGCGGCG[CG]CTGGGCCTGATTATCACTATAAACAGGCGTCCGCGGAGGGCGGGGGCGGAGGCCCGCGTC	67
68	cg24504927	12	85667353		TTAACCACAGAGTTGTTCTTGATTGTAAGGGACTTCGGCCCACTTGGTTGAAGTGGAGAGGC[CG]GTCCTCATTCCAGACGGTCCCGCACGGCAGTCGCTCATGGCTCCCTCCAGGCCGGGAGCCA	68
69	cg03276408	13	112712475		GCAGGGACCCGGCTGGCCGCCTCCCCTTGCAGGAAACAGGTGTTTGAACGCGATAGCGGC[CG]CCAGTCAACTAAGGCATTAAAAGCTCGCTTTATAACATCGATTTCCTGGAGTGCGGTGGG	69
70	cg27501878	20	61638588	BHLHE23	TGCTTCTCGACTCCCCGGCCGCCGCCTCCGGCCGCCCCGGGGATTCCGCCACCAAACGCA[CG]CGTCCCAGGTGGGCACCCGCCTCGGTCCGTCTTTGAGTCTGACCCTAGCGCCAAGAGTCCC	70
71	cg07495363	2	198651076	BOLL;BOLL	GGACTCCGAGCCGGGGCGTCTCAGGGGCAGAGCGCACGGCACAGCGGGGGCGGGCGTGGGGG[CG]TGCGGAGCGAGGGCTCGGTTCTGGGACCCCTCTCGCCTTCTCTCAGACGGCTGGGGAGGGA	71
72	cg06750832	1	111217691	KCNA3	CTGCGGGGCGCGCGCCGCGCCTCCGCGTCCCCTTAGGATTCCCGCCCACCGCGCGGGCGCGCG[CG]TCCCGCTCTCGGGGGCAGCCGCCGGGCCTGCATTTCTTGCAGCCCTCAAGGCCCCTCGGT	72
73	cg22474464	20	21492914	NKX2-2	GGAACCAGATCTTGACCTGCGTGGGCGTGAGGCGGATGAGGCTGGCCAGGTGTTCGCGCT[CG]GGCGCCGACAGGTACCGCTGCTGCCGAAAGCGCCGCTCCAGCTCGTAGGTCTGCGCCTTG	73
74	cg20146541	1	248020697	TRIM58	ACTGCGGCCACAGCTTCTGCCTCAGGTGCATCTCCGAGTTCTGCGAGAAGTCGGACGGCG[CG]CAGGGCGGCGTCTACGCCTGTCCGCAGTGCCGGGGCCCCTTCCGGCCCTCGGGCTTTCGC	74
75	cg02864844	7	149917263		GGATGCAGACGCACCAGCGGCTGCTCACACTCCCTCCACAAACCTGCCGGAGTCTCCACT[CG]CCCGCCAACTGTAGCCTCCATCTGCGCCCCACGCCCCCGCACAAGCCCCCTCGTCGCGGG	75
76	cg12323723	X	136656581		GGCGCAAAAGGGGCCGCCCCCGTGCCGGGAACAGACTTTGAAGTGGGTTTTTAGCGCGCA[CG]TGTGAGAGCCGGGCCAGGGCCGGGAGCGGGGACCCGCTGGGAGGAAAGAGGAGGCTCCGGC	76

77	cg14996220	12	85673270	ALX1	CTCCCACCCATCGCCACCGCGTCCACACTCCGCCGCATTCCAGAGAAAAAGAAAAGGCCA[CG]GCCTCGTAAAGCTCCCAGCTGGCCCGGACCCCGAGCTCCTCCCGGTTAGAAGCCGAAGCG	77
78	cg23151000	2	38302892	CYP1B1	GATTTCCTTTAAAGTACCTACCACGCCACCCGCTACCTGTAATAATCCATCTGAAGAGGT[CG]CCGGGCAGCGCCTCGGCAGACAGACTGACCTGGGGGAGGTGCGGTTTCCAGTGGCGCGG	78
79	cg00963169	1	50513927	ELAVL4	CGGCGCAGGCCCCGCACCCCCGACTCTGCCCGCCCTCTGTTACGGACACCCGCTGGGCCA[CG]TGGTCGCGACTGGCTTCTCCCAGCGGCCAGCCTGGCCACCCGACTCCCAGGGAGGGGGA	79
80	cg13320291	10	11803070	GFRA1;GFRA1;GFRA1	TCGGCCCTCCGCCCCGGGGCATCCTGGCCTGAGCAACGACCCGGGCTCCCGGGCCACCCC[CG]GCTCCAGCCACCCGCTCCGCCCGGCTGAAACTCAGGCGCTTTCCGAGGAGAAGTGCGGCG	80
81	cg13352750	7	27225123	HOXA11AS;HOXA11	TTCTCACCGAAAGCACGTAATCGCCGGTGTAACTCATGTTGGCTGGGGGGCCTCCCGGCG[CG]CGCGGAGAGGCTGGGGTGCGCCCCCATGCAGCATGCTTGTGCTCAATTGCAGGGTCCTCG	81
82	cg26521404	7	27204981	HOXA9	GCTGAAGTCGGGGTGCTCGGCCAGCGTCGCCGCCTGCCGGGGAGGCTGGCCCAGGGTCCC[CG]GCGCATAGCGGCCAACGCTCAGCTCATCCGCGGCGTCGGCGCCCAGCAGGAACGAGTCCA	82
83	cg07573209	15	76630095	ISL2	GGGTTGGGTTGGGGCTGGAGTAGCCGAGGCCGGCCTGGGTCCGGCAGTCAGGCCTGACG[CG]GCCCCGCCCCCTTCCCCGGCAGAGAAGCCCGGGACGGCCATGTGCGTGGGCTGCGGGAGT	83
84	cg10835584	18	55108852	ONECUT2	GCGCTTCCGGCGATCCGCCTGGGCGGCTGGGTCCGCGAAGCCAATGCGCTGAACGGTGCC[CG]AGTCTTCCTAACTATCCTGTGCTTGGCCGTTGCCACTGGGCCCTGGTGACTAAGCCCAAG	84
85	cg09474331	19	54926805	TTYH1;TTYH1	TACCGGCCCTCAGCTTGGGTGCATCTCCTCCACCAGCTGCCCCGCGCCGACTTCCAGCTC[CG]CCCGGTGCCCAGCGTTTTCGCGCCCCAAGAGCAGGAATACCAGCAGGTGGGACCGGGCGC	85
86	cg20870512	7	1272515	UNCX	GTCCCAAATTCCCCGGGCCGCGGGCTAATTATCGGGAGCTTGATGTTGATAAGTAAAGCGC[CG]GAGTGCGGGGCGAAGCATGTGTGGGGCTCCCTGTCTCCGGCGCCGCCGCCGCGC	86
87	cg21678445	18	22930283	ZNF521	GAGGGGCGGTCAACTTCTGGGAATGGCCAAGAGGGGTGCTCTGAGGCCCGAGCCGGGGTC[CG]GTGCCCGCGGCCGGCGGGGTCTGTTTACTCCGGCGACGTGCGGAGAGGACCCAAGC	87
88	cg05560435	5	172671526		CGCCGCCTTTCCCTCCTCGCCTCTTCCTTCCTCCAACCTGCTGTGC[CG]TTACCGACAGCCAAGTTTCCACGCCCGGCGGAGCGCATTGTGAACAGGAGCTGACAAATT	88
89	cg08382226	6	108440339		CCGACTCGGGATGACAATTGACGGGGATCAAGGGATTGCCCATTCTGTGCCTGTAAGAAC[CG]ATTCGTGCCAGAGAAACTCATCAAGTGGAGGCGGAGAATAAAGACCGTTCGGGGGTAAAT	89
90	cg25510609	15	85233236	AKAP13;AKAP13;AKAP13	GGCAGCTCAAAGCTGCTGAGCTCTGCCCTGACGGCATGCGCCCCGCTTCTGTGAGAGCA[CG]CTTCCTGGCATGGGGGTCGAATCATGTCGATAAAATGGGCTCGGTTTCATGAAGTACCCC	90
91	cg19650157	13	78493297	EDNRB;EDNRB	GAAAGCCGCTACTCCCTGGCTGGCTGAGCTACAGCTCCCGCAGCGCGCCCAGGAGTGCGC[CG]GAGATTCGGAAAACCCGCAGAGACTTCTCAAGTCAGCAGGAACTTGGAAACCGCTGTTCCC	91
92	cg01283246	5	135266135	FBXL21;FBXL21	CTCGCTCGGGCGTGTTCCTGCGCCGACCGGACGGCCGGACTCCAGCACCTTGGCCCGGCC[CG]CGAACGCTGAGCACGCGCGGAAACCCTTTAAAGGTAGCACATTTTTCGGGTGTCGCGGGG	92
93	cg20872937	18	74961968	GALR1	CTGGAAAAGCCGGGAGGGAGTCGGAGGCGCCAGCCCACTGGGGAGGGTGGCGCTGGGCGCG[CG]GGATGCGCGGGGAGCCTTCTCTGCAGGAGCCGCACAGTGCACTGCTGCGCGCTGGGCAGT	93
94	cg03698009	7	27204349	HOXA9	TCTCCCGTAGCCCTGCGGGGCCGCTCTTCACTGCTCTCCAGACTTGGGGCCCTATCTGAGG[CG]TCCCAAACACCAACTTCTGGCTCCTGGCCCCAACTCGAGAGGCTTCCAGCGAGGACGAAG	94
95	cg01423964	1	111217575	KCNA3;KCNA3	GAGCCCACCGCCTGTTGCAGCCAAAGCCGCGATGCTCTGTCTGGGTCTGGCGCGGTCAGC[CG]GGCTCCCGCACGGGGACGCCTCCTCCCTCCTTCTCGCGCTCTCCGCCCCCTCCCCTGCGG	95
96	cg19923650	5	172659730	NKX2-5;NKX2-5;NKX2-5	CGAAGTTGTTGTTGGCGGCGGCAGTGGCCGGCTGCGCTGGGGAAGGCCCGGCGGGGTAAG[CG]GCAGTGCAGCTGTAGCCAGGGCTGCAGGCCGCGCGCCGTAACCCGGATAGGCGGGGTAG	96
97	cg13324546	8	23564031	NKX2-6	GTCCATTCTCAGGTACTGAAAGTTTTCGGGGCTCTTCCGCACCCGCGGATGTGGCGAAGC[CG]CGGGGCAGCTCCGCTCGCGCTCCAGTCGCAGGATGTCCTTGACCGAGAAGGGGGTGGAGG	97
98	cg27357571	21	34398226	OLIG2	CGCCCCCGCCCCCGCCACCGCTGCCGCCGTCGCCGCTGCCACCGGGCTATAAAAACCGGC[CG]AGCCCCTAAAGGTGCGGATGCTTATTATAGATGCGACGCCGACACCAGCGCCCGGTGCCAGG	98
99	cg08448701	20	21686282	PAX1	CAGTGACGGGAAACCAATGAGCTGCCAACTCGCGCGTCTCCGGCGTGACTGCCGAGATTGA[CG]TGGAGGACACGTCAAATTGATTCCCGCACGCTGCAGCCTCCCGGTCAGACGAATTTCTCC	99
100	cg16002355	4	111544387	PITX2;PITX2;PITX2	CTCCGCCTCCTCCCAGACCCTTCTCCGGGTGCGACTGAGTGGCTCCGCACCAATCAGGA[CG]CCCCCGAGCCGCGGTGGAGGGACTGTCCTGCCTGCACCTATCAGCAGTGCGGGGCCCGGGCT	100

101	cg05302386	14	52734525	PTGDR;PTGDR	TTAGCACCCGGGGCGCGGGGCCCTCGCCCTTCGCCAGCCTCTCACTCCAGCCCTCTGCTCC[CG]CACGCCATGAAGTCGCCGTTCTACCGCTGCCAGAACACCACCTCTGTGGAAAAAGGCAAC	101
102	cg17964510	2	175199694	SP9	AGGCGCAGCCAGCGGCACTTCAAAGCGGGTGCTCCTCGCACTTAGGCTGAGTTTAGCCGG[CG]GGAGCCTGGAGTCCGCTCGGCACGAGCGCGGGGAGCGCGGGAGCCGCGCGGGACCCAAGCA	102
103	cg24031355	10	22634439	SPAG6;SPAG6;SPAG6	GGCCGCGTTCCGGTTCCGGTAGGTTGCCCGGGAGACGCGGGTACACAGAGAAGCGGCTCC[CG]TCGGAGGCCGAGTCGTCGCCACGATCGCCCCCTTGGTGGACTCGCAGGCCGAGCGGCTTC	103
104	cg24961583	3	27765409		GGGAGCGCATTTTCCGGCTGAGATGTCGGGACTCTGCTTCCCCAACCGAACGCGATCACA[CG]GGAAACTCTTCGCCCACAACAGATGAGATGGCCAAAGGATTGCTGAGTGCGCACACGCAA	104
105	cg02167020	10	102899949		GGACTCATGCAGAAGAGGACATTCCGCAGGTAGGTACAATCCCAGCGCTGGGGCCTGGGGG[CG]TCCGGGGGGCGGCCTTTGAGCTTCCCGGATACCGCTCGCTGCTCCCGGAGCCTGTTCGGC	105
106	cg25026529	1	91183051	BARHL2	GGGGGAGTTAAAAAAAATTTCTGAGAAAACTCGGAACTTGCGCTCCAGGAACGACTGCGCA[CG]TGGCGCGGCGGTGGCGGCGCGGAGGACCCAGGCGAAGGCGAAGGCGAAGGCGCA	106
107	cg05783139	2	198650985	BOLL;BOLL	GACTGCGTCGCCTCGGGTGGCGGTGCGGGCGGCGCTGCAAGCCGGAGAGGGG[CG]CGGGAGGGCGAGTTTCGGCTGTGGCCCTGGGACTCCGAGCCGGGGCGTCTCAGGGGCAGA	107
108	cg25691167	7	19184961	FERD3L	GAGGGCGTCTCGGGGAGGCCAGGGACAGGTCTGCGACGAAGTCCAGCACCGTAGTGTCCA[CG]CAGCTCTCCGGATAGGCCGCCATCGCTTCGGCTTGCCCTCCTCTCATCGGTTTTCCGC	108
109	cg25951981	4	45995743	GABRA4	GTATACCGCTCCACACCCTTTCGTGCCCGCGCGCTGAAGGTTCTGGGGTTCGTATCCGGG[CG]CTTGCGCTGCAAGACTCGGCAAGTTTGTTCCGACTGTAACTCCGGGGATGAGGAACGGGGG	109
110	cg11601252	15	68122139	LBXCOR1	CACAGAGGAGAGGGTTGGGGGCAGCGGAAAATCGGGCAGTCGAGGCAGCCGAACCCGGA[CG]ATGTCCCCCCACCCACCCGAAGGTCGCAGCCTGGGCCGCGTTCTCAGCAGGAGTCGGGC	110
111	cg10698928	8	65290320	MIR124-2	CAGAAGGACCATCTGCGGACTCGTTTTCACTGCTCCCAGCTCCCGAAATCGTTTCTCTGGA[CG]CGGCAGTATCCGCAGCGCATGCACCCTCCTCTGGACTTCCGCAGCCCAGACCTGCGCTTA	111
112	cg25832771	8	72756058	MSC	GGGGGCACCCAGGGCAGGCTGGTCTTGAGCCTGGAGAAGGCTTTGCTCAGCACGCGCATC[CG]GGCACGCTCACGGGCGTTGGCCGCGTTCCGCTGCGACTGCTTGCACTCTGCGGCTGAGCC	112
113	cg18507379	5	140787507	PCDHGA4;PCDHGA6;PCDHGA9;PCDHGA1;PCDHGA8;PCDHGA5;PCDHGB1;PCDHGB6;PCDHGB4;PCDHGA3;PCDHGA2;PCDHGA7;PCDHGB2;PCDHGB6;PCDHGB5;PCDHGB3	TCGGAGACCGAATTCAAAATGAAAAACCGGGCTGCTGTCCCGCACGGAGCCTCTGGGCCC[CG]CTGTCGGCCAGTGCAGAGCAAGCGCTGACGCCGGGGATCCGTCAGCCTCTGGCCTGGGAT	113
114	cg14733048	8	109095782	RSPO2;RSPO2	CTTTCTCCACGGTCACTTCACAGCTAAGATTCTTTCTTTCCGAGCTGTAGAAGGCAGAA[CG]CTCTCGGGAGGACGAAGTGATCCGAAGGGGATGTGGCAAGCGCACTTTCCGATGGAGATGC	114
115	cg09735723	10	106402042	SORCS3	GGGAGGGATCTGTGCTCACTTTCTCCAATACTTGGCTTGGAGGGTCAGTTTTCCTGTTTG[CG]GGGTGCTTGAATTCTTGGATGAGAAAAAGGGCTGACTTGGGGCGGGAGCCGCTGAACAGA	115
116	cg05099508	10	22634432	SPAG6;SPAG6;SPAG6	GCAAGACGGCCGGCGTTCCGGTTCCGGTAGGTTGCCCGGGAGACGCGGGTACACAGAGAAGG[CG]GCTCCCGTCGGAGGCCGAGTCGTCGCCACGATCGCCCCCTTGGTGGACTCGCAGGCCGAG	116
117	cg14638883	6	166582201	T	TCTGAGAAGTGTCCTCCTCGCTCTTATAAAAACAGGACTTGTTGCCGAGGTCAGCGGG[CG]CATCGAGTGTGCCAGGCGTGTGCGTGGTTTCTGCTGTGTCATTGCTTTCACGGAAGGTGG	117
118	cg26673012	7	35293753	TBX20;TBX20	GGAACTACGGACAGTGACCCCTGGCGCTCGCGCTCCCTGCGCCTTAATTTGCTGGCGGCGGG[CG]ATCCCGGAGGCCCGCAGCCAGTCAGCGCGTCTCACGTCACCGCTTCCTGATTCCGCGC	118

119	cg14763548	20	25062447	VSX1;VSX1	GCTCGGGGCCCCTGGGCGGCAGGAACGGCACGTCCGCTAGGAGCAGGCAGGGTGCTCGAG[CG]GCCGCCGCGGCGCTGCGTGCCGAAGCCACAGAGGAGGCCGAGTCCCAGCGGTAGGGCCCCA
120	cg22884556	2	45157296		AATCACTGTAGTTAAAAATGTATGGGATTTTTGCCGTCGGCGCGG[CG]GGCCCAGTGTGGGACTGCGGCTGGGAGCCCGGGCCCTCCGGGAGTGAAGATACCTTTGGA
121	cg21401879	2	45162036		GCGGGCTCCCCCTGGCCACATCCCGGGCCTCTCACACAAGAAGAATAGTTCTGTTTCCGC[CG]TAAACCCCCACACAAAGGCTGCCCGGCCACCGGGTCCCCTGTCCCCCTTCCCGGGGACTT
122	cg09797577	6	28778226		GTAGGCTGGACCAGGAAGGAGACCTGGTTCGTTTCGCCCAGGCTGTCACGGCTTCAAGAG[CG]CCTCTCCGCTATTTCCGTCGCTCGACAGACGGGCTGAGCTCTTTGGAGTGATGTTGGGTT
123	cg20567847	7	155167038		CCGGCGGGGGCTGGGCCGGGCGGGTGGGTTTCTGAGCGCAGCGCTTGGAGCTGGGGGAG[CG]GGAGCAGGGGCGCCCGGCGGCGGGCCGGGACCCGGCTTTTCCGGCTACCCGTGGGCCA
124	cg04828133	7	155259845		GGAGGTTAATAGCGACTGACGACAAAGGGCCAAGGTGCAATTCCTAAAGCCGGGATTCGC[CG]GGTGAGGCAGAAATCAGCCTCCGGCGGAGATGGGTCCCCCCTTCCCGACGCGCCCCTGCAC
125	cg04521510	7	27242044		CTTCGGTTTTCCGGCCAGACCCGGAAAAACGAAAACACAGCTTGGGGAGCCCCCACTAGC[CG]GCGCCTGTGCCAGCTCACCTCTGGCCATGGCGCAGCTGCCGGTGCACACGGCGGCCAAGG
126	cg24304093	11	14926738		TCGGCCTGGCCAGCGCGGGGGGGTCTTCTTCCCAGTCCGCGGAAGCCAGGTCCGGGAGACGG[CG]GGGCTCACTGTGCCTTCGGACCTACACTCTGCTCGGGGGCCGGATTTTCTGCAAAGCGTCC
127	cg24292235	12	8171463		AGCCTAGGATTCGACTTTGAATGGTCCGTTAATGTGGTCGGCAAAACGTGACTCGGTTCAT[CG]GGCGCTCCCTGTAAGCAAGACAAGCACCCACCTGCGGTCAGAGCAGGGGTCCGGCTCGGCG
128	cg25092681	5	16180033	Mar-11	GCGGGAGGGAGGAGCGCGCGCGGCGGGGGTGACGGGGCGCGGGGTGGGCTGGGGGG[CG]CCGATCAGTGGGACGGAGTTCGGGGTTCGGCTCCGAGCGCGGCGGGCTGGAAGTGGGGG AT
129	cg02984614	17	31618409	ACCN1;ACCN1	GGGGGCGGAACACCAGGCCTCCTGACTGCCGGGGAGTGGGCGCAGAGGGAGCGGGTTCG[CG]CGGAGGGCAACTCACGGAAGGAGAAGTTGTGCGGCCCGCAGAGCTCGCCGCGGTACTTGCA
130	cg07113642	20	2781262	CPXM1;CPXM1	AGCAGGAGCCCCCACATGGCGGGGATTGAGTGCCAGGGGGCGCGGGCTACGGCGGGTGG[CG]GGTCGGTCTCTTCCTGCCGAGTGCGCCGAGCCCCCCGCCCCTTCCTGCCCCCCGCCCCTC
131	cg04609163	11	125774090	DDX25;PUS3	CGTAAAGCGCCGGCGGGGAGTCCGGGGGGCTCCCGCCTGGAGGGGCTGTGTGAGCGGCGGGGC[CG]CGGGGCGGCGGCGGGGGGGCGCTCTCCACTCTGCGGAAGCTGCCCCCCTCTGCCCTCCGGTCC
132	cg11193865	14	101193038	DLK1	CGGCCCCGTGCTGTACAGTGTGAGGGAACGTGTACCAAACGCCTCGCGGGATACCTGTGCC[CG]TCTAGCCAAGAGTGCACCCGTGTGCGCGAGCGGGCTTCTGGGACGCCGCCGTCGTCGGGG
133	cg09124223	7	153749738	DPP6;DPP6	GGGGAGGGGGCGCAGGAGGCTGAGCCAGGCAGAGTCGCCAGCGGAGACTCGCCGAGTGGCCG[CG]CGGGAGGAGCGGCCCCCGGCCTGGGCTTGCCTTGCTGCTGCTGCCTCCCC
134	cg14768785	3	172166517	GHSR;GHSR	ACACCTCCCCTTTCCCCCACCAACTCCCCCAAAGTTTCTCCCAACACATCCTCCGGCCCGG[CG]CCCACACGCATACCTGTCACCAGCCCTGCCTCGCATTTGCGTTCTCGATCCAGTTCCATC
135	cg12477716	17	47073436	IGF2BP1;IGF2BP1	GTATCCGGGACTCCGAAACGCGCGGCGAGCAGCCCCTCCCCCACCGCCCAGAGGGGGTGT[CG]ACCGCCCACGTGTCGCCCTTGCCCCTTGCCCAGTCGGGTCCTTCCCTCGGGCTCCGGGAGCCGGAG
136	cg24826867	16	85932853	IRF8	GAACGCGGGCGGCGAGACGGCGGCAGGTAGGCACAGTGGGCGGGTAGGG[CG]CCCGTGTCCCGCGCGGTCCGGTCCCGCGGGGTCCCCGACGCCAGGCGGGGCGTGGGGGTG
137	cg24663256	4	21950307	KCNIP4;KCNIP4	ACATTCATGTCTAGGGACGCAGGGTGCAGAAGCGAGACTCGAGAGTCCACCGGCCAGAGGGG[CG]TCTGTCCACGGGTCTGCACGGGAGCGCACCGCCGCTCGGCCCGGGGGCGTCCGTGGCGCT
138	cg06653045	17	68165064	KCNJ2	TAATTCGCGAACAGTCGGGAGAACAAACAGCCAAGCGGCGCTGCAGTGGCCGCACTTGCG[CG]CGTCTCAATCCTGGGGGCTCTGCGCGCCGCCCCAGTCCCTCGCCCCATTGACTCAGTGG
139	cg13592399	14	52535758	NID2;NID2	CCGGTCCCCCTCCATGCTCGCTCGGCCGTGGCGCTTACCCGCTGCACAACGCGTCCCGCCC[CG]GCCTCCAGCCCACTCTCCGGCGCGCCAGCCTCGAACCTGGATCTCCGCGGGGCGCCTGG
140	cg20008332	2	5833169	SOX11	CGCTGGAAAATGCTGAAGGACAGCGAGAAGATCCCGTTCATCCGGGAGGCGGGAGCGGCTG[CG]GCTCAAGCACATGGCCGACTACCCCGACTACAAGTACCGGCCCCGGAAAAAGCCCAAAAT
141	cg02071005	2	175200844	SP9	TTCCTCTTGCCCTCCTCCTTCTTGCTCCCTCCCCCATCCCACCACTCTAGGAAGAGCCGG[CG]CTTCGGAACGACCCCGTTGGCCATGCTGGCGGCGACCTGCAACAAGATCGGCAACACGAG
142	cg23302682	6	166580952	T	ATCATCTCATTGGTGAGCTCCTTGAAGCGCAGCCACAGCTCGCTCTCCTCCAGGCCCACG[CG]CAGTTCGCGCTCTGTGGGTCGCCCTTCTCGCTGCCCGCCTGCAGCTCATTCTCCACGGC

143	cg07978472	4	122686493	TMEM155;LOC100192379	GCCAAGGAGCTGAGGAAATCCGGCGCAGACTCTCCCAGCTGGCACCAAAGCCTTCCGCTT**[CG]**CCGAGATCCTCTCAGGTGCTCTTGAGGACGCGAGCGGACTTCCCTAGGAGCGAACTTCCGC	143
144	cg14732324	5	528621		CCCTCGGGGACAGCCCGGCCGGCCACGCCGCCGAACTCGCCATGGGCCTCTCTCCGCATC**[CG]**CATGTGCATCCGCAACCGCTTCCGTCCCGCTGAGCGCACGAACCCTCTCGCTCCTGTCCC	144
145	cg00228475	6	58149279		TCCTCAAGAGAGGTAGGGTCCGTTCCCCCCGCGGGGCCGGTTAGCTCAGTTGGTTAGAG**[CG]**TGGCGCTAATAACGCCAAGGTCGCGGGTTCGATCCCCGTACGGGCCACAGGCTTTTCTAA	145
146	cg12812583	8	23567310		GAGCAGGCTCCCCAGCGTAGCGAGTCCTTGTTATGGAAAGGGTCGTTTCGGCTCAGGATG**[CG]**CGCTCCCGGCGTAGACCTGGGGATAGGGGTCCCTGTCGCGCTCGCCCCACCCCTGCAGGG	146
147	cg01163842	14	95235125	CSC	GTGGCGGCGGGACACCCCGCGCAGGCCAACAAAAGGAGGGGAGCCGCTCGCTCCCGCTTC**[CG]**CGTTTTCATTCAACTTCCTGGGCCTAAAGCGCCCTCCAGCAGCCTGCGGGCCGCCATCGG	147
148	cg17394649	6	29760164	HCG4	AAACGCCGTCTGTGGGGAGTAGCTAGGGGCCTGCCCGGCGGGGCGCAGGAACCCGGTTG**[CG]**GTGCCCGGGAGCGAGGGTCGGGAGGGTCTCAGCCCCCTCCTTGCTCCCAGGCTTCCACTCCT	148
149	cg10723962	6	26240782	HIST1H4F	GCGTGACAACATACAGGGCATCACGAAGCCCGCCATCCGTCGCTTGGCCCGAGCGCGGCGG**[CG]**TGAAACGCATTTCGGGCCTCATTTATGAGGAGACCCGCGGTGTTCTTAAGGTGTTCCTGG	149
150	cg17466857	7	27225528	HOXA11AS;HOXA11	GGCTTGTCCGATTTGCACGGTGACTTGATTACACTCTCTCATTCATGGTCACTTCCGAAG**[CG]**CTTTAGTGCCTTCCGTCCCTAAACCGCCAACAGCCAGAACGGCTTCTCCCCGCGGTTTGT	150

Table 2

Table 2 below provides exemplary primers which may be used to amplify or sequence MVPs defined in Table 1 above.

Chr	Sense Start	Antisense Start	Sense sequence	Sense SEQ ID NO	Antisense sequence	Antisense SEQ ID NO
2	119,606,705	119,606,804	TTGGAGGATTAGTATTYGATTATGTTGAA	152	AATCCRCTAAAACTCTAAAATAACACCC	405
19	54,926,752	54,926,852	CCTCAACTTAAATACATCTCCTCC	153	TTGTTGGTATTTTTGTTTTTGGG	406
7	70,597,038	70,597,139	GGGGTGGGGTGGATTT	154	CTCAAATCTCAATTACTCTCAAAATAAA	407
4	111,544,332	111,544,433	CCTCCTCCCAAACCCTTCT	155	TGTTGATAGGTGTAGGTAGGATAG	408
7	27,205,053	27,205,155	GTTTAGGGTTTTAGTGGTGGT	156	CCACCAAATTATTACATAAAATCTACAA	409
7	5,569,082	5,569,184	GGAAATYGGGAGGTTTTTGTGTA	157	AAAACRCCCCAAACACCAAATAAAA	410
1	237,205,102	237,205,205	ATCTCTACTTAAAAATTAACCACCC	158	TTTTGGATTGGGTTAGGGTATT	411
15	68,122,063	68,122,166	RCCTTACCAAATTCCTCACAA	159	GATTTTYGGGGTGGGTGGG	412
13	78,493,251	78,493,354	TTTGGTTGGTTGAGTTATAGTTTT	160	ACAACRATTTCCAAATTCCTACTAAC	413
7	70,597,035	70,597,140	GGYGGGGTGGGGTGGA	161	RCTCAAATCTCAATTACTCTCAAA	414
8	65,290,282	65,290,387	RTTTTCACTACTCCAACTCCC	162	TGATTGTAAGYGTAGGTTTGGGT	415
18	74,961,907	74,962,016	TTTGGAAAAGTYGGGAGGGAG	163	CAACAATACACTATACRACTCCTACAA	416
6	28,367,508	28,367,619	YGGAATTTATTAAAAGTGATTTATAAAGGT	164	ACCCCTTTCRCTCCCTTCCTA	417
2	223,163,755	223,163,866	CCCAAAACTTAATCAAAAACCCT	165	AGAGATGGGAAGAGAAAGTGG	418
6	166,582,153	166,582,264	CCTCCTCRCTCTCTTATAAAAACAAAACT	166	TTTTATTTTTYGTGAAAGTAATGATATAGTAGAA	419
8	26,723,555	26,723,669	GGGGAATTTTGGAGGATGTATT	167	ATCCTAACCCTAAAATCCCTAAAT	420
6	108,440,301	108,440,416	GGGGATTAAGGGATTGTTTATTTTG	168	CCTTCTCAATAAATACATTTACCCC	421
8	132,054,523	132,054,639	GTTTAGTGTYGGTGTTAATGATAGATG	169	AATAACCACCACCRCTCCTCC	422
5	115,152,343	115,152,460	AGGTTTAGATTTGTGGGGTTTA	170	RCCACCTTTCTTAAATAACTCT	423
10	106,401,968	106,402,085	TTAAAGATGAGTGGGGGAGG	171	CCCCAAATCAACCCTTTTTC	424
11	20,618,161	20,618,279	AAAGTAATTAAGGTYGTAGTGATTGGT	172	RCCTATTTATAATAATAATCAAACCCA	425
19	54,926,741	54,926,860	GGGTTATYGGTTTTTAGTTTGGGTG	173	RATCCCACCTACTAATATTCC	426
2	223,163,760	223,163,880	AGTTTGGTTAGGAGTTTTGAGT	174	ACCTCRACACCACCAAAAATAAA	427
22	22,862,820	22,862,940	CCTACACTCCCCCACAAAA	175	TGAGGGGAGATTTGAGGGAT	428

11	120,434,955	120,435,076	YGGGGGAAAATGTTTTTATTGA	176	AATTCCCRAACCCCCTCCC	429
8	23,563,877	23,563,998	CCTCACAAAAACCCCCAAAA	177	GGAAAATTTTTAGTATTTGAGAATGGA	430
2	198,650,902	198,651,024	TATTTGGYGGGTGGGGAGAA	178	TCRAAATCCCAAAACCACAACC	431
6	10,882,276	10,882,398	ACCTCTAACAACTACCCCCT	179	TTTTTGGGATTGTGGTGGAG	432
20	25,065,180	25,065,302	CCACCCAAAACCCAACRTCAAATTA	180	GGGTGGATYGTGGGTTAGTTTT	433
7	27,204,952	27,205,074	GTTTGTYGGGGAGGTTGGTTT	181	AACCACCACTAAAACCCTAAA	434
5	172,671,485	172,671,608	CCTCTTCCTTCCTTCCAAATC	182	GGTGTGTGTGTGTAATATAATAATTTGT	435
3	147,113,701	147,113,826	GGGAAAGGATAAGGGAAGGG	183	ACAAACCAACCACATTTACTTCT	436
12	85,667,301	85,667,427	AGAGTTGTTTTTGATTGTAAGGGA	184	AAAACCACCTTCCTAACTCC	437
7	27,232,754	27,232,882	ACCACTCACCCRCACAAACAC	185	ATGGAGGTTTAGGTYGGTGTAAA	438
8	145,106,354	145,106,483	RTATAAAAACAAAACACATCCTATTAAC	186	TTATGGGGGYGGGTGATGGA	439
11	112,833,711	112,833,841	GGGGYGGGTTAGGGAGTATTT	187	ACTTCCCRCCCAACCAACTTC	440
7	155,259,749	155,259,880	AAATTGTAGTTTGTTAGGTTGAGT	188	ACCCATCTCCCCRAAAACTAAT	441
4	174,450,311	174,450,443	AAAAATAAAAATTCTCCTCATaactacaa	189	GAAATGAGTTTGGTAGGTGGTT	442
8	132,054,519	132,054,651	AAAAACTTAATACCRATACTAATAACAAATAAAC	190	GGTGGAGGAGGTGGTGATTA	443
10	118,030,953	118,031,086	AACTCCCRAACCACCCCC	191	YGTTGATTATGGTTGGTTTGTT	444
1	111,217,615	111,217,749	CTCTCCRCCCCCTCCCCTA	192	GAGGGGTTTTGAGGGTTGTA	445
15	86,233,181	86,233,315	CTCAAAACTACTAAACTCTACCCTA	193	TGAATTTTATTGTTATGTGGGGTATT	446
1	248,020,537	248,020,671	GGGAGGYGGGTTATGGTTTGG	194	RAAAATACACCTAAAACAAAAACTAT	447
2	66,667,006	66,667,141	TTTGAGAAATGTGAATTAGTTATTTGT	195	ACTTTTTAACATCTATTTTACTAACCTATT	448
5	16,180,236	16,180,371	TTGGATGGGTYGGTGGGATG	196	CCCTTTCCCRCTCCACTACTC	449
3	42,306,888	42,307,023	ACCACAACTAACTCTTAATATCCT	197	GGGTGGTGGTAATTTGGTT	450
7	24,323,740	24,323,875	AACTACCCRAAACCCCTCCT	198	TGGTTGGGYGGTAAGTATTATGTTG	451
10	15,761,845	15,761,980	AAAATCTAACRATCCCCAACTACC	199	AGGGTTAGAGTTTTTGGGTTAG	452
8	72,755,998	72,756,133	GGGGGTATTTAGGGTAGGTTG	200	CCCTCCCRACCAAAAACTCA	453
12	54,408,585	54,408,721	YGTTTGGAAAGAAATGGAGGTA	201	ACAACAACCRAATAATCCCCATTC	454
3	147,113,681	147,113,817	CAAAAACCTTCCCACACCC	202	TTATATTTGTTTTTGGGAGGAGTG	455
11	71,955,309	71,955,445	AAAACCCAAACACCRCCCAAAA	203	TGGTTYGGATTGGGGTAGGAT	456
6	29,760,051	29,760,188	GGGGGTTTTGGTTTTGATTTAG	204	CCTCCCRACCCTCCTCCC	457
2	66,667,480	66,667,618	AAGATTGAGTTTTGGGTTTGTT	205	TCCTAATTTCTTACCTCATTACACT	458
11	71,955,292	71,955,431	GAAGGGGYGGGGTTTAGGAG	206	ACAAAATAAAAACTCCRAAAATAAAATCCC	459
1	50,886,838	50,886,977	TTCTCCTCCTACRCCTACTACCT	207	GTGGTGTYGGTTTTTAAGGGTT	460

14	95,235,085	95,235,226	GTAGGTTAATAAAAGGAGGGGAG	208	ACCCAACCCTACTTAACTCTC	461
7	5,569,280	5,569,421	ATTATTTATGGTGAGTTGYGAGAATAGT	209	ACCCRACTTCCTTTATCCCCA	462
10	15,761,835	15,761,977	GGGGGTTGGTGGAATTTG	210	ACTAAAACCCCCTAAACCAACC	463
1	119,530,560	119,530,702	TCTCCACCCTCCCCCT	211	GGTGGGTGGGAAGTAGGAT	464
4	174,450,301	174,450,443	AGTTAGTTATGGAAGTAGGGGT	212	RAAATAAATCTAATAAATAATTTTCCCC	465
7	5,569,274	5,569,417	GATATTATTATTTATGGTGAGTTGYGAGAATA	213	RACTTCCTTTATCCCCAATCTa	466
2	198,650,997	198,651,141	AGTTTYGGTTGTGGTTTTGGGA	214	TCTCTCCTCCCCAACCRTCTAAAA	467
3	42,306,941	42,307,085	GGATGATATTATTTGGTTTTGTTTAGAG	215	ATCAAATTCCCCAAAACCCT	468
7	19,184,912	19,185,056	GGGGAGGTTAGGGATAGGTT	216	TTAAAAATCCCCACCAACAAC	469
14	52,734,499	52,734,643	CAACCTTCACTCCAACCCT	217	GGTTAGTAGGTTGTTTAGGAGG	470
13	109,147,820	109,147,965	GGGGAAYGTGGAAAGGAGGG	218	RTAAAACTAAACTCCAACTCCC	471
1	50,513,833	50,513,979	AAAAATACRCCTCAAAAACCAAATAAAAAC	219	TTGGGAGTYGGGGTGGTTAG	472
14	101,192,985	101,193,133	GTGTTGTATAGTGTGAGGGAA	220	CAATCCCTATAACCCCCTCC	473
20	21,686,229	21,686,377	RAAAACCAATAAACTACCAACTC	221	YGATTTTAGGTTTAGGGTGAATTTT	474
7	24,324,907	24,325,055	GTTTTTGTTYGTGTGTTTGGGTG	222	CCACCCACCTCTACCTAAT	475
2	66,666,426	66,666,575	TGGGGTTGGGATTTTTAGGT	223	AAAACTAATTCCTACRAATTCCTCCTA	476
5	76,923,868	76,924,017	GAGTTTTGGTAGGTGTTGGT	224	AACCTCTTTAAAACCTTCCCTAA	477
2	66,667,385	66,667,534	CTCTACAAAATAAAATCAAAAACACCA	225	AGAAATATTTTAGTGTGAATTAAATAAGTTG	478
10	106,401,990	106,402,139	TCTATACTCACTTTCTCCAATACTT	226	tgggggtagggGAGTT	479
20	2,781,198	2,781,347	AACAAACACAAAAACCCCCACA	227	GTTGTGGGYGGGTAGGAGGT	480
11	112,833,653	112,833,803	CAACTCCAAATACCRTTATACCTACCCTA	228	TGGGAGGGATTYGAGTTGGTTG	481
1	248,020,634	248,020,786	TGGATTGYGGTTATAGTTTTTGTTTTAGG	229	CCACCAAACCCRCCAACTACC	482
13	78,493,250	78,493,403	CCCTAACTAACTAAACTACAACTCC	230	TAGGTTGGGTAAAGGAAGGA	483
6	108,440,232	108,440,386	TCTAAAAATTAACTCTAACTTCCCCA	231	GGTTTTTATTTTTYGTTTTTATTTGATGAGTT	484
2	182,322,443	182,322,597	TGTTGTTGTTGTGTTTGGGG	232	TCAATCCAACTCTACTCACCAT	485
11	20,618,126	20,618,280	RAAAACAAAATCAATTACTATTTTTCATCT	233	YGTTTGTTTATAGTGATAATTAGGTTTA	486
8	145,106,394	145,106,549	GGGGAGGYGGGTTTAGTGTAG	234	CCTAACCAAAAACCRAATTTAATTTAACC	487
15	86,233,182	86,233,337	TTAAAGTTGTTGAGTTTTGTTTTGA	235	ACCTCCCCATACTTTAATCCT	488
6	29,760,062	29,760,217	TCCTAACCCAAACCTAAACAAA	236	AAGTTTGGGAGTAAGGAGGG	489
12	85,667,262	85,667,417	TCAAAATAAAAACTCCTCCACCTAT	237	TTTTGGTTTTYGGTTTGGAGGG	490
4	147,561,686	147,561,841	TGTAGGTATGGTTYGAGGAGGT	238	AAAACCRAAACTAAATAACAAAATAAACTTC	491
6	28,778,138	28,778,295	GTGAGGGAAGAGAGGTGTTT	239	CAAACCAAAACCCAACATCAC	492

7	27,225,023	27,225,180	TGTAGTTATTTTAGGGGAAGTAATAGAT	240	AACCCTACAATTAAACACAAACAT	493
14	24,045,512	24,045,669	GGTAGGGGYGGGGGTGG	241	RCAAACACCTACCAAAACCA	494
13	28,366,707	28,366,864	RTTCAACCTCCTAAACAAAAACAA	242	GTAGTATTAGYGAGTTTATTAGGAAGGAG	495
4	147,561,678	147,561,835	ACCCATCCTACAAACATAACTC	243	GAGGTTGAATGGTAAAGTAGGTT	496
10	22,518,210	22,518,368	GAAGATTTYGGGGAAGGAGTGG	244	AAATAAATTTCCAAAAACCAAACAAAA	497
7	27,225,011	27,225,169	ACTTCCTTTCTTTATAACCACCTC	245	TGAGTATAAGTATGTTGTATGGGG	498
14	95,234,999	95,235,157	RAAAAACAAAACAATTCAAATCAA	246	YGTTTTAGGTTTAGGAAGTTGAATG	499
1	50,513,832	50,513,991	GGAAGATGYGTTTTAAGAATTAGGTAGAA	247	CTCTCCCCCTCCCTAAAAAT	500
10	106,401,430	106,401,589	GGGTYGGGGTATTTTAGTTTTTGT	248	CACCTCCTCCCRAAAACCCT	501
X	136,656,476	136,656,635	GAGGTTTGGGGATTGGTTG	249	CCTCCTCTTTCCTCCCAA	502
2	66,666,980	66,667,139	CCCTCTTCCCTCTCTTAACAC	250	TTTTTAATATTTATTTTGTTGATTTGTTTGG	503
10	106,401,346	106,401,505	AAAACTACTACCCCAACAAAAC	251	TTGGGTGATTGGGGGTT	504
12	62,584,934	62,585,094	GGGGAAGGGAGAYGTGTGTA	252	CAAAAACTACAAAAACATCRCTAAATATTACC	505
1	111,217,518	111,217,678	CCCACCRCCTATTACAACCAAAAC	253	GTGGGYGGGAATTTTAAGGGG	506
2	66,666,422	66,666,582	ACCTTAAAATTAAAATCCCTAAATACAAC	254	GTTGGGGAGGGTTGGTT	507
5	115,152,387	115,152,547	RCTTAAAATCACTAAAAATATACCAAC	255	GAGATAAYGGGGTTTTTGGGAAG	508
5	180,596,604	180,596,764	GTGTAAgttttygtagtgtagtggt	256	CCAAATAAAAATAATAATAAAATATTCCAAACT	509
22	50,216,544	50,216,704	TCCCAAAATCCCACACTACA	257	GGTTTATATTTAAGGTTAGGAAGAAGG	510
7	27,225,490	27,225,651	ACTTAATTACACTCTCTCATTCATAATC	258	AGATGAGGGGAGAGGTGG	511
11	120,434,988	120,435,149	RCAACTTCTACCTTTTATTACAAAC	259	GGAAAGGAGGGTTAYGGGTAAAG	512
1	237,205,044	237,205,207	AATTTTTATTTGATTATGAATAGAGGTAATTT	260	CCTCCTAAACTAAACTAAAACACTAAC	513
20	2,781,222	2,781,385	GGGGATTGAGTGTTAGGGG	261	ACCCTCTCCTCCTCTCC	514
7	5,569,852	5,570,016	CCACCCTACRATCCCCATTAAC	262	YGGGAGAAGTTTTTGGGTT	515
2	175,199,604	175,199,769	TCTACTCTACCTACRCCCTCATTAAA	263	TTGTTYGGAAAAATTGTTTGGGTTT	516
18	44,787,470	44,787,635	CRAAATCCTCCCTCCTACCTC	264	GGGTGGGATAGGATAGGGT	517
11	14,926,608	14,926,773	CACAAACCCAAAAACCCCA	265	GAGTAGAGTGTAGGTTYGAAGGTAT	518
20	25,062,422	25,062,588	GTTAGGAGTAGGTAGGGTGT	266	CCAAAACAAAAATCTAACTACRAAAATCC	519
6	391,841	392,007	GTGAGGTTGATATTAGAGAGGAT	267	AACCTTAAAACCCAAACCAAC	520
7	19,184,852	19,185,019	RCAAAAACTAAATCCCCCAAAAA	268	GAAAATYGATGAGAGGTAGGGTTAAG	521
5	140,787,404	140,787,572	TTGAGGGGGATGTATATTTGTATT	269	AAAAATCCCAAACCAAAAACTAAC	522
18	74,961,845	74,962,014	RACTCTATCCACCACCAA	270	GTAGTGTATTGTGTGGTTTTTGT	523
6	27,513,424	27,513,594	AACCATTATAAATTCACCCCAAC	271	TTTGGGTAAGTTTTGAGGAAAG	524

6	26,240,675	26,240,845	GGTAAAGGTTTAGGAAAGGGAG	272	CTCCAAAAACACCTTAAAAACAC	525
6	28,778,185	28,778,355	AAACCTAATTCRTTTCACCCAAACTATC	273	GGAAGTTAGGTAATTTTTGAAGTTTTT	526
2	66,666,605	66,666,776	TCCACCRAATCCTAAATATACAATAAAAA	274	YGGAGGTTTTGAGTTATGGA	527
7	27,204,213	27,204,386	TTAAGGAGAGGGGYGGTTAGTT	275	RAATTAAAACCAAAAACCAAAAAT	528
20	25,065,065	25,065,239	ACATTTTCATAACCTCCTACAATAAA	276	TTTGGTTTTGGTTGAYGTTGAGT	529
7	20,817,774	20,817,949	CTACAAAACRAAAACAAAACACAAAATAA	277	GGTGGTGGTTTGGGGT	530
11	125,773,960	125,774,135	AACCAATTCRCTAACCTTTCTAACATC	278	gggGTAGTTTTYGTAGAGTGGAG	531
7	27,204,899	27,205,074	CCTTAAACTAAAAACTACACRAACTAAAAT	279	GGTTATTATTGGGGTTTTGGGTAA	532
5	172,671,421	172,671,597	AGTGTTGAATTGATGTTGGAAA	280	ACAACACAACAATTTATCAACTACT	533
18	55,108,733	55,108,909	TCTACACCCCCTCCCC	281	GGTTTAGTTATTAGGGTTTAGTGGT	534
14	70,653,838	70,654,015	TTTTTGATTATTTAGGAGTTTGGTTG	282	ctttcctctctctctctctct	535
6	99,841,571	99,841,748	AAACACAAATACAAAACTATACATAACT	283	GGATTAGTGGTTTTGTTTGGAAAA	536
14	24,045,441	24,045,620	CCCCRAACCCAACCAACA	284	GGGTAYGTTAGAGTGTGTTTTATTATTAG	537
2	66,666,618	66,666,799	GGATGTGTAGTGGAGGGG	285	CCTTCCTCCTAAAACCCTAAAAT	538
10	135,043,371	135,043,552	gggggtAGTATTGGGGGT	286	AATCCTCAAAAATTCTATTCTTAAACC	539
4	122,686,466	122,686,647	CCCAACTAACACCAAAACCTT	287	ATGTTGGGGGTGGAATTTT	540
20	37,352,962	37,353,143	TCCAATCCCCCACCCC	288	GGTGGTTAGTGTATTGYGGAGTTG	541
7	27,204,196	27,204,378	ATCCTAAACCTAAACAACCAAAAA	289	GTTAGGAGTTAGAAGTTGGTGT	542
8	26,723,434	26,723,616	CCAAAACCCCAAACTCCAA	290	TTTTAGGTGATTGYGAGGTAATTTGT	543
1	50,886,642	50,886,825	RCTCTCACCTAAACCCCC	291	GTTGTAGTTGTTTTAYGGTATTGTTGA	544
6	27,513,336	27,513,519	GTTATTGATGTGTTTTGAATGAGTT	292	aaaaccccaataaatttcaaatcc	545
7	5,569,687	5,569,870	ACACCCCRAACCCCAAAAC	293	AATGGGGATYGTAGGGTGGG	546
10	22,518,180	22,518,364	CCAACACCCTAACRAAACCTAAAC	294	GGATTTTTAGGGATTAAGTAAAGAAATTAT	547
7	27,232,735	27,232,920	GTTAGTTTGTAGTTGYGGGGTTATTTA	295	CAAACCCTCCACRCTTCTACAAAAA	548
2	162,283,572	162,283,757	ACAATACATTCRTATCATCACCCACC	296	TGGTTTAGTTTGGGYGGAGAGTA	549
2	45,161,903	45,162,088	aaatctaccccaccctacc	297	GGAAGGGGGATAGGGGAT	550
6	166,582,063	166,582,250	GTTTTTATTATTTGGAAAAGGAAGGTT	298	AAAACAATAACACAACAAAAACCA	551
8	72,756,006	72,756,193	CCAAAACAAACTAATCTTAAACCTAAA	299	GGTTTYGTGTGGTTGGGGG	552
14	29,254,632	29,254,819	CCCTCTTTCCCCTTCAATCTT	300	TGAGGTTTGTAGTGAAGGGT	553
15	76,629,983	76,630,170	ATCATCTTTCCTTATAATCRCAAATCTTAAAAA	301	TGGTYGTGGATTTGATTTTTGTAGT	554
8	72,754,933	72,755,121	ACTTCTTTAATATCAAAATCCRATCTTCC	302	AGAGGATATGTAAATTTTTAGAATGTTG	555
8	23,567,163	23,567,352	atactttctaaCCCCTCTCRAAAAATA	303	GTGYGATAGGGATTTTTATTTTTAGGTTTA	556

12	62,584,867	62,585,056	ACCCRCATCTACCCTCACCT	304	GGGAAGYGGGGTTGTAGGTA	557
7	5,569,805	5,569,994	CAATAACRCCCCCAACCCCCAC	305	GGAGATGGGGGATATTTTAYGTTAGTT	558
19	58,238,678	58,238,868	CTAAAACACAAAAACTACAAACACT	306	TTGATGTTTTGTAGGGATGGA	559
5	16,180,017	16,180,207	ygggtgggttgggg	307	CCTACCCCCCTACTTCCCA	560
13	28,366,672	28,366,863	AGAGGTTATTGTTTTAGTTTAGGTTT	308	CAACACTAACRAATCCACCAAAAA	561
1	248,020,475	248,020,666	ACTCCTCCCCCTATACAAAC	309	TGTATTTGAGGTAGAAGTTGTGG	562
4	147,559,282	147,559,473	AAACCCTACAACCCCTCC	310	GTTTAGGTGATTTTGATTTTAGGTTTT	563
3	147,113,702	147,113,894	GGAAAGGATAAGGGAAGGGT	311	TCCAAAACTTTCAACACCATAC	564
4	147,559,273	147,559,465	AGTGAGAGAAAGTTTTGTAGTTTT	312	AACCCTAATCCCCAAATTCCC	565
6	10,882,217	10,882,409	GGGYGGGGTTTTGTTTTTGGT	313	CTCCACCTCCRTTCCTAAAACTATAATAA	566
12	54,408,539	54,408,731	ACAATACCCRCTCCCCCACC	314	AGTTTTATTTGTAGTAGTYGAATGGTTTT	567
18	22,930,230	22,930,422	RATCAACTTCTAAAAATAACCAAAAA	315	YGGAGGTGGGGAGTAGTT	568
8	72,754,903	72,755,096	GTGGTTTTATATAGTTTGTTGGTTG	316	CTAATAAAATCTTAACTACTCCAAATCAA	569
2	66,667,490	66,667,683	TCTAAACTTATTTTCAACTTATTTAATTCAC	317	GTGAGAAGGAGAGGGGAG	570
20	21,492,888	21,493,082	GATGAGGTTGGTTAGGTGTT	318	TCCAAATCCCCRAAACCCTC	571
6	391,785	391,979	CCCTCAACTCCRAATCCAAAAC	319	GTTTAGYGTAGATGGGGTGGGA	572
7	24,324,845	24,325,039	CTTCTATACCTACAAATACTAAATAACAAA	320	GGTGATGAGGTTGATGTAGTGT	573
7	70,596,935	70,597,129	AAACCRATATACCCTAAAAACCCC	321	AATTATTTTTAAAGTGGGGGTAGTAT	574
5	528,491	528,685	RACAAACCCTCCCAACA	322	ggAGGGATAGGAGYGAGAGGG	575
2	66,667,301	66,667,496	TGTTTGAGTGTTTGTTTGTAGAT	323	ACCCAAAACTCAATCTTACTTCT	576
4	174,450,245	174,450,440	GGTTGTAGGATAGGGTTATGTTG	324	ATAAATCTAATAAATAATTTTCCCCACC	577
20	21,686,179	21,686,374	GGAGGGGGTAGAGTTTTAGG	325	CCCCAAACCCAAAATAAACTTC	578
1	149,719,383	149,719,579	ATGTTGAGGGTGTTAYGGTTTTATT	326	CACCTACCAAATTTACAAATCCC	579
7	1,272,352	1,272,548	GTTTATTGAGGTGTATGTTAGGTATAAT	327	CCCRAAACCCCACACATACTT	580
2	119,607,734	119,607,930	AATTCATTTTACCTTTCATATAATATAAAACC	328	TTTGGGTGGAGGTTATGGA	581
15	68,122,089	68,122,286	GGGTTGGGGGTAGYGGAAAAT	329	AACAACCRAACCCACCAAAAA	582
15	76,630,037	76,630,234	GTTGGGTTGGGGTTGGA	330	ACACTCRACACACTTAAAACAAAC	583
3	129,693,494	129,693,691	CCTCCCCACTAACCTCAC	331	GAAGGAGAGGGGTTGGG	584
5	172,659,564	172,659,761	TCCCTTCCCTACCAAACTC	332	GGTTTGTAGTTTTGGTTATAGTTGT	585
7	149,917,234	149,917,431	CCCTCCACAAACCTACCAAA	333	GAAGGTAGAGAATTTGGGTTTTT	586
14	60,975,848	60,976,045	CCAATCCRCCCACCCCAATAAC	334	TTGAGTGTGTTAGGYGTTTGTTG	587
2	162,283,566	162,283,764	AGTGGAGTAGTGTATTYGTGTTATTATTTA	335	RCTCCAATAACCCAACCTAA	588

20	21,081,785	21,081,983	GGGATGGGGAAATTATTTGATTAG	336	CCTATTTCTACCCCCACCC	589
8	23,563,886	23,564,084	GATTTTTAGGAGGTTTYGAATTATTTAGTTT	337	CCCTTCTCRATCAAAAACATCCTAC	590
16	85,932,830	85,933,028	GTAGGTATAGTGGGYGGGTAGG	338	AATCAACRACTTAAACATTCAAATTAAACT	591
13	53,420,219	53,420,417	CCCCTTTACRCACCTCCTTCTT	339	GTTGYGTTTGTTAGTGTAGGAAGTT	592
2	119,606,711	119,606,910	AACCAATATCCRACTATACTAAAACCTA	340	GGAAAATTGGAGGTAGGGG	593
7	27,205,063	27,205,262	CCAATAATAACCATCACCRTACCCAA	341	GGAGTGATTTATGYGTTATTGTTTTGT	594
22	50,216,594	50,216,793	TTTATTTTTGYGGGGGAATATAAGGAG	342	AAACTACACCCRAAAACCTCTAAAT	595
22	50,922,498	50,922,697	CAACAAAACTCAAAACATTCCC	343	GAGGTTTTTATAGGGTAGGATTAGT	596
22	22,862,796	22,862,997	GGGGATYGTGGGATTTGGTT	344	AACAACAACTAATCCRAACTAATAAAAAC	597
1	111,217,055	111,217,259	GGGAATTGGTAAAGGGTTTTTAG	345	AAAACCACCTACTAAAACCRAAAATAA	598
6	28,367,514	28,367,722	CCATCAAAAATAACCCACAAAAAC	346	GGGTAAAGAAGAGGAGAAGATAAA	599
1	111,217,355	111,217,565	GGGYGTTGAGGAGGGTGG	347	RCCAAACCCAAACAAAACA	600
2	119,607,767	119,607,977	ATTTGTTATAGTAATGGGATGATAAATG	348	ATTCCCAATAACATATTCATTTACAAAA	601
7	149,917,184	149,917,394	AGTTATYGATTTTTGTAAGGGATGTAGA	349	CATAAATCAAACCRCAAAAATCTCCAATA	602
2	63,281,084	63,281,295	AAGGTYGAGATTTGGGTTTGTTAG	350	CCTCCTCCCRCATAAAAATATCAAAATAA	603
1	50,886,613	50,886,825	TTTGTTGTTTTTATTTGYGGTTTGGT	351	ACTACAACTACTCTACRACACTACC	604
5	135,266,111	135,266,323	RAACTCCAACACCTTAACC	352	ggAGGGGTYGTTTTAGGGGG	605
7	155,166,900	155,167,114	GGTAGGAAGAAGGGYGATGTTTTT	353	ACTAAAACCCRCACCTAACCCA	606
14	95,239,509	95,239,725	ACTAAACAAACCATCAAAACCC	354	GTTTGGGTGATAATAGGGAAAATTA	607
10	118,030,867	118,031,084	TTTTAAGGGTTTGTTTTAGAGTTTG	355	CTAACCATAACTAACCTACCCC	608
3	147,113,558	147,113,776	AGTTGGGATTTGAGGAAATTTAG	356	CCTTCAAAACCAAATACAAACTTAT	609
11	20,618,134	20,618,352	AATCAATTACTATTTTTCATCTTTAACAAAA	357	aaagaagaagaagaagaagaaagaTTAG	610
10	118,030,866	118,031,086	ACTTCAAAAATTTACTCCAAAACCT	358	YGTTGATTATGGTTGGTTTGT	611
11	71,955,207	71,955,427	AATACAACRCAAATACAACCAACCC	359	GATAGGGATTTYGGGGATAGGG	612
11	14,926,603	14,926,824	GGGTGTATAGGTTTAGGGGT	360	CTACAAATAAAAACTTCTTTCCAAATAAAC	613
5	15,500,202	15,500,423	CCTCCCCRAAACTCCAACTAT	361	GAGTYGGGTAGGTTTTTGGGTT	614
2	63,281,053	63,281,282	TCTTTTATTCCCAATTCRACTTTCTTT	362	GAAGATGTTAGGGTAGYGAGTTTTG	615
14	29,254,647	29,254,879	AGTTTTTAGGGAGGGGGGAG	363	CAAAACAATAATTTCTCAACTTTTCC	616
1	111,217,544	111,217,780	YGATGTTTTGTTTGGGTTTGG	364	AATACAACATAAAAACTCTTTCRCTAACAC	617
11	31,848,615	31,848,857	GTTGGGTTGGGAGAAGTTT	365	ACTCCCTCCCCTTCTATTTTCA	618
4	111,544,173	111,544,425	GGGTTTTGTYGTAGTTTAGTTTTAGG	366	AATACAAACAAAACAATCCCTCCAC	619
7	27,205,054	27,205,314	TTTAGGGTTTTTAGTGGTGGTT	367	ACTTTTACTATAAAAATTATAACTACAAAACATC	620

19	58,238,712	58,238,972	GGAGTYGGAGAAAGGGTGATT	368	ATCAAACTATCCCTAACCRAAATTCTA	621
5	180,596,452	180,596,712	CATCTCAACCTTCCAAATACTAAA	369	TGAAGTAATGAGATGAAAAGTATAAGAG	622
2	63,283,871	63,284,132	TGGYGTGTTTTTGTTTATTGGAGTATT	370	RAAAACTTAAACCAATCCAAC	623
15	65,116,211	65,116,473	ACAATCCRAAAACAACAACTACACT	371	TGGTTGTGGAGGAGTTGAG	624
5	140,787,349	140,787,617	ATTTAAAACACAAAACATAAAAATATCTACTA	372	TTTTGTGTGGGAGTTGGTT	625
14	95,239,425	95,239,697	GGAGGTAGGTTTYGGGAAAGG	373	ACATATTTACTACATTTCCRACCTAAAC	626
10	135,043,275	135,043,552	ACCCATCCCCTAACCTAAC	374	GGTTTTTAGGAGTTTTGTTTTTAGAT	627
19	58,238,625	58,238,903	CCCACACCCACCCAC	375	GGAGGAGATGTTGTTTTTAGTG	628
6	26,240,714	26,240,993	AAAATACTACRTAACAACATACAAAACATCAC	376	GTTGTTAAATAAAAAGTYGGGGTGAG	629
20	25,065,005	25,065,285	TGTGTTTTGGGTTATTTTGTGTT	377	ACTTTTTATCTCTTACAAACRTCTCCTAAAC	630
7	155,259,636	155,259,918	AAAATTCTACTAAAAAATCATTCTATCTCC	378	GGGTAGTAGTGTGTGTAGGG	631
5	172,659,477	172,659,760	GGATTTAGGGTTATGTTGGGAG	379	ACCTACAACCCTAACTACAACTA	632
7	20,817,773	20,818,057	GTTGTAGAAYGGGAGTAGGGTATAG	380	CAACTCCTCRAAATACCCAATACA	633
14	52,734,368	52,734,652	GGAGGTTTTTGTYGTGGGGAATA	381	CAACCCCAAAACCAACAAAT	634
20	61,638,444	61,638,729	CACAAACCRAAAATAAAAACTCTAAACC	382	TTTTTAGYGGGATAGGGTGTTGG	635
18	77,547,903	77,548,190	CTAACAAAACRACCCAACCAAAAA	383	GTTAGGAGGGATTYGGGAGGT	636
11	120,434,917	120,435,205	TTAGAYGGTATTAGGTAGTTGAATTTAGT	384	AAACCCTTCCCAACCCT	637
9	37,002,424	37,002,713	GGTTAGTAAGAATGTTATAGTTTTATTTTGT	385	ACTCCATCAACRACATCCTAAACA	638
1	91,182,909	91,183,198	AACAATCTATAAATACTTTCRACACAACT	386	yggtgttaggttgtgggt	639
18	44,787,350	44,787,639	AAAACTAAAAATTCCTAAAATCCCTTTA	387	GGAGGGGTGGGATAGGA	640
1	111,217,246	111,217,537	CAACAAATAATCCCCRAACACCA	388	TGGTTGTAATAGGYGGTGGGT	641
7	1,272,400	1,272,691	RTTTCCCTACACCCAACAC	389	TGGGTATGYGGGTGTTTTAGGA	642
11	125,774,055	125,774,347	GGGTTTTYGTTTGGAGGGTTGT	390	CATCCRCTAACCAATAAACTTCCTTAAA	643
7	24,323,635	24,323,929	GGGGTGTGGGTGGTTTT	391	CTCCCACCCCTAAACAAAC	644
X	136,656,432	136,656,726	RACCCCTCCAACCTTT	392	GTGGGTGTGGGAGGTTT	645
1	91,182,902	91,183,197	GGGAAGTGGTAATTTGTGGATA	393	rataccaaactataaacaaaccc	646
7	27,225,460	27,225,755	TGTTTGGTGGTTTGTTYGATTTGTA	394	CCCCTATACCTCTATCTCTACC	647
2	182,322,235	182,322,530	CTTCCCTCTCTCCTTCCTTTA	395	GTGTTGTGGGGGTTTTGG	648
2	63,283,941	63,284,236	CAAACTACTACCCCTTCAACT	396	GGGTGTTAAGATAAGATATGTTTAGT	649
11	20,618,153	20,618,450	TTTTTGATAAAGTAATTAAGGTYGTAGTGA	397	ACCAAACCCTAAAACAACAAAT	650
14	70,653,836	70,654,133	ATTTCTTAACCACCCAAAAACTTA	398	GGGTTAAATAGAAGAAATGTTTTTAATG	651
1	119,530,348	119,530,646	ATTAAGGTTTTAAYGGAGAAGGTATGT	399	RCTAATTCCTAACAACTAAACCAAC	652

17	47,073,268	47,073,566	CCCCTAACRCCATCCCAACC	400	GTTTTTGGGGAGGGGGGAA	653
8	23,567,042	23,567,341	GAAAAGATATTTTGTGGGgggatt	401	ACCCCTATCCCCAAATCTAC	654
17	47,073,229	47,073,528	YGAGGGGGTTTTAGGGT	402	TCCCTCRACTTTAACCAACTCA	655
2	198,650,889	198,651,188	ACCCTAAACCTACCACCTAAC	403	GTGGAGGAGAGGGGGAAG	656
13	109,147,667	109,147,966	TCCTCCTACTCCCAAAATCT	404	YGTGGGGTTGAGTTTTAGTT	657

Table 3

Table 3 below provides statistical information of exemplary assays involving all 150 MVPs as defined in Table 1 above, the top 3 ranked MVPs (SEQ ID NOs:1-3), the top 5 ranked MVPs (SEQ ID NOs:1-5) and the top 10 ranked MVPs (SEQ ID NOs:1-10).

	All 150	**Top 3**	**Top 5**	**Top 10**
Sensitivity	0.93	0.61	0.66	0.70
Specificity	0.97	0.71	0.74	0.80
PPV	0.98	0.79	0.82	0.86
NPV	0.89	0.51	0.55	0.61
AUC	0.95	0.66	0.70	0.75

References

1. Siegel R, Ward E, Brawley O, Jemal A: Cancer statistics, 2011: the impact of eliminating socioeconomic and racial disparities on premature cancer deaths. *CA Cancer J Clin* 2011, 61:212-236.

2. Jemal A, Siegel R, Ward E, Hao Y, Xu J, Thun MJ: Cancer statistics, 2009. *CA Cancer J Clin* 2009, 59:225-249.

3. Khadra MH, Pickard RS, Charlton M, Powell PH, Neal DE: A prospective analysis of 1,930 patients with hematuria to evaluate current diagnostic practice. *J Urol* 2000, 163:524-527.

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

5. Lyratzopoulos G, Abel GA, McPhail S, Neal RD, Rubin GP: Gender inequalities in the promptness of diagnosis of bladder and renal cancer after symptomatic presentation: evidence from secondary analysis of an English primary care audit survey. *BMJ Open* 2013, 3.

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

7. Denzinger S, Burger M, Walter B, Knuechel R, Roessler W, Wieland WF, Filbeck T: Clinically relevant reduction in risk of recurrence of superficial bladder cancer using 5-aminolevulinic acid-induced fluorescence diagnosis: 8-year results of prospective randomized study. *Urology* 2007, 69:675-679.

8. Zaak D, Kriegmair M, Stepp H, Stepp H, Baumgartner R, Oberneder R, Schneede P, Corvin S, Frimberger D, Knuchel R, Hofstetter A: Endoscopic detection of transitional cell carcinoma with 5-aminolevulinic acid: results of 1012 fluorescence endoscopies. *Urology* 2001, 57:690-694.

9. Schlake A, Crispen PL, Cap AP, Atkinson T, Davenport D, Preston DM: NMP-22, urinary cytology, and cystoscopy: a 1 year comparison study. *Can J Urol* 2012, 19:6345---6350.

10. Burns MB, Lackey L, Carpenter MA, Rathore A, Land AM, Leonard B, Refsland EW, Kotandeniya D, Tretyakova N, Nikas JB, et al: APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature* 2013, 494:366-370.

11. Kelly JD, Fawcett DP, Goldberg LC: Assessment and management of non-visible haematuria in primary care. *BMJ* 2009, 338:a3021.

12. Lotan Y, Roehrborn CG: Sensitivity and specificity of commonly available bladder tumor markers versus cytology: results of a comprehensive literature review and meta-analyses. *Urology* 2003, 61:109-118; discussion 118.

13. van Rhijn BW, van der Poel HG, van der Kwast TH: Urine markers for bladder cancer surveillance: a systematic review. *Eur Urol* 2005, 47:736-748.

14. Tilki D, Burger M, Dalbagni G, Grossman HB, Hakenberg OW, Palou J, Reich O, Roupret M, Shariat SF, Zlotta AR: Urine markers for detection and surveillance of non-muscle-invasive bladder cancer. *Eur Urol* 2011, 60:484-492.

15. Beukers W, Hercegovac A, Vermeij M, Kandimalla R, Blok AC, van der Aa MM, Zwarthoff EC, Zuiverloon TC: Hypermethylation of the polycomb group target gene PCDH7 in bladder tumors from patients of all ages. *J Urol* 2013, 190:311-316.

16. Hoque MO, Kim MS, Ostrow KL, Liu J, Wisman GB, Park HL, Poeta ML, Jeronimo C, Henrique R, Lendvai A, et al: Genome-wide promoter analysis uncovers portions of the cancer methylome. *Cancer Res* 2008, 68:2661---2670.

17. Kandimalla R, van Tilborg AA, Zwarthoff EC: DNA methylation-based biomarkers in bladder cancer. *Nat Rev Urol* 2013.

18. Kandimalla R, Masius R, Beukers W, Bangma CH, Orntoft TF, Dyrskjot L, van Leeuwen N, Lingsma H, van Tilborg AA, Zwarthoff EC: A 3-plex methylation assay combined with the FGFR3 mutation assay sensitively detects recurrent bladder cancer in voided urine. *Clin Cancer Res* 2013, 19:4760-4769.

19. Yu J, Zhu T, Wang Z, Zhang H, Qian Z, Xu H, Gao B, Wang W, Gu L, Meng J, et al: A novel set of DNA methylation markers in urine sediments for sensitive/specific detection of bladder cancer. *Clin Cancer Res* 2007, 13:7296-7304.

20. Friedrich MG, Weisenberger DJ, Cheng JC, Chandrasoma S, Siegmund KD, Gonzalgo 438 ML, Toma MI, Huland H, Yoo C, Tsai YC, et al: Detection of methylated apoptosis-associated genes in urine sediments of bladder cancer patients. *Clin Cancer Res* 440 2004, 10:7457-7465.

21. Hoque MO, Begum S, Topaloglu O, Chatterjee A, Rosenbaum E, Van Criekinge W, Westra WH, Schoenberg M, Zahurak M, Goodman SN, Sidransky D: Quantitation of promoter methylation of multiple genes in urine DNA and bladder cancer detection. *J Natl Cancer Inst* 2006, 98:996-1004.

22. Su SF, de Castro Abreu AL, Chihara Y, Tsai Y, Andreu-Vieyra C, Daneshmand S, Skinner EC, Jones PA, Siegmund KD, Liang G: A panel of three markers hyper- and hypomethylated in urine sediments accurately predicts bladder cancer recurrence. *Clin Cancer Res* 2014, 20:1978-1989.

23. Andersson E, Steven K, Guldberg P: Size-based enrichment of exfoliated tumor cells in urine increases the sensitivity for DNA-based detection of bladder cancer. *PLoS One* 2014, 9:e94023.

24. Fackler MJ, Lopez Bujanda Z, Umbricht C, Teo WW, Cho S, Zhang Z, Visvanathan K, Jeter S, Argani P, Wang C, et al: Novel methylated biomarkers and a robust assay to detect circulating tumor DNA in metastatic breast cancer. *Cancer Res* 2014, 74:2160-2170.

25. Philipp AB, Stieber P, Nagel D, Neumann J, Spelsberg F, Jung A, Lamerz R, Herbst A, Kolligs FT: Prognostic role of methylated free circulating DNA in colorectal cancer. *Int J Cancer* 2012, 131:2308-2319.

26. Campan M, Moffitt M, Houshdaran S, Shen H, Widschwendter M, Daxenbichler G, Long T, Marth C, Laird-Offringa IA, Press MF, et al: Genome-scale screen for DNA methylation-based detection markers for ovarian cancer. *PLoS One* 2011, 6:e28141.

27. Leng S, Do K, Yingling CM, Picchi MA, Wolf HJ, Kennedy TC, Feser WJ, Baron AE, Franklin WA, Brock MV, et al: Defining a gene promoter methylation signature in sputum for lung cancer risk assessment. *Clin Cancer Res* 2012, 18:3387-3395.

28. Belinsky SA: Gene-promoter hypermethylation as a biomarker in lung cancer. *Nat Rev Cancer* 2004, 4:707-717.

29. Kulis M, Esteller M: DNA methylation and cancer. *Adv Genet* 2010, 70:27 - 56. 469

30. Tewhey R, Warner JB, Nakano M, Libby B, Medkova M, David PH, Kotsopoulos SK, Samuels ML, Hutchison JB, Larson JW, et al: Microdroplet-based

PCR enrichment for large-scale targeted sequencing. *Nat Biotechnol* 2009, 27:1025-1031.

31. Paul DS, Guilhamon P, Karpathakis A, Butcher LM, Thirlwell C, Feber A, Beck S: Assessment of RainDrop BS-seq as a method for large-scale, targeted bisulfite sequencing. *Epigenetics* 2014, 9.

32. Guilhamon P, Eskandarpour M, Halai D, Wilson GA, Feber A, Teschendorff AE, Gomez V, Hergovich A, Tirabosco R, Fernanda Amary M, et al: Meta-analysis of IDH-mutant cancers identifies EBF1 as an interaction partner for TET2. *Nat Commun* 2013, 4:2166.

33. Komori HK, LaMere SA, Torkamani A, Hart GT, Kotsopoulos S, Warner J, Samuels ML, Olson J, Head SR, Ordoukhanian P, et al: Application of microdroplet PCR for large-scale targeted bisulfite sequencing. *Genome Res* 2011, 21:1738-1745.

34. Lowe R, Rakyan VK: Marmal-aid - a database for Infinium HumanMethylation450. *BMC Bioinformatics* 2013, 14:359.

35. R: A language and environment for statistical computing. 485

36. Morris TJ, Butcher LM, Feber A, Teschendorff AE, Chakravarthy AR, Wojdacz TK, Beck S: ChAMP: 450k Chip Analysis Methylation Pipeline. *Bioinformatics* 2013.

37. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, 488 Beck S: A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics* 2013, 29:189-196.

38. Morris TJ, Butcher LM, Feber A, Teschendorff AE, Chakravarthy AR, Wojdacz TK, Beck S: ChAMP: 450k Chip Analysis Methylation Pipeline. *Bioinformatics* 2014, 30:428-430.

39. Butcher LM, Beck S: Probe Lasso: A novel method to rope in differentially methylated regions with 450K DNA methylation data. *Methods* 2015, 72:21-28.

40. Deaton AM, Webb S, Kerr AR, Illingworth RS, Guy J, Andrews R, Bird A: Cell type-specific DNA methylation at intragenic CpG islands in the immune system. *Genome Res* 2011, 21:1074-1086.

41. Hajdinjak T: UroVysion FISH test for detecting urothelial cancers: meta-analysis of diagnostic accuracy and comparison with urinary cytology testing. *Urol Oncol* 2008, 26:646-651.

42. Gerlinger M, Catto JW, Orntoft TF, Real FX, Zwarthoff EC, Swanton C: Intratumour Heterogeneity in Urologic Cancers: From Molecular Evidence to Clinical Implications. *Eur Urol* 2014.

43. Hoffmann AC, Wild P, Leicht C, Bertz S, Danenberg KD, Danenberg PV, Stohr R, Stockle M, Lehmann J, Schuler M, Hartmann A: MDR1 and ERCC1 expression predict outcome of patients with locally advanced bladder cancer receiving adjuvant chemotherapy. *Neoplasia* 2010, 12:628-636.

44. Venkatesan A, Chu P, Kerlikowske K, Sickles EA, Smith---Bindman R: Positive predictive value of specific mammographic findings according to reader and patient variables. *Radiology* 2009, 250:648-657.

45. Mandel JS, Church TR, Bond JH, Ederer F, Geisser MS, Mongin SJ, Snover DC, Schuman LM: The effect of fecal occult-blood screening on the incidence of colorectal cancer. *N Engl J Med* 2000, 343:1603-1607.

46. Hoffman RM, Gilliland FD, Adams-Cameron M, Hunt WC, Key CR: Prostate-specific antigen testing accuracy in community practice. *BMC Fam Pract* 2002, 3:19.

47. Chung W, Bondaruk J, Jelinek J, Lotan Y, Liang S, Czerniak B, Issa JP: Detection of bladder cancer using novel DNA methylation biomarkers in urine sediments. *Cancer Epidemiol Biomarkers Prev* 2011, 20:1483-1491.

48. Catalona WJ, Smith DS, Ratliff TL, Dodds KM, Coplen DE, Yuan JJ, Petros JA, Andriole GL: Measurement of prostate-specific antigen in serum as a screening test for prostate cancer. *N Engl J Med* 1991, 324:1156-1161.

49. 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.

50. Wolff EM, Chihara Y, Pan F, Weisenberger DJ, Siegmund KD, Sugano K, Kawashima K, Laird PW, Jones PA, Liang G: Unique DNA methylation patterns distinguish noninvasive and invasive urothelial cancers and establish an epigenetic field defect in premalignant tissue. *Cancer Res* 2010, 70:8169-8178.

51. Olkhov-Mitsel, E and Bapat, B: Strategies for discovery and validation of methylated and hydroxymethylated DNA biomarkers. *Cancer Medicine* 2012, 1(2): 237–260.

52. Eng, J: Receiver Operating Characteristic Analysis: A Primer. *Academic Radiology* 2005, 12(7): 909-916.

53. Herman, J. G., Graff, J. R., Myohanen, S., Nelkin, B. D. & Baylin, S. B.: Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proc. Natl Acad. Sci. USA* 1996, 93: 9821–9826.

54. Frommer, M. et al.: A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl Acad. Sci. USA* 1992, 89: 1827–1831.

55. Xiong, Z. & Laird, P. W.: COBRA: a sensitive and quantitative DNA methylation assay. *Nucleic Acids Res.* 1997, 25: 2532–2534.

56. Gonzalgo, M. L. & Jones, P. A.: Rapid quantitation of methylation differences at specific sites using methylationsensitive single nucleotide primer extension (Ms-SNuPE). *Nucleic Acids Res.* 1997, 25: 2529–2531.

57. Singal, R. & Grimes, S. R.: Microsoft Word macro for analysis of cytosine methylation by the bisulfite deamination reaction. *Biotechniques* 2001, 30: 116–120.

58. Anbazhagan, R., Herman, J. G., Enika, K. & Gabrielson, E.: Spreadsheet-based program for the analysis of DNA methylation. *Biotechniques* 2001, 30: 110–114.

59. Li, L. C. & Dahiya, R.: MethPrimer: designing primers for methylation PCRs. *Bioinformatics* 2002, 18: 1427–1431.

60. Aronesty, E.: Comparison of Sequencing Utility Programs. *The Open Bioinformatics Journal*, 2013, 7: 1-8.

61. Krueger, F. & Andrews, S. R.: Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*, 2011, 27(11):1571-2.

CLAIMS

1. A method of diagnosing bladder cancer in an individual comprising:
   (a) providing DNA from a sample from the individual;
   (b) determining whether each one of a group of MVPs selected from a panel comprising the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG] is methylated, wherein the group comprises at least 25 of the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG]; and
   (c) diagnosing bladder cancer in the individual when at least 25 of the MVPs of the group of (b) are methylated.

2. The method according to claim 1, wherein the group of MVPs comprises at least 40 of the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG], and wherein bladder cancer is diagnosed when at least 25 of the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG] are methylated.

3. The method according to claim 2, wherein the group of MVPs comprises at least 50 of the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG], or comprises at least 100 of the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG].

4. The method according to claim 3, wherein the group of MVPs comprises all 150 of the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG].

5. The method according to any one of claims 2 to 4, wherein bladder cancer is diagnosed in the individual when at least 40 of the MVPs selected from the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG] are methylated, or when at least 50 of the MVPs selected from the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG] are methylated, or when at least 100 of the MVPs are methylated, or when all 150 MVPs are methylated.

6. The method according to any one of the preceding claims, wherein the MVPs determined to be methylated include the MVPs identified in SEQ ID NOS 1 to 3 and denoted by [CG], or include the MVPs identified in SEQ ID NOS 1 to 5 and denoted by [CG], or include the MVPs identified in SEQ ID NOS 1 to 10 and denoted by [CG], or include the MVPs identified in SEQ ID NOS 1 to 40 and denoted by [CG].

7. The method according to claim 1, wherein the group of MVPs comprises all 150 of the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG], wherein bladder cancer is diagnosed in the individual when at least 40 of the MVPs selected from the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG] are methylated, and wherein the MVPs determined to be methylated include the MVPs identified in SEQ ID NOS 1 to 10 and denoted by [CG].

8. The method according to any one of the preceding claims, wherein the step of determining whether each one the MVPs is methylated comprises bisulphite converting the DNA.

9. The method according to any one of claims 1 to 8, wherein the step of determining whether each one the MVPs is methylated comprises:

1) performing a sequencing step to determine the sequence of MVPs;
2) hybridising DNA to an array comprising probes capable of discriminating between methylated and non-methylated forms of MVPs and applying a detection system to the array to discriminate methylated and non-methylated forms of the MVPs; or
3) performing an amplification step using methylation-specific primers, wherein the status of an MVP as methylated or non-methylated is determined by the presence or absence of an amplified product.

10. The method according to claim 9(a) or 9(b), wherein before the sequencing or hybridization steps an amplification step is performed, wherein loci comprising each MVP are amplified.

11. The method according to claim 9(c) or claim 10, wherein the amplification step is performed by PCR.

12. The method according to claim 10 or claim 11, wherein a capturing step is performed before the sequencing or hybridization steps, and wherein the capturing step involves binding polynucleotides comprising the MVP loci to binding molecules specific to the MVP loci and collecting complexes comprising MVP loci and binding molecules; and wherein:

i. the capturing step occurs before the step of bisulphite converting the DNA;

ii. the capturing step occurs after the step of bisulphite converting the DNA but before the amplification or hybridization steps; or

iii. the capturing step occurs after the step of bisulphite converting the DNA and after the amplification step.

13. The method according to claim 12, wherein the binding molecules are oligonucleotides specific for each MVP, preferably DNA or RNA molecules each having a sequence which is complementary to the corresponding MVP.

14. The method according to any one of claims 11 to 13, wherein the binding molecule is coupled to a purification moiety.

15. The method according to claim 14, wherein the purification moiety comprises a first purification moiety and the step of collecting complexes comprising MVP loci and binding molecules comprises binding the first purification moiety to substrates comprising a second purification moiety, wherein first and second purification moieties form an interaction complex.

16. The method according to claim 15, wherein the first purification moiety is biotin and the second purification moiety is streptavidin; or wherein the first purification moiety is streptavidin and the second purification moiety is biotin.

17. The method according to any one of claims 10 to 16, wherein the step of amplifying loci comprising MVPs comprises the use of primers which are independent of the methylation status of the MVP.

18. The method according to any one of claims 3 to 17, wherein the step of amplifying loci comprising MVPs is performed by microdroplet PCR amplification.

19. The method according to any one of the preceding claims wherein the biological sample obtained from the individual is a sample of urine, blood, serum, plasma or cell-free DNA.

20. The method according to any one of the preceding claims, wherein the method achieves a ROC sensitivity of 95% or greater and a ROC specificity of 90% or greater; preferably a ROC sensitivity of 96% and a ROC specificity of 97%.

21. The method according to any one of claims 1 to 19, wherein the method achieves a ROC AUC of 95% or greater, preferably 98%.

22. The method according to any one of claims 1 to 19, wherein the method achieves a negative predictive value (NPV) of 95% or greater, preferably 97%.

23. The method according to any one of the preceding claims, wherein the step of diagnosing bladder cancer in the individual further comprises:

I. stratifying the grade of the tumor; and/or

II. determining the risk of recurrence of the tumor; and/or

III. determining the risk of progression of non-muscular invasive disease; and/or

determining the likely response to treatment therapy.

24. A method of treating bladder cancer in an individual comprising:

(a) obtaining DNA from a sample from the individual and determining whether each one of a group of MVPs selected from a panel comprising the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG] is methylated, wherein the group comprises at least 25 of the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG];

(b) diagnosing bladder cancer in the individual when at least 25 MVPs of the group of (a) are methylated; and

(c) administering one or more bladder cancer treatments to the individual.

25. A method of treating bladder cancer in an individual comprising:

(a) providing DNA from a sample from the individual and determining whether each one of a group of MVPs selected from a panel comprising the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG] is methylated, wherein the group comprises at least 25 of the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG];

(b) diagnosing bladder cancer in the individual when at least 25 MVPs of the group of (a) are methylated; and

(c) administering one or more bladder cancer treatments to the individual.

26. A method of treating bladder cancer in an individual comprising:

(a) determining whether each one of a group of MVPs selected from a panel comprising the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG] is methylated in DNA from a sample from the individual, wherein the group comprises at least 25 of the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG];

(b) diagnosing bladder cancer in the individual when at least 25 MVPs of the group of (a) are methylated; and

(c) administering one or more bladder cancer treatments to the individual.

27. A method of treating bladder cancer in an individual comprising administering one or more bladder cancer treatments to the individual, wherein the individual has been diagnosed with bladder cancer by steps comprising:

(a) providing DNA from a sample from the individual and determining whether each one of a group of MVPs selected from a panel comprising the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG] is methylated, wherein the group comprises at least 25 of the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG]; and

(b) diagnosing bladder cancer in the individual when at least 25 MVPs of the group of (a) are methylated.

28 A method of diagnosing bladder cancer in an individual comprising:

(a) obtaining data which identify whether each one of a group of MVPs selected from a panel comprising the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG] is methylated, wherein the group comprises at least 25 of the MVPs identified in SEQ ID NOS 1 to 150 and denoted by [CG]; and

(b) diagnosing bladder cancer in the individual when at least 25 MVPs of the group of (a) are methylated;

wherein the data were obtained by a method comprising:

i. obtaining DNA from the sample; and

ii. determining whether MVPs are methylated in the DNA.

29. The method according to any one of the preceding claims, wherein the cancer is a non-muscle invasive bladder cancer (NMIBC).

30. The method according to any one of claims 1 to 29, wherein the cancer is a muscle invasive bladder cancer (MIBC).

31. An array capable of discriminating between methylated and non-methylated forms of MVPs; the array comprising oligonucleotide probes specific for a methylated form of each MVP in a MVP panel and oligonucleotide probes specific for a non-

methylated form of each MVP in the panel; wherein the panel consists of at least 25 MVPs selected from the MVPs identified in SEQ ID NOS 1 to 150.

32. An array according to claim 31, provided that the array is not an Infinium HumanMethylation450 BeadChip array, and/or provided that the number of MVP-specific oligonucleotide probes of the array is less than 482,421, preferably 482,000 or less, 480,000 or less, 450,000 or less, 440,000 or less, 430,000 or less, 420,000 or less, 410,000 or less, or 400,000 or less.

33. An array according to claim 31 or claim 32, wherein the panel consists of at least 40 MVPs selected from the MVPs identified in SEQ ID NOS 1 to 150; preferably at least 50 MVPs, at least 60 MVPs, at least 70 MVPs, at least 80 MVPs, at least 90 MVPs, at least 100 MVPs, at least 110 MVPs, at least 120 MVPs, at least 130 MVPs, at least 140 MVPs, at least 145 MVPs, or all 150 MVPs identified in SEQ ID NOS 1 to 150.

34. An array according to any one of claims 31 to 33, wherein the panel includes the MVPs defined by SEQ ID NOS 1 to 3, or the MVPs defined by SEQ ID NOS 1 to 5, or the MVPs defined by SEQ ID NOS 1 to 10, or the MVPs defined by SEQ ID NOS 1 to 20, or the MVPs defined by SEQ ID NOS 1 to 30, or the MVPs defined by SEQ ID NOS 1 to 40, or the MVPs defined by SEQ ID NOS 1 to 50, or the MVPs defined by SEQ ID NOS 1 to 60, or the MVPs defined by SEQ ID NOS 1 to 70, or the MVPs defined by SEQ ID NOS 1 to 80, or the MVPs defined by SEQ ID NOS 1 to 90, or the MVPs defined by SEQ ID NOS 1 to 100, or the MVPs defined by SEQ ID NOS 1 to 100, or the MVPs defined by SEQ ID NOS 1 to 120, or the MVPs defined by SEQ ID NOS 1 to 130, or the MVPs defined by SEQ ID NOS 1 to 140, or the MVPs defined by SEQ ID NOS 1 to 150.

35. An array according to claim 34, wherein the panel includes all MVPs defined by SEQ ID NOS 1 to 150.

36. An array according to any one of claims 31 to 35, further comprising one or more oligonucleotides comprising a MVP selected from any of the MVPs defined in SEQ ID NOS 1 to 150, wherein the one or more oligonucleotides are hybridized to corresponding oligonucleotide probes of the array.

37. An array according to claim 36, wherein the one or more oligonucleotides comprise at least 25 MVPs selected from the MVPs identified in SEQ ID NOS 1 to 150; preferably at least 50 MVPs, at least 60 MVPs, at least 70 MVPs, at least 80 MVPs, at least 90 MVPs, at least 100 MVPs, at least 110 MVPs, at least 120 MVPs, at least 130 MVPs, at least 140 MVPs, at least 145 MVPs, or all 150 MVPs identified in SEQ ID NOS 1 to 150.

38. An array according to claim 36 or 37; wherein the one or more oligonucleotides comprise the MVPs defined by SEQ ID NOS 1 to 10, or the MVPs defined by SEQ ID NOS 1 to 20, or the MVPs defined by SEQ ID NOS 1 to 30, or the MVPs defined by SEQ ID NOS 1 to 40, or the MVPs defined by SEQ ID NOS 1 to 50, or the MVPs defined by SEQ ID NOS 1 to 60, or the MVPs defined by SEQ ID NOS 1 to 70, or the MVPs defined by SEQ ID NOS 1 to 80, or the MVPs defined by SEQ ID NOS 1 to 90, or the MVPs defined by SEQ ID NOS 1 to 100, or the MVPs defined by SEQ ID NOS 1 to 110, or the MVPs defined by SEQ ID NOS 1 to 120, or the MVPs defined by SEQ ID NOS 1 to 130, or the MVPs defined by SEQ ID NOS 1 to 140, or the MVPs defined by SEQ ID NOS 1 to 150.

39. An array according to claim 38; wherein the one or more oligonucleotides comprise all MVPs defined by SEQ ID NOS 1 to 150.

40. A hybridized array, wherein the array is obtainable by hybridizing to an array according to any one of claims 31 to 35 a group of oligonucleotides each comprising a different MVP selected from any of the MVPs defined in SEQ ID NOS 1 to 150, and wherein the group comprises at least 25 oligonucleotides.

41. A hybridized array according to claim 40, wherein the group comprises at least 40 oligonucleotides; optionally wherein the group comprises at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 110, at least 120, at least 130, at least 140, at least 145, or at least 150 oligonucleotides.

42. A hybridized array according to claim 40 or claim 41, wherein the group comprises at least 25 oligonucleotides comprising the MVPs defined by SEQ ID NOS 1 to 25, or wherein the group comprises at least 40 oligonucleotides comprising the MVPs defined by SEQ ID NOS 1 to 40, or wherein the group comprises at least 50 oligonucleotides comprising the MVPs defined by SEQ ID NOS 1 to 50, or wherein the group comprises at least 60 oligonucleotides comprising the MVPs defined by SEQ ID NOS 1 to 60, or wherein the group comprises at least 70 oligonucleotides comprising the MVPs defined by SEQ ID NOS 1 to 70, or wherein the group comprises at least 80 oligonucleotides comprising the MVPs defined by SEQ ID NOS 1 to 80, or wherein the group comprises at least 90 oligonucleotides comprising the MVPs defined by SEQ ID NOS 1 to 90, or wherein the group comprises at least 100 oligonucleotides comprising the MVPs defined by SEQ ID NOS 1 to 100, or wherein the group comprises at least 110 oligonucleotides comprising the MVPs defined by SEQ ID NOS 1 to 110, or wherein the group comprises at least 120 oligonucleotides comprising the MVPs defined by SEQ ID NOS 1 to 120, or wherein the group comprises at least 130 oligonucleotides comprising the MVPs defined by SEQ ID NOS 1 to 130, or wherein the group comprises at least 140 oligonucleotides comprising the MVPs defined by SEQ ID NOS 1 to 140, or wherein the group comprises at least 145 oligonucleotides comprising the MVPs defined by SEQ ID NOS 1 to 145, or wherein the group comprises at least 150 oligonucleotides comprising the MVPs defined by SEQ ID NOS 1 to 150.

43. A hybridized array according to claim 42, wherein the group comprises at least the 150 oligonucleotides comprising the MVPs defined by SEQ ID NOS 1 to 150.

44. A process for making the hybridized array according to claim 40, comprising contacting an array according to any one of claims 31 to 35 with a group of oligonucleotides each comprising a different MVP selected from any of the MVPs defined in SEQ ID NOS 1 to 150, and wherein the group comprises at least 25 oligonucleotides.

45. A process for making the hybridized array according to claim 41, comprising contacting an array according to any one of claims 31 to 35 with a group of oligonucleotides as defined according to claim 41; or a process for making the hybridized array according to claim 42 or 43, comprising contacting an array according to any one of claims 31 to 35 with a group of oligonucleotides as defined according to either claim 42 or claim 43.

46. A kit comprising an array according to any one of claims 31 to 35.

47. The kit according to claim 46, further comprising a DNA modifying regent that is capable of modifying a non-methylated cytosine in a MVP dinucleotide but is not capable of modifying a methylated cytosine in a MVP dinucleotide, optionally wherein the dinucleotide is CpG.

48. The kit according to claim 46 or 47, wherein the DNA modifying regent is a bisulphite reagent.

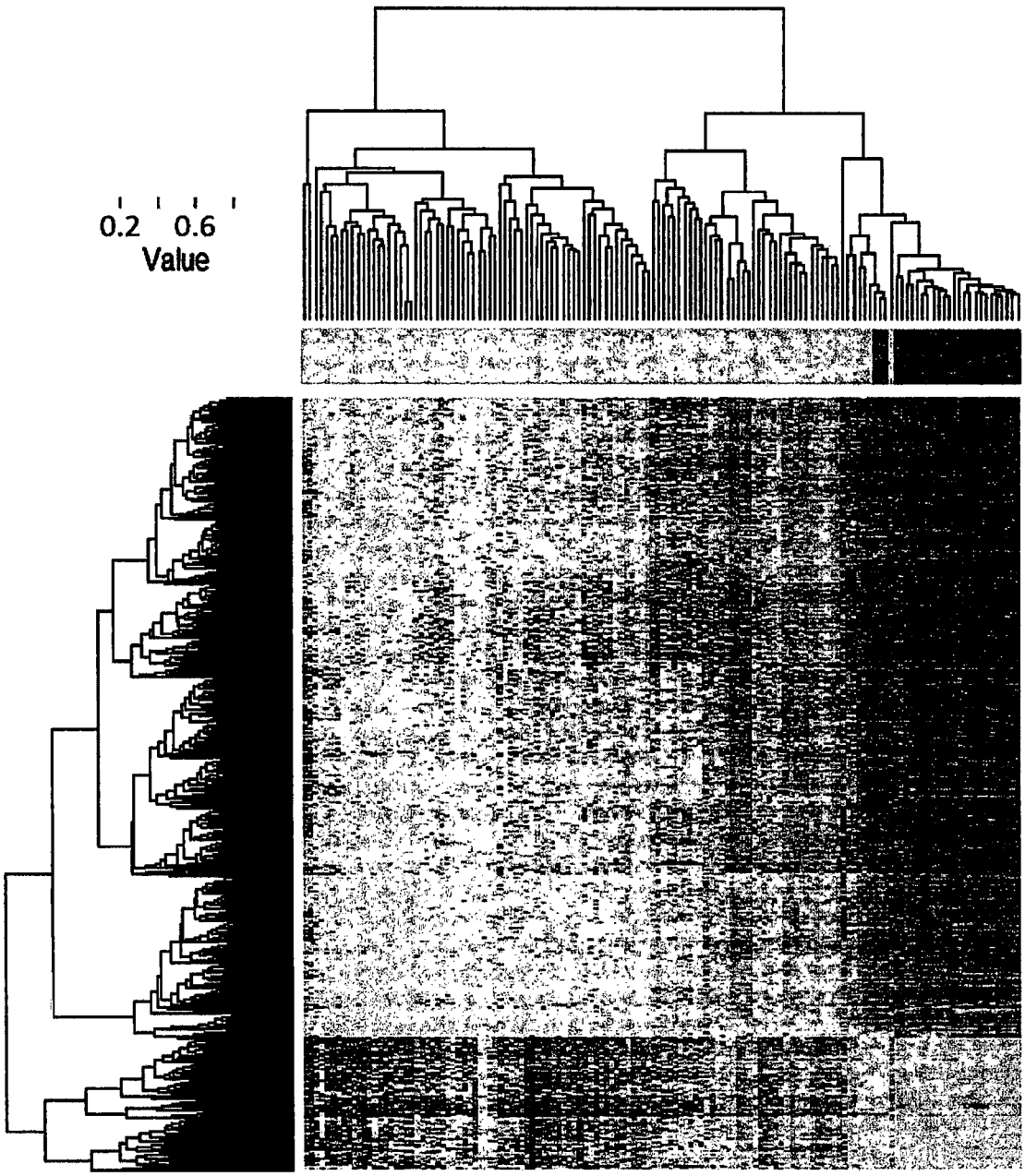

Figure 1

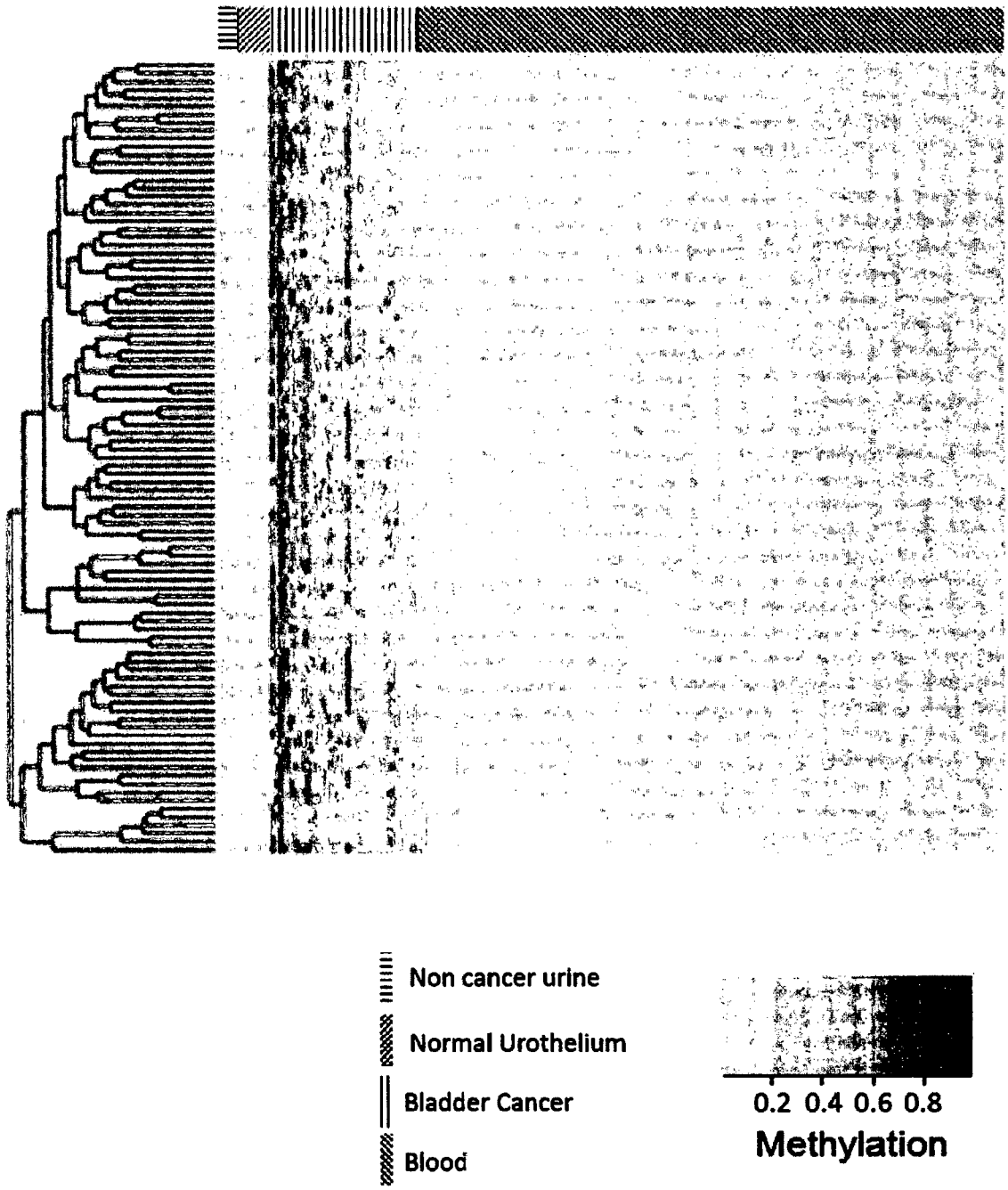
Non cancer urine
Normal Urothelium
Bladder Cancer
Blood
0.2 0.4 0.6 0.8
Methylation

Figure 2

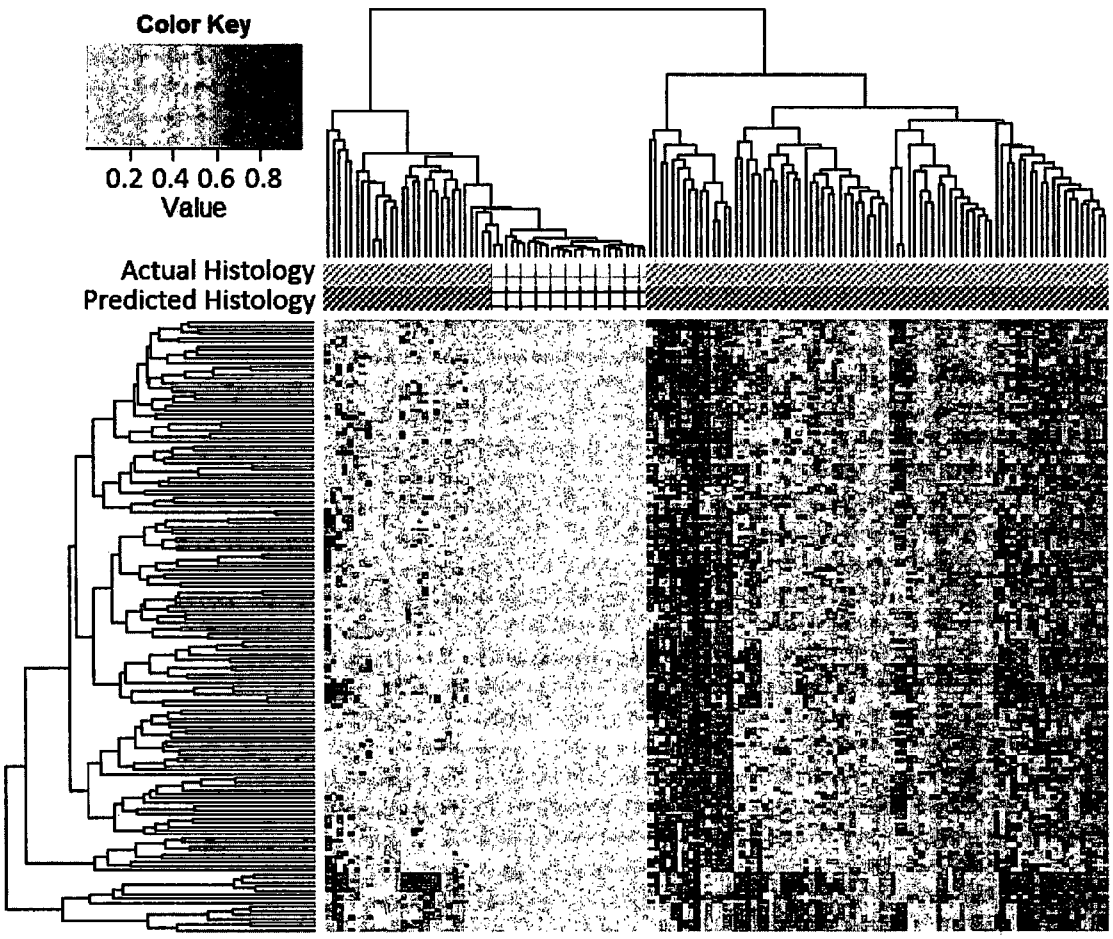
Color Key
0.2 0.4 0.6 0.8
Value
Actual Histology
Predicted Histology

Figure 3

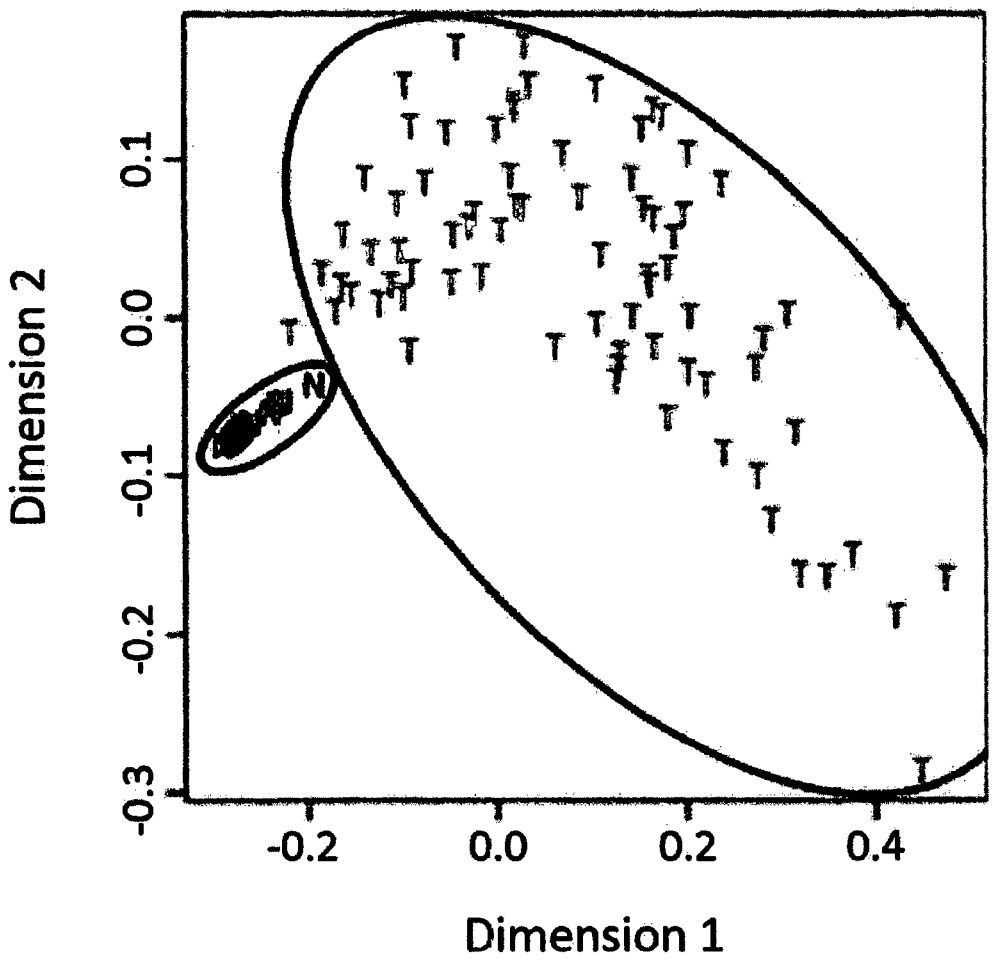

Figure 4

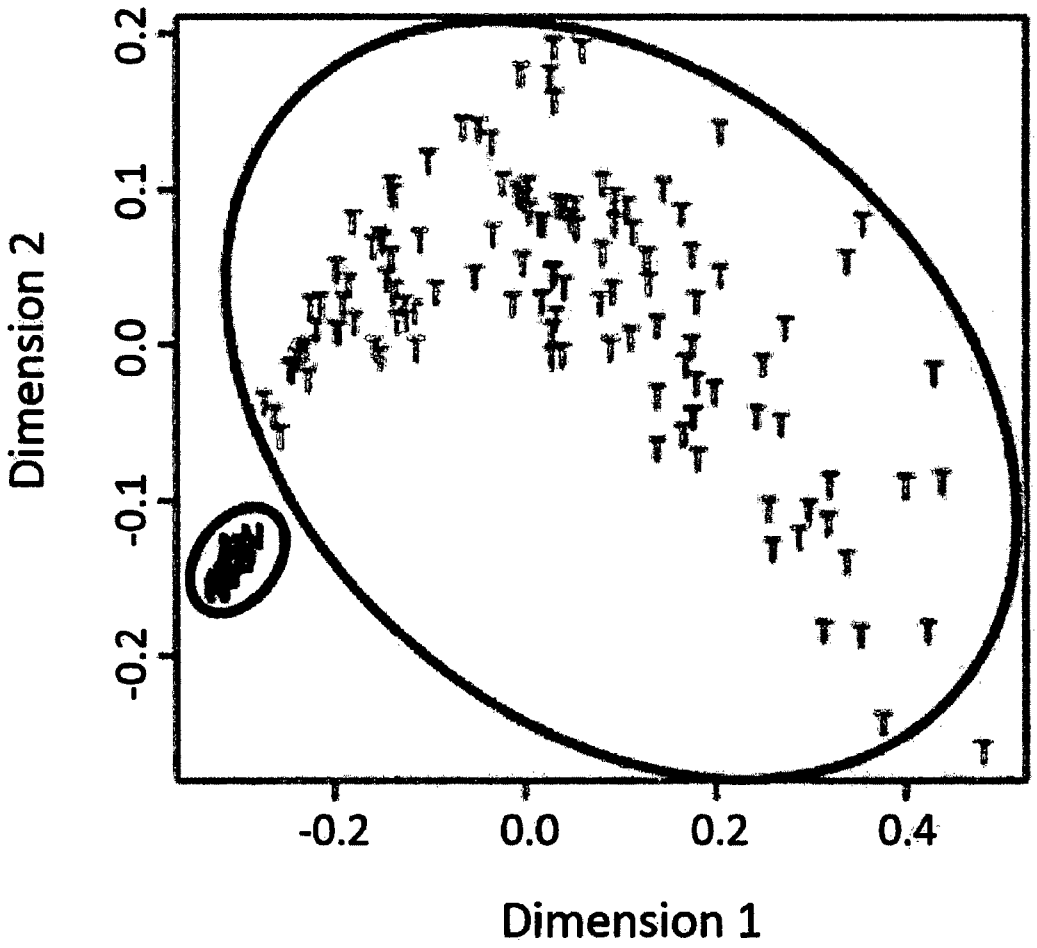

Figure 4 (continued)

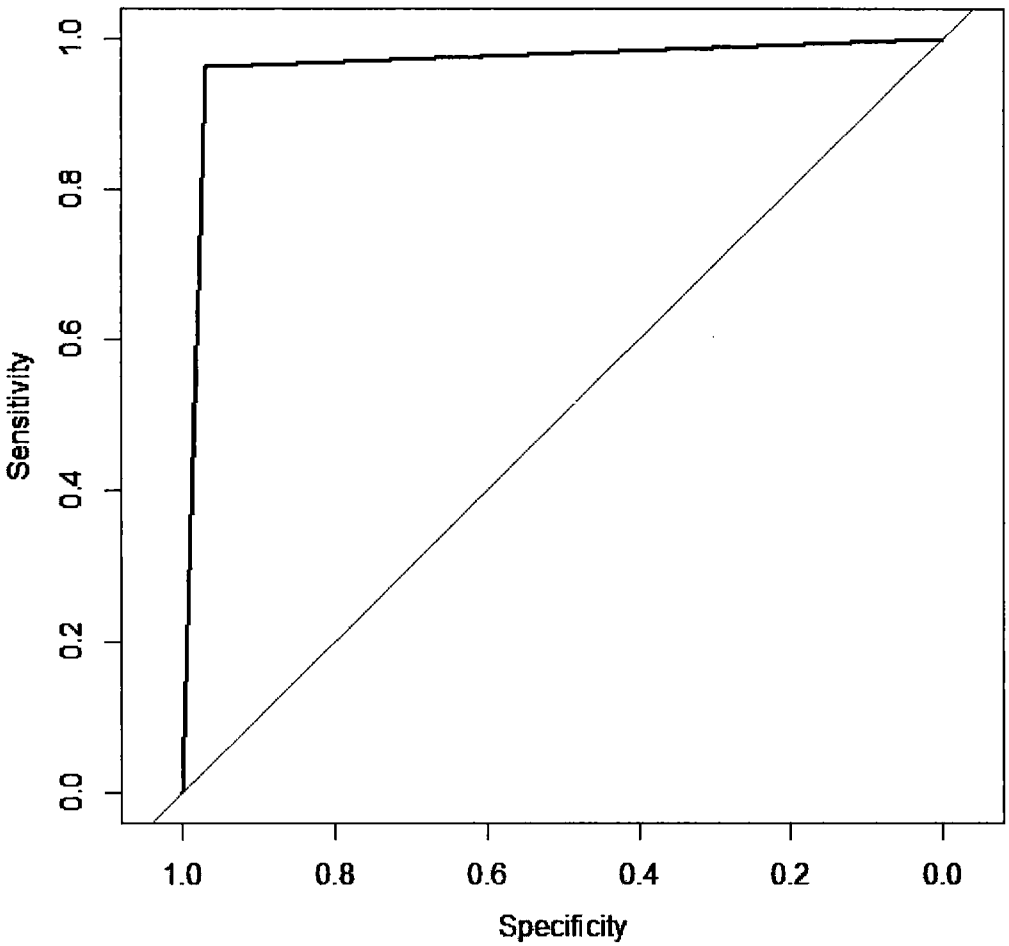

Figure 5

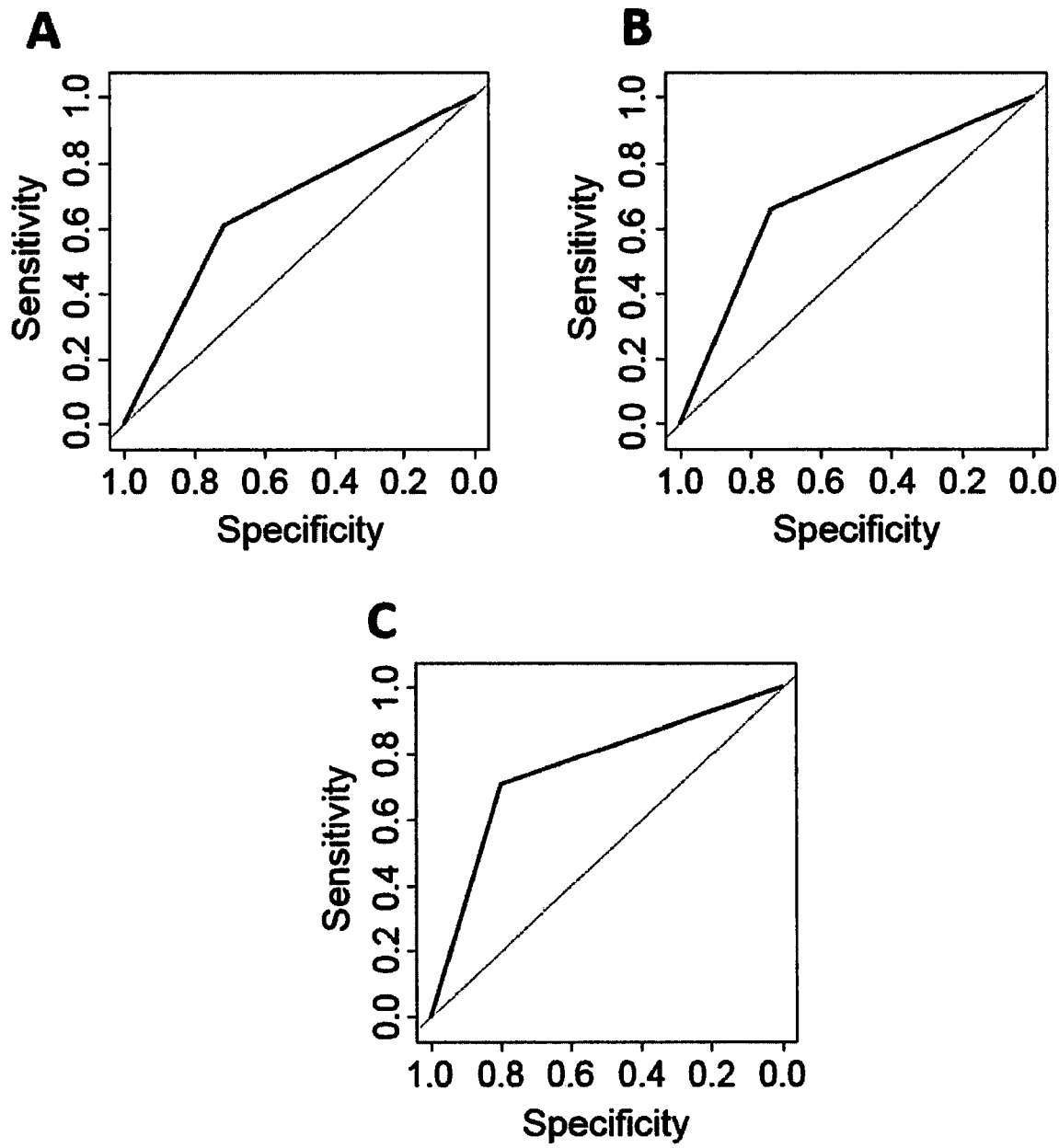
A
Sensitivity
0.0 0.2 0.4 0.6 0.8 1.0
1.0 0.8 0.6 0.4 0.2 0.0
Specificity
B
Sensitivity
0.0 0.2 0.4 0.6 0.8 1.0
1.0 0.8 0.6 0.4 0.2 0.0
Specificity
C
Sensitivity
0.0 0.2 0.4 0.6 0.8 1.0
1.0 0.8 0.6 0.4 0.2 0.0
Specificity

Figure 6

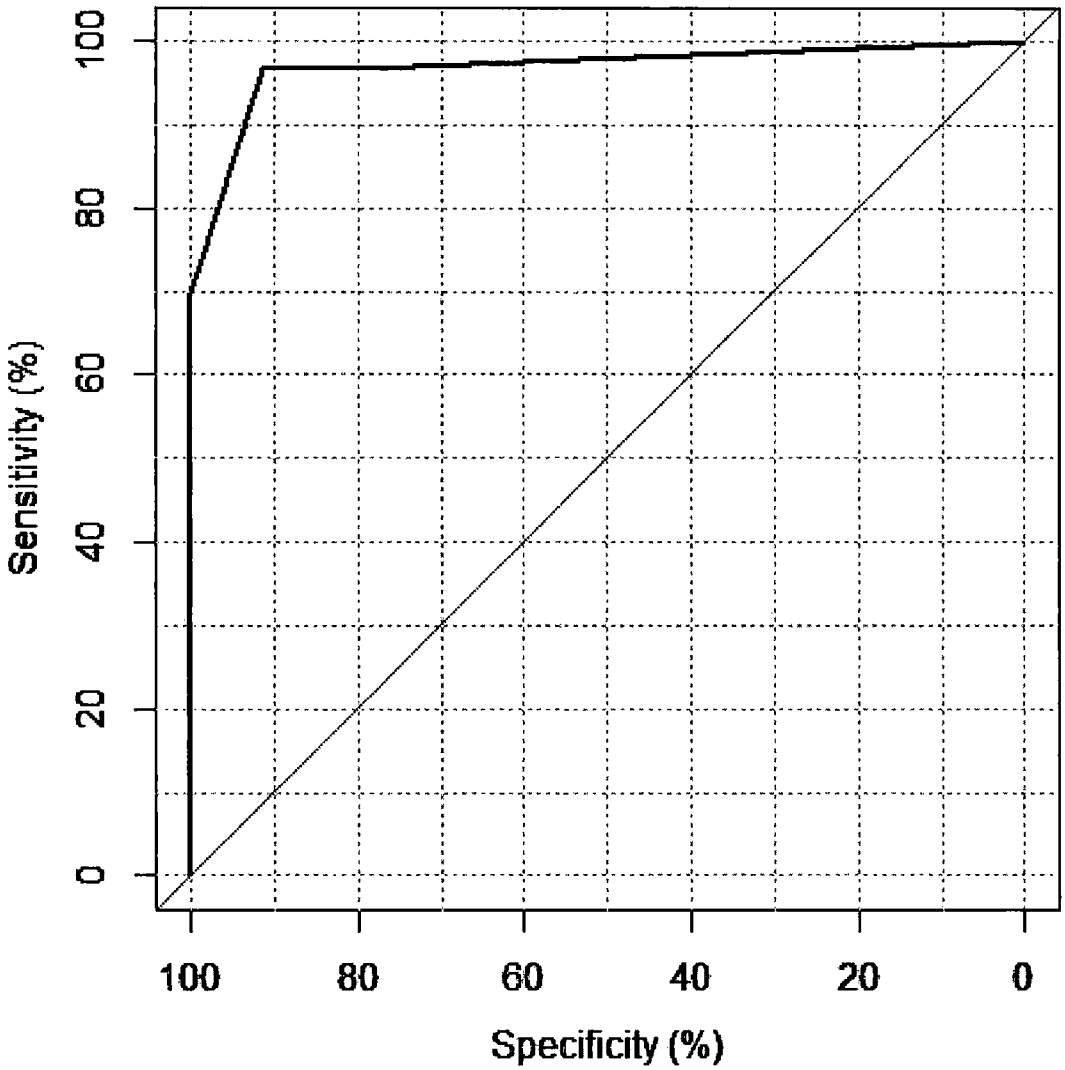

Figure 7

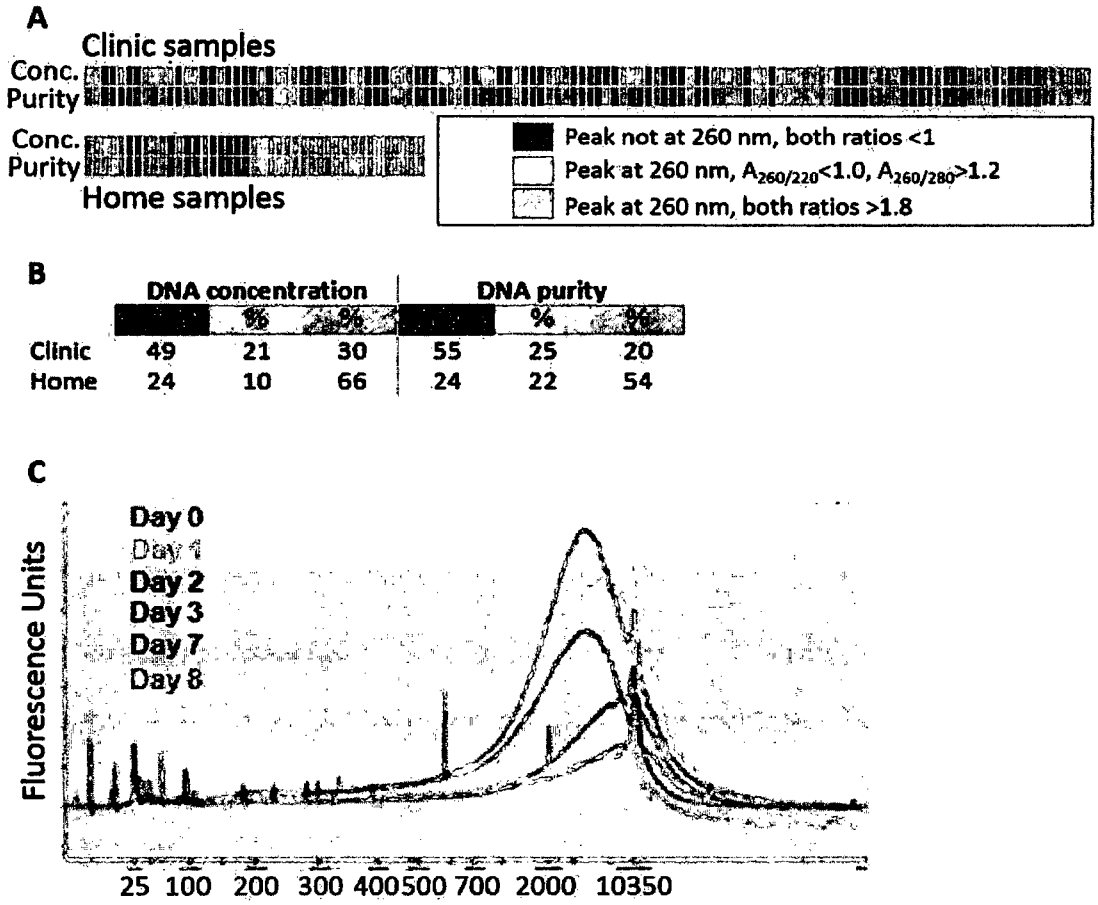
A
Clinic samples
Conc.
Purity
Conc.
Purity
Home samples
Peak not at 260 nm, both ratios <1
Peak at 260 nm, A260/220<1.0, A260/280>1.2
Peak at 260 nm, both ratios >1.8
B
DNA concentration
DNA purity
Clinic 49 21 30 55 25 20
Home 24 10 66 24 22 54
C
Day 0
Day 1
Day 2
Day 3
Day 7
Day 8
Fluorescence Units
25 100 200 300 400 500 700 2000 10350

Figure 8

D

Digest protocol	n	Yield				Purity	
		Median conc. (ng/µL)	Lower quartile (ng/µL)	Upper quartile (ng/µL)	Range (ng/µL)	$A_{260/230}$>1.79 (% of samples)	$A_{260/280}$>1.79 (% of samples)
1 h at 56°C	43	16.8	8.2	105.3	3.8-482.9	30	40
48 h at 21°C	35	29.9	14.2	55.6	4.8-264.7	43	89

E

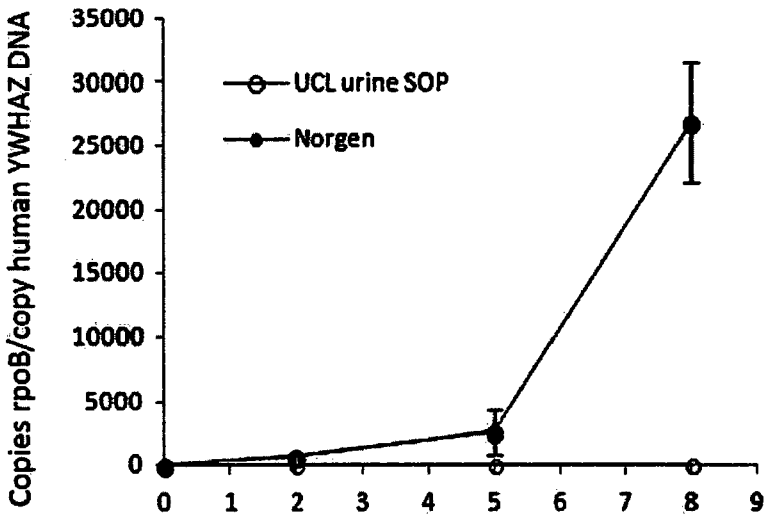

Figure 8 (continued)

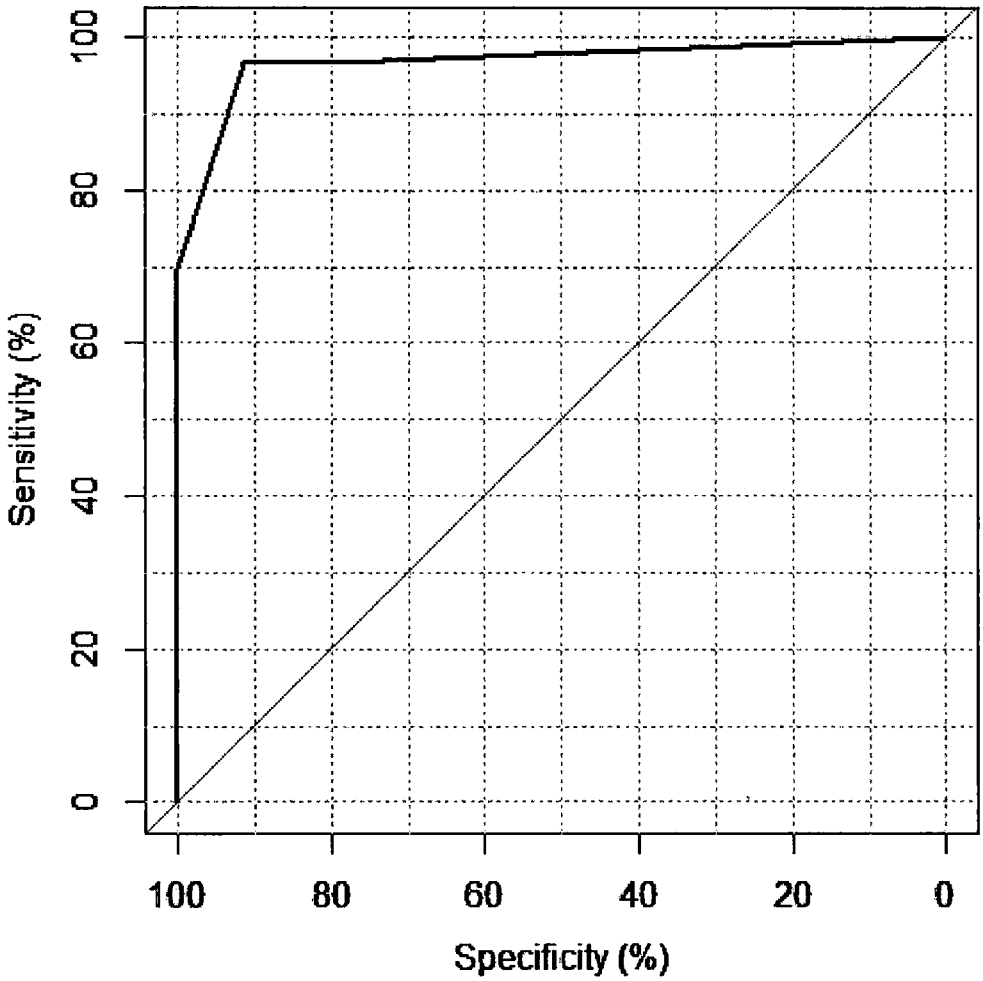

Figure 9

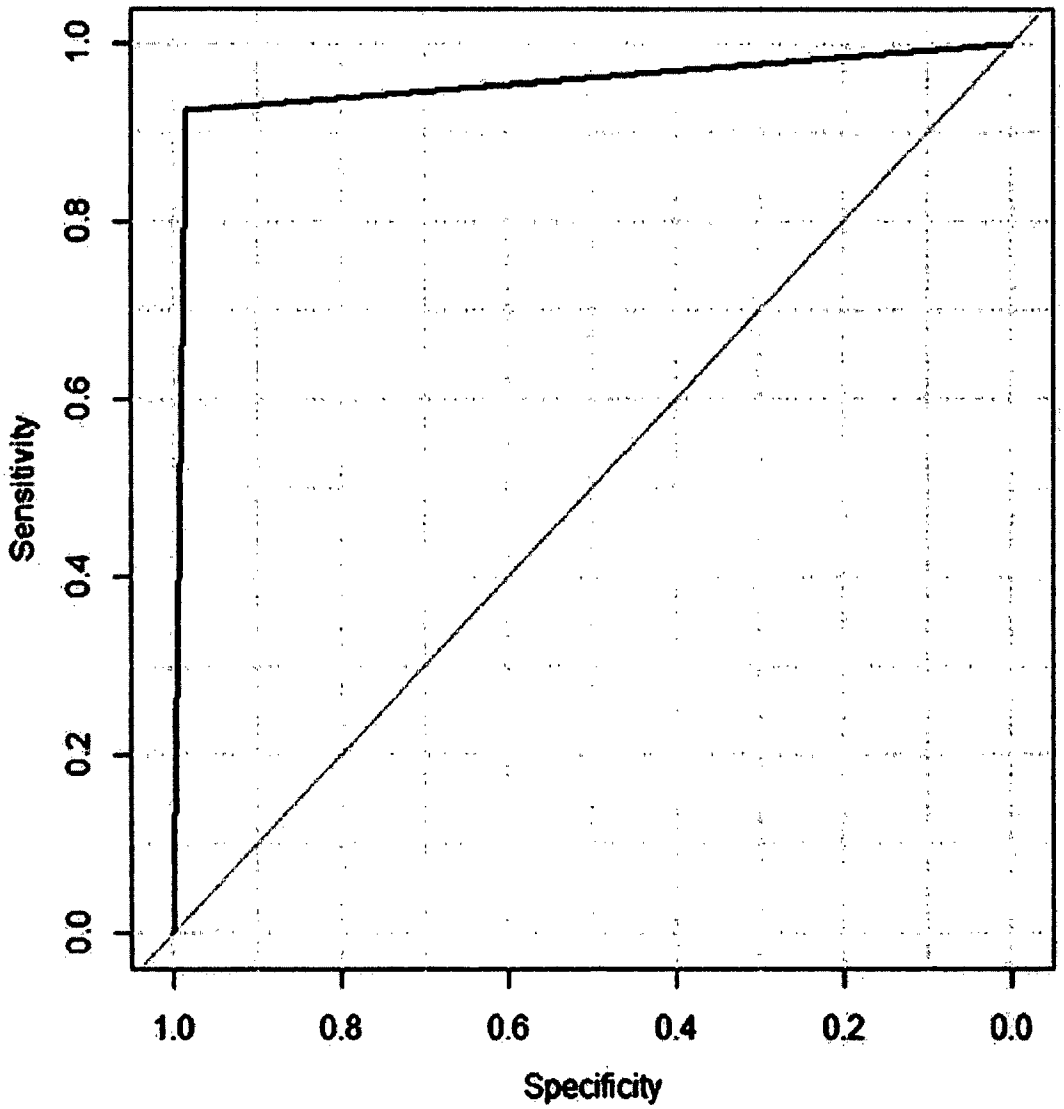

Figure 10

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2016/051903

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12Q1/68
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, BIOSIS, Sequence Search, EMBASE, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2010/123354 A2 (UNIV ERASMUS MEDICAL CT [NL]; ZWARTHOFF ELLEN CATHARINA [NL]; VAN TILB) 28 October 2010 (2010-10-28) page 157 - page 158; claims 1,2 page 155; table 12a -----	1-48
X	ANDREW FEBER ET AL: "Using high-density DNA methylation arrays to profile copy number alterations", GENOME BIOLOGY, BIOMED CENTRAL LTD., LONDON, GB, vol. 15, no. 2, 3 February 2014 (2014-02-03), page R30, XP021185518, ISSN: 1465-6906, DOI: 10.1186/GB-2014-15-2-R30 page 4; figure 3 ----- -/--	31-48

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
13 September 2016	21/09/2016

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Reuter, Uwe

2

Form PCT/ISA/210 (second sheet) (April 2005)

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	T. REINERT ET AL: "Comprehensive Genome Methylation Analysis in Bladder Cancer: Identification and Validation of Novel Methylated Genes and Application of These as Urinary Tumor Markers", CLINICAL CANCER RESEARCH, vol. 17, no. 17, 25 July 2011 (2011-07-25) , pages 5582-5592, XP055299578, US ISSN: 1078-0432, DOI: 10.1158/1078-0432.CCR-10-2659 the whole document -----	1-48
A	YOSHITOMO CHIHARA ET AL: "Diagnostic markers of urothelial cancer based on DNA methylation analysis", BMC CANCER, BIOMED CENTRAL, LONDON, GB, vol. 13, no. 1, 4 June 2013 (2013-06-04), page 275, XP021153060, ISSN: 1471-2407, DOI: 10.1186/1471-2407-13-275 the whole document -----	1-48

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/GB2016/051903

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2010123354 A2	28-10-2010	CA 2759312 A1	28-10-2010
		EP 2421988 A2	29-02-2012
		EP 3048176 A2	27-07-2016
		US 2012101023 A1	26-04-2012
		WO 2010123354 A2	28-10-2010