



(51) International Patent Classification:

C12Q 1/6886 (2018.01)

(21) International Application Number:

PCT/EP2020/056605

(22) International Filing Date:

12 March 2020 (12.03.2020)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

19305300.6 13 March 2019 (13.03.2019) EP

(71) Applicants: INSERM (INSTITUT NATIONAL DE LA SANTÉ ET DE LA RECHERCHE MÉDICALE)

[FR/FR]; 101, rue de Tolbiac, 75013 Paris (FR). UNIVERSITÉ DE BORDEAUX [FR/FR]; 35 Place Pey Berland, 33000 BORDEAUX (FR). CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (CNRS) [FR/FR]; 3, Rue Michel Ange, 75016 Paris (FR).

(72) Inventors: BIKFALVI, Andreas; Université de Bordeaux, U1029 - LAMC, Allée Geoffroy Saint Hilaire, 33615 PESSAC Cedex (FR). COOLEY, Lindsay; Université de Bordeaux, U1029, LAMC - ACML, Campus de Talence, Allée Geoffroy Saint Hilaire, Bât. B2, R-d-C, 33615 PESSAC Cedex (FR). RUDEWICZ, Justine; Université de Bordeaux, Centre de Génomique Fonctionnel, Centre de Bioinformatique de Bordeaux 146, rue Léo Saignat Bordeaux cedex., 33076 Bordeaux cedex (FR). NIKOLSKI, Marie; CNRS - LaBRI, UMR 5800 351 cours de la Libération, 33405 TALENCE CEDEX (FR). SOULEYREAU, Wilfried; Université de Bordeaux, U1029, Campus de Talence, Allée Geoffroy Saint Hilaire, Bât. B2, RdC, 33615 PESSAC Cedex (FR).

(74) Agent: INSERM TRANSFERT; 7 rue Watt, 75013 Paris (FR).

(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, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP,

KR, KW, 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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, 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))

(54) Title: NEW GENE SIGNATURES FOR PREDICTING SURVIVAL TIME IN PATIENTS SUFFERING FROM RENAL CELL CARCINOMA

(57) Abstract: Renal Cell Carcinoma (RCC) encompasses a heterogeneous group of cancers derived from renal tubular epithelial cells and has a worldwide mortality. However, mortality rates have barely improved over the last 20 years. Novel biomarkers and biomarkers are thus urgently required for this cancer. The inventors have devised a strategy to produce mouse cancer cell lines of progressively enhanced aggressiveness and specialization. The mouse renal cancer cell line RENCA was serially passaged in vivo using multiple implantation strategies designed to replicate different aspects of primary tumour growth and metastasis. Transcriptomic and epigenomic data has been acquired for the derived cell lines and primary analyses have been performed. The inventors then selected plurality of genes with no reported role in RCC, and checked their relevance in patient data samples. This approach contributes to identify several gene signatures that are suitable for predicting survival time in patients suffering from RCC.



NEW GENE SIGNATURES FOR PREDICTING SURVIVAL TIME IN PATIENTS SUFFERING FROM RENAL CELL CARCINOMA

FIELD OF THE INVENTION:

The present relates to new gene signatures that are suitable for predicting survival time in patients suffering from renal cell carcinoma (RCC).

BACKGROUND OF THE INVENTION:

Renal Cell Carcinoma (RCC) encompasses a heterogeneous group of cancers derived from renal tubular epithelial cells and has a worldwide mortality of over 140,000 people per year. The disease encompasses multiple histological and molecular subtypes, of which clear cell RCC (ccRCC) is the most common. The incidence and prevalence of RCC are rising, along with increases in related risk factors such as hypertension, diabetes and obesity¹. However, mortality rates have barely improved over the last 20 years according to Surveillance, Epidemiology and End Results (SEER) data. Gandaglia et al² report a continuing upward trend in both incidence and mortality even in patients with localized disease. These data are in stark contrast with markedly improving survival rates in many other cancers and highlights RCC as one of the cancers in which current therapeutic approaches have failed to make the advances hoped for. Novel approaches to this problem are thus urgently required. When disease is localized to the kidney, surgical resection is the preferred option. However, therapeutic options for metastatic disease are limited. ccRCC metastasizes primarily to the lungs (secondarily to liver and bone), and 5 year survival is less than 10%^{3,4,5}. Furthermore, 40% of patients with seemingly localized disease will also relapse later with localized or metastatic disease. Localised recurrence is also difficult to treat, difficult to predict, and has a poor prognosis^{6,7}.

The challenges associated with treatment of RCC include high levels of resistance to traditional chemotherapeutic drugs⁸. The majority of currently available targeted therapies focus on inhibiting angiogenesis driven by the VEGF/VEGFR axis⁹. While progress has been made at extending life somewhat, such therapies are rarely curative, and will act primarily to “inhibit” the disease making eventual drug resistance almost inevitable. The high cost and failure rate of this approach is significant. Second line treatments include mTOR inhibitors and immunotherapeutic agents¹⁰. These can be successful, but are effective for only a limited subset of patients¹¹. The pathophysiology of RCC still far from understood, and there is a clear need to identify key mechanisms in RCC progression in order to open up novel therapeutic avenues

targeting different aspects of RCC biology. Furthermore, clinical treatment of RCC is hampered by a lack of relevant biomarkers. Currently, no fully validated molecular biomarkers for RCC are in clinical practice. Response to currently available treatments and long term disease free survival is highly variable and problematic to predict. Patient diagnosis, prognosis, and clinical decisions are currently made based on histological information such as Fuhrman grade and tumour stage. Therapy selection is based on limited guidelines and response to previous treatments. In this respect, clinical treatment of RCC lags behind other cancers for which molecular knowledge is invaluable in guiding clinical decisions e.g. hormone receptor status in breast cancer.

SUMMARY OF THE INVENTION:

The present relates to new gene signatures that are suitable for predicting survival time in patients suffering from renal cell carcinoma (RCC). In particular, the present invention is defined by the claims.

DETAILED DESCRIPTION OF THE INVENTION:

An object of the present invention relates to a method for predicting the survival time of a patient suffering from a renal cell carcinoma (RCC) comprising determining the expression level of at least one gene in a sample obtained from the patient wherein said gene is selected from the group consisting of ATP1A3, AVIL, CCL8, CES2, CFB, COL6A3, CRABP2, DNAJC12, F3, FAIM2, IL2RG, IL34, KRT13, KRT17, LRG1, MAGEB1, MAGEB2, MAGEC2, MAGEC3, MMP3, NAT8, PLAC, PPARGC1A, PPEF1, PTGIR, PTPN14, RAC2, SAA1, SAA2, SAMS1, SERPINA3, SLC11A1, SLC2A3, TPRG1, TSPAN32, UCK2, WAS, WT1, BASP1, DPYSL3, GLIPR2, HMGA2, RAB15, SNAI2 and PCBP3.

In some embodiments, the method of the present invention comprising determining the expression levels of KRT13, LRG1, SAA2, SAA1, SAMS1, SERPINA3, SLC2A3, UCK2, and WT1 in the sample obtained from the patient.

In some embodiments, the method of the present invention comprising determining the expression levels of NAT8, PPARGC1A, TPRG1, and WAS in the sample obtained from the patient.

In some embodiments, the method of the present invention comprising determining the expression level of ATP1A3 in the sample obtained from the patient.

In some embodiments, the method of the present invention comprising determining the expression levels of AVIL, CES2, FAIM2, KRT17, PTGIR, and RAC2 in the sample obtained from the patient.

5 In some embodiments, the method of the present invention comprising determining the expression levels of MAGEC2, MAGEC3, CCL8, CFB, COL6A3, DNAJC12, IL2RG, IL34, and PLAC8 in the sample obtained from the patient.

In some embodiments, the method of the present invention comprising determining the expression levels of BASP1, DPYSL3, GLIPR2, HMGA2, RAB15, SNAI2, and PCBP3 in the sample obtained from the patient.

10 As used herein, the term “renal cell carcinoma” or “RCC” has its general meaning in the art and refers to a cancer originated from the renal tubular epithelial cells in the kidney. According to the pathological features, the cancer is classified into clear cell type, granular cell type, chromophobe type, spindle type, cyst-associated type, cyst-originating type, cystic type, or papillary type. In some embodiments, the renal cell carcinoma (RCC) is at Stage
15 I, II, III, or IV as determined by the TNM classification, but however the present invention is accurately useful for predicting the survival time of patients when said cancer has been classified as Stage II or III by the TNM classification, i.e. non metastatic renal cell carcinoma (RCC).

The method of the present invention is particularly suitable for predicting the duration
20 of the overall survival (OS), progression-free survival (PFS) and/or the disease-free survival (DFS) of the cancer patient. Those of skill in the art will recognize that OS survival time is generally based on and expressed as the percentage of people who survive a certain type of cancer for a specific amount of time. Cancer statistics often use an overall five-year survival rate. In general, OS rates do not specify whether cancer survivors are still undergoing treatment
25 at five years or if they've become cancer-free (achieved remission). DSF gives more specific information and is the number of people with a particular cancer who achieve remission. Also, progression-free survival (PFS) rates (the number of people who still have cancer, but their disease does not progress) includes people who may have had some success with treatment, but the cancer has not disappeared completely. As used herein, the expression “short survival time”
30 indicates that the patient will have a survival time that will be lower than the median (or mean) observed in the general population of patients suffering from said cancer. When the patient will have a short survival time, it is meant that the patient will have a “poor prognosis”. Inversely, the expression “long survival time” indicates that the patient will have a survival time that will be higher than the median (or mean) observed in the general population of patients suffering

from said cancer. When the patient will have a long survival time, it is meant that the patient will have a “good prognosis”.

In some embodiments, the sample is a tissue tumor sample. The term “tumor tissue sample” means any tissue tumor sample derived from the patient. Said tissue sample is obtained for the purpose of the in vitro evaluation. In some embodiments, the tumor sample may result from the tumor resected from the patient. In some embodiments, the tumor sample may result from a biopsy performed in the primary tumour of the patient or performed in metastatic sample distant from the primary tumor of the patient. In some embodiments, the tumor tissue sample encompasses (i) a global primary tumor (as a whole), (ii) a tissue sample from the centre of the tumor, (iii) lymphoid islets in close proximity with the tumor, (iv) the lymph nodes located at the closest proximity of the tumor, (v) a tumor tissue sample collected prior surgery (for follow-up of patients after treatment for example), and (vi) a distant metastasis. In some embodiments, the tumor tissue sample, encompasses pieces or slices of tissue that have been removed from the tumor, including following a surgical tumor resection or following the collection of a tissue sample for biopsy, for further quantification of several expression levels of different genes, notably through histology or immunohistochemistry methods, through flow cytometry methods and through methods of gene or protein expression analysis, including genomic and proteomic analysis. The tumor tissue sample can, of course, be subjected to a variety of well-known post-collection preparative and storage techniques (e.g., fixation, storage, freezing, etc.). The sample can be fresh, frozen, fixed (e.g., formalin fixed), or embedded (e.g., paraffin embedded).

In the present specification, the name of each of the genes of interest (i.e. ATP1A3, AVIL, CCL8, CES2, CFB, COL6A3, CRABP2, DNAJC12, F3, FAIM2, IL2RG, IL34, KRT13, KRT17, LRG1, MAGEB1, MAGEB2, MAGEC2, MAGEC3, MMP3, NAT8, PLAC, PPARGC1A, PPEF1, PTGIR, PTPN14, RAC2, SAA1, SAA2, SAMSN1, SERPINA3, SLC11A1, SLC2A3, TPRG1, TSPAN32, UCK2, WAS, DPYSL3, GLIPR2, HMGA2, RAB15, SNAI2, PCBP3 and WT1) refers to the internationally recognized name of the corresponding gene, as found in internationally recognized gene sequences and protein sequences databases, including in the database from the HUGO Gene Nomenclature Committee that is available notably at the following Internet address: <http://www.gene.ucl.ac.uk/nomenclature/index.html>.

In the present specification, the name of each of the genes of interest may also refer to the internationally recognized name of the corresponding gene, as found in the internationally recognized gene sequences and protein sequences database Genbank. Through these internationally recognized sequence databases, the nucleic acid and the amino acid sequences

corresponding to each of the gene of interest described herein may be retrieved by the one skilled in the art.

As used herein, the term "expression level" refers, e.g., to a determined level of expression of gene of interest. The expression level of expression indicates the amount of expression product in a sample. The expression product of a gene of interest can be the nucleic acid of interest itself, a nucleic acid transcribed or derived therefrom, or the a polypeptide or protein derived therefrom.

In some embodiments, the expression level of the gene is determined at nucleic acid level. Typically, the expression level of a gene may be determined by determining the quantity of mRNA. Methods for determining the quantity of mRNA are well known in the art. For example the nucleic acid contained in the samples (e.g., cell or tissue prepared from the subject) is first extracted according to standard methods, for example using lytic enzymes or chemical solutions or extracted by nucleic-acid-binding resins following the manufacturer's instructions. The extracted mRNA is then detected by hybridization (e. g., Northern blot analysis, in situ hybridization) and/or amplification (e.g., RT-PCR). Other methods of Amplification include ligase chain reaction (LCR), transcription-mediated amplification (TMA), strand displacement amplification (SDA) and nucleic acid sequence based amplification (NASBA).

Nucleic acids having at least 10 nucleotides and exhibiting sequence complementarity or homology to the mRNA of interest herein find utility as hybridization probes or amplification primers. It is understood that such nucleic acids need not be identical, but are typically at least about 80% identical to the homologous region of comparable size, more preferably 85% identical and even more preferably 90-95% identical. In some embodiments, it will be advantageous to use nucleic acids in combination with appropriate means, such as a detectable label, for detecting hybridization.

Typically, the nucleic acid probes include one or more labels, for example to permit detection of a target nucleic acid molecule using the disclosed probes. In various applications, such as in situ hybridization procedures, a nucleic acid probe includes a label (e.g., a detectable label). A "detectable label" is a molecule or material that can be used to produce a detectable signal that indicates the presence or concentration of the probe (particularly the bound or hybridized probe) in a sample. Thus, a labeled nucleic acid molecule provides an indicator of the presence or concentration of a target nucleic acid sequence (e.g., genomic target nucleic acid sequence) (to which the labeled uniquely specific nucleic acid molecule is bound or hybridized) in a sample. A label associated with one or more nucleic acid molecules (such as a probe generated by the disclosed methods) can be detected either directly or indirectly. A label

can be detected by any known or yet to be discovered mechanism including absorption, emission and/ or scattering of a photon (including radio frequency, microwave frequency, infrared frequency, visible frequency and ultra-violet frequency photons). Detectable labels include colored, fluorescent, phosphorescent and luminescent molecules and materials, catalysts (such as enzymes) that convert one substance into another substance to provide a detectable difference (such as by converting a colorless substance into a colored substance or vice versa, or by producing a precipitate or increasing sample turbidity), haptens that can be detected by antibody binding interactions, and paramagnetic and magnetic molecules or materials.

Particular examples of detectable labels include fluorescent molecules (or fluorochromes). Numerous fluorochromes are known to those of skill in the art, and can be selected, for example from Life Technologies (formerly Invitrogen), e.g., see, The Handbook—A Guide to Fluorescent Probes and Labeling Technologies). Examples of particular fluorophores that can be attached (for example, chemically conjugated) to a nucleic acid molecule (such as a uniquely specific binding region) are provided in U.S. Pat. No. 5,866,366 to Nazarenko et al., such as 4-acetamido-4'-isothiocyanatostilbene-2,2'-disulfonic acid, acridine and derivatives such as acridine and acridine isothiocyanate, 5-(2'-aminoethyl) aminonaphthalene-1-sulfonic acid (EDANS), 4-amino-N-[3 vinylsulfonyl]phenyl]naphthalimide-3,5 disulfonate (Lucifer Yellow VS), N-(4-anilino-1-naphthyl)maleimide, antl1ranilamide, Brilliant Yellow, coumarin and derivatives such as coumarin, 7-amino-4-methylcoumarin (AMC, Coumarin 120), 7-amino-4-trifluoromethylcoumarin (Coumarin 151); cyanosine; 4',6-diarninidino-2-phenylindole (DAPI); 5',5"-dibromopyrogallol-sulfonephthalein (Bromopyrogallol Red); 7-diethylamino-3-(4'-isothiocyanatophenyl)-4-methylcoumarin; diethylenetriamine pentaacetate; 4,4'-diisothiocyanatodihydro-stilbene-2,2'-disulfonic acid; 4,4'-diisothiocyanatostilbene-2,2'-disulfonic acid; 5-[dimethylamino] naphthalene-1-sulfonyl chloride (DNS, dansyl chloride); 4-(4'-dimethylaminophenylazo)benzoic acid (DABCYL); 4-dimethylaminophenylazophenyl-4'-isothiocyanate (DABITC); eosin and derivatives such as eosin and eosin isothiocyanate; erythrosin and derivatives such as erythrosin B and erythrosin isothiocyanate; ethidium; fluorescein and derivatives such as 5-carboxyfluorescein (FAM), 5-(4,6-dichlorotriazin-2-yl)arninofluorescein (DTAF), 2',7'-dimethoxy-4',5'-dichloro-6-carboxyfluorescein (JOE), fluorescein, fluorescein isothiocyanate (FITC), and QFITC Q(RITC); 2',7'-difluorofluorescein (OREGON GREEN®); fluorescamine; IR144; IR1446; Malachite Green isothiocyanate; 4-methylumbelliferone; ortho cresolphthalein; nitrotyrosine; pararosaniline; Phenol Red; B-

phycoerythrin; o-phthalaldehyde; pyrene and derivatives such as pyrene, pyrene butyrate and succinimidyl 1-pyrene butyrate; Reactive Red 4 (Cibacron Brilliant Red 3B-A); rhodamine and derivatives such as 6-carboxy-X-rhodamine (ROX), 6-carboxyrhodamine (R6G), lissamine rhodamine B sulfonyl chloride, rhodamine (Rhod), rhodamine B, rhodamine 123, rhodamine X isothiocyanate, rhodamine green, sulforhodamine B, sulforhodamine 101 and sulfonyl chloride derivative of sulforhodamine 101 (Texas Red); N,N,N',N'-tetramethyl-6-carboxyrhodamine (TAMRA); tetramethyl rhodamine; tetramethyl rhodamine isothiocyanate (TRITC); riboflavin; rosolic acid and terbium chelate derivatives. Other suitable fluorophores include thiol-reactive europium chelates which emit at approximately 617 nm (Heyduk and Heyduk, *Analyt. Biochem.* 248:216-27, 1997; *J. Biol. Chem.* 274:3315-22, 1999), as well as GFP, LissamineTM, diethylaminocoumarin, fluorescein chlorotriazinyl, naphthofluorescein, 4,7-dichlororhodamine and xanthene (as described in U.S. Pat. No. 5,800,996 to Lee et al.) and derivatives thereof. Other fluorophores known to those skilled in the art can also be used, for example those available from Life Technologies (Invitrogen; Molecular Probes (Eugene, Oreg.)) and including the ALEXA FLUOR® series of dyes (for example, as described in U.S. Pat. Nos. 5,696,157, 6,130,101 and 6,716,979), the BODIPY series of dyes (dipyrrrometheneboron difluoride dyes, for example as described in U.S. Pat. Nos. 4,774,339, 5,187,288, 5,248,782, 5,274,113, 5,338,854, 5,451,663 and 5,433,896), Cascade Blue (an amine reactive derivative of the sulfonated pyrene described in U.S. Pat. No. 5,132,432) and Marina Blue (U.S. Pat. No. 5,830,912).

In addition to the fluorochromes described above, a fluorescent label can be a fluorescent nanoparticle, such as a semiconductor nanocrystal, e.g., a QUANTUM DOTTM (obtained, for example, from Life Technologies (QuantumDot Corp, Invitrogen Nanocrystal Technologies, Eugene, Oreg.); see also, U.S. Pat. Nos. 6,815,064; 6,682,596; and 6,649,138). Semiconductor nanocrystals are microscopic particles having size-dependent optical and/or electrical properties. When semiconductor nanocrystals are illuminated with a primary energy source, a secondary emission of energy occurs of a frequency that corresponds to the bandgap of the semiconductor material used in the semiconductor nanocrystal. This emission can be detected as colored light of a specific wavelength or fluorescence. Semiconductor nanocrystals with different spectral characteristics are described in e.g., U.S. Pat. No. 6,602,671. Semiconductor nanocrystals that can be coupled to a variety of biological molecules (including dNTPs and/or nucleic acids) or substrates by techniques described in, for example, Bruchez et al., *Science* 281:2013-2016, 1998; Chan et al., *Science* 281:2016-2018, 1998; and U.S. Pat. No. 6,274,323. Formation of semiconductor nanocrystals of various compositions are disclosed in,

e.g., U.S. Pat. Nos. 6,927, 069; 6,914,256; 6,855,202; 6,709,929; 6,689,338; 6,500,622; 6,306,736; 6,225,198; 6,207,392; 6,114,038; 6,048,616; 5,990,479; 5,690,807; 5,571,018; 5,505,928; 5,262,357 and in U.S. Patent Publication No. 2003/0165951 as well as PCT Publication No. 99/26299 (published May 27, 1999). Separate populations of semiconductor nanocrystals can be produced that are identifiable based on their different spectral characteristics. For example, semiconductor nanocrystals can be produced that emit light of different colors based on their composition, size or size and composition. For example, quantum dots that emit light at different wavelengths based on size (565 nm, 655 nm, 705 nm, or 800 nm emission wavelengths), which are suitable as fluorescent labels in the probes disclosed herein are available from Life Technologies (Carlsbad, Calif.).

Additional labels include, for example, radioisotopes (such as ^3H), metal chelates such as DOTA and DPTA chelates of radioactive or paramagnetic metal ions like Gd^{3+} , and liposomes.

Detectable labels that can be used with nucleic acid molecules also include enzymes, for example horseradish peroxidase, alkaline phosphatase, acid phosphatase, glucose oxidase, beta-galactosidase, beta-glucuronidase, or beta-lactamase.

Alternatively, an enzyme can be used in a metallographic detection scheme. For example, silver in situ hybridization (SISH) procedures involve metallographic detection schemes for identification and localization of a hybridized genomic target nucleic acid sequence. Metallographic detection methods include using an enzyme, such as alkaline phosphatase, in combination with a water-soluble metal ion and a redox-inactive substrate of the enzyme. The substrate is converted to a redox-active agent by the enzyme, and the redoxactive agent reduces the metal ion, causing it to form a detectable precipitate. (See, for example, U.S. Patent Application Publication No. 2005/0100976, PCT Publication No. 2005/003777 and U.S. Patent Application Publication No. 2004/0265922). Metallographic detection methods also include using an oxido-reductase enzyme (such as horseradish peroxidase) along with a water soluble metal ion, an oxidizing agent and a reducing agent, again to form a detectable precipitate. (See, for example, U.S. Pat. No. 6,670,113).

Probes made using the disclosed methods can be used for nucleic acid detection, such as ISH procedures (for example, fluorescence in situ hybridization (FISH), chromogenic in situ hybridization (CISH) and silver in situ hybridization (SISH)) or comparative genomic hybridization (CGH).

In situ hybridization (ISH) involves contacting a sample containing target nucleic acid sequence (e.g., genomic target nucleic acid sequence) in the context of a metaphase or

interphase chromosome preparation (such as a cell or tissue sample mounted on a slide) with a labeled probe specifically hybridizable or specific for the target nucleic acid sequence (e.g., genomic target nucleic acid sequence). The slides are optionally pretreated, e.g., to remove paraffin or other materials that can interfere with uniform hybridization. The sample and the probe are both treated, for example by heating to denature the double stranded nucleic acids. The probe (formulated in a suitable hybridization buffer) and the sample are combined, under conditions and for sufficient time to permit hybridization to occur (typically to reach equilibrium). The chromosome preparation is washed to remove excess probe, and detection of specific labeling of the chromosome target is performed using standard techniques.

For example, a biotinylated probe can be detected using fluorescein-labeled avidin or avidin-alkaline phosphatase. For fluorochrome detection, the fluorochrome can be detected directly, or the samples can be incubated, for example, with fluorescein isothiocyanate (FITC)-conjugated avidin. Amplification of the FITC signal can be effected, if necessary, by incubation with biotin-conjugated goat antiavidin antibodies, washing and a second incubation with FITC-conjugated avidin. For detection by enzyme activity, samples can be incubated, for example, with streptavidin, washed, incubated with biotin-conjugated alkaline phosphatase, washed again and pre-equilibrated (e.g., in alkaline phosphatase (AP) buffer). For a general description of in situ hybridization procedures, see, e.g., U.S. Pat. No. 4,888,278.

Numerous procedures for FISH, CISH, and SISH are known in the art. For example, procedures for performing FISH are described in U.S. Pat. Nos. 5,447,841; 5,472,842; and 5,427,932; and for example, in Pir1kel et al., *Proc. Natl. Acad. Sci.* 83:2934-2938, 1986; Pinkel et al., *Proc. Natl. Acad. Sci.* 85:9138-9142, 1988; and Lichter et al., *Proc. Natl. Acad. Sci.* 85:9664-9668, 1988. CISH is described in, e.g., Tanner et al., *Am. J. Pathol.* 157:1467-1472, 2000 and U.S. Pat. No. 6,942,970. Additional detection methods are provided in U.S. Pat. No. 6,280,929.

Numerous reagents and detection schemes can be employed in conjunction with FISH, CISH, and SISH procedures to improve sensitivity, resolution, or other desirable properties. As discussed above probes labeled with fluorophores (including fluorescent dyes and QUANTUM DOTS®) can be directly optically detected when performing FISH. Alternatively, the probe can be labeled with a nonfluorescent molecule, such as a hapten (such as the following non-limiting examples: biotin, digoxigenin, DNP, and various oxazoles, pyrrazoles, thiazoles, nitroaryls, benzofurazans, triterpenes, ureas, thioureas, rotenones, coumarin, coumarin-based compounds, Podophyllotoxin, Podophyllotoxin-based compounds, and combinations thereof), ligand or other indirectly detectable moiety. Probes labeled with such non-fluorescent

molecules (and the target nucleic acid sequences to which they bind) can then be detected by contacting the sample (e.g., the cell or tissue sample to which the probe is bound) with a labeled detection reagent, such as an antibody (or receptor, or other specific binding partner) specific for the chosen hapten or ligand. The detection reagent can be labeled with a fluorophore (e.g.,
5 QUANTUM DOT®) or with another indirectly detectable moiety, or can be contacted with one or more additional specific binding agents (e.g., secondary or specific antibodies), which can be labeled with a fluorophore.

In other examples, the probe, or specific binding agent (such as an antibody, e.g., a primary antibody, receptor or other binding agent) is labeled with an enzyme that is capable of
10 converting a fluorogenic or chromogenic composition into a detectable fluorescent, colored or otherwise detectable signal (e.g., as in deposition of detectable metal particles in SISH). As indicated above, the enzyme can be attached directly or indirectly via a linker to the relevant probe or detection reagent. Examples of suitable reagents (e.g., binding reagents) and chemistries (e.g., linker and attachment chemistries) are described in U.S. Patent Application
15 Publication Nos. 2006/0246524; 2006/0246523, and 2007/ 01 17153.

It will be appreciated by those of skill in the art that by appropriately selecting labelled probe-specific binding agent pairs, multiplex detection schemes can be produced to facilitate detection of multiple target nucleic acid sequences (e.g., genomic target nucleic acid sequences) in a single assay (e.g., on a single cell or tissue sample or on more than one cell or tissue
20 sample). For example, a first probe that corresponds to a first target sequence can be labelled with a first hapten, such as biotin, while a second probe that corresponds to a second target sequence can be labelled with a second hapten, such as DNP. Following exposure of the sample to the probes, the bound probes can be detected by contacting the sample with a first specific binding agent (in this case avidin labelled with a first fluorophore, for example, a first spectrally
25 distinct QUANTUM DOT®, e.g., that emits at 585 nm) and a second specific binding agent (in this case an anti-DNP antibody, or antibody fragment, labelled with a second fluorophore (for example, a second spectrally distinct QUANTUM DOT®, e.g., that emits at 705 nm). Additional probes/binding agent pairs can be added to the multiplex detection scheme using other spectrally distinct fluorophores. Numerous variations of direct, and indirect (one step, two
30 step or more) can be envisioned, all of which are suitable in the context of the disclosed probes and assays.

Probes typically comprise single-stranded nucleic acids of between 10 to 1000 nucleotides in length, for instance of between 10 and 800, more preferably of between 15 and 700, typically of between 20 and 500. Primers typically are shorter single-stranded nucleic

acids, of between 10 to 25 nucleotides in length, designed to perfectly or almost perfectly match a nucleic acid of interest, to be amplified. The probes and primers are “specific” to the nucleic acids they hybridize to, i.e. they preferably hybridize under high stringency hybridization conditions (corresponding to the highest melting temperature T_m , e.g., 50 % formamide, 5x or 6x SCC. SCC is a 0.15 M NaCl, 0.015 M Na-citrate).

The nucleic acid primers or probes used in the above amplification and detection method may be assembled as a kit. Such a kit includes consensus primers and molecular probes. A preferred kit also includes the components necessary to determine if amplification has occurred. The kit may also include, for example, PCR buffers and enzymes; positive control sequences, reaction control primers; and instructions for amplifying and detecting the specific sequences.

In some embodiments, the methods of the invention comprise the steps of providing total RNAs extracted from cumulus cells and subjecting the RNAs to amplification and hybridization to specific probes, more particularly by means of a quantitative or semi-quantitative RT-PCR.

In some embodiments, the level is determined by DNA chip analysis. Such DNA chip or nucleic acid microarray consists of different nucleic acid probes that are chemically attached to a substrate, which can be a microchip, a glass slide or a microsphere-sized bead. A microchip may be constituted of polymers, plastics, resins, polysaccharides, silica or silica-based materials, carbon, metals, inorganic glasses, or nitrocellulose. Probes comprise nucleic acids such as cDNAs or oligonucleotides that may be about 10 to about 60 base pairs. To determine the level, a sample from a test subject, optionally first subjected to a reverse transcription, is labelled and contacted with the microarray in hybridization conditions, leading to the formation of complexes between target nucleic acids that are complementary to probe sequences attached to the microarray surface. The labelled hybridized complexes are then detected and can be quantified or semi-quantified. Labelling may be achieved by various methods, e.g. by using radioactive or fluorescent labelling. Many variants of the microarray hybridization technology are available to the man skilled in the art (see e.g. the review by Hoheisel, Nature Reviews, Genetics, 2006, 7:200-210).

In some embodiments, the nCounter® Analysis system is used to detect intrinsic gene expression. The basis of the nCounter® Analysis system is the unique code assigned to each nucleic acid target to be assayed (International Patent Application Publication No. WO 08/124847, U.S. Patent No. 8,415,102 and Geiss et al. Nature Biotechnology. 2008. 26(3): 317-325; the contents of which are each incorporated herein by reference in their entireties). The code is composed of an ordered series of colored fluorescent spots which create a unique

barcode for each target to be assayed. A pair of probes is designed for each DNA or RNA target, a biotinylated capture probe and a reporter probe carrying the fluorescent barcode. This system is also referred to, herein, as the nanoreporter code system. Specific reporter and capture probes are synthesized for each target. The reporter probe can comprise at a least a first label

5 attachment region to which are attached one or more label monomers that emit light constituting a first signal; at least a second label attachment region, which is non-over-lapping with the first label attachment region, to which are attached one or more label monomers that emit light constituting a second signal; and a first target- specific sequence. Preferably, each sequence specific reporter probe comprises a target specific sequence capable of hybridizing to no more

10 than one gene and optionally comprises at least three, or at least four label attachment regions, said attachment regions comprising one or more label monomers that emit light, constituting at least a third signal, or at least a fourth signal, respectively. The capture probe can comprise a second target-specific sequence; and a first affinity tag. In some embodiments, the capture probe can also comprise one or more label attachment regions. Preferably, the first target-

15 specific sequence of the reporter probe and the second target- specific sequence of the capture probe hybridize to different regions of the same gene to be detected. Reporter and capture probes are all pooled into a single hybridization mixture, the "probe library". The relative abundance of each target is measured in a single multiplexed hybridization reaction. The method comprises contacting the tissue sample with a probe library, such that the presence of

20 the target in the sample creates a probe pair - target complex. The complex is then purified. More specifically, the sample is combined with the probe library, and hybridization occurs in solution. After hybridization, the tripartite hybridized complexes (probe pairs and target) are purified in a two-step procedure using magnetic beads linked to oligonucleotides complementary to universal sequences present on the capture and reporter probes. This dual

25 purification process allows the hybridization reaction to be driven to completion with a large excess of target-specific probes, as they are ultimately removed, and, thus, do not interfere with binding and imaging of the sample. All post hybridization steps are handled robotically on a custom liquid-handling robot (Prep Station, NanoString Technologies). Purified reactions are typically deposited by the Prep Station into individual flow cells of a sample cartridge, bound

30 to a streptavidin-coated surface via the capture probe, electrophoresed to elongate the reporter probes, and immobilized. After processing, the sample cartridge is transferred to a fully automated imaging and data collection device (Digital Analyzer, NanoString Technologies). The expression level of a target is measured by imaging each sample and counting the number of times the code for that target is detected. For each sample, typically 600 fields-of-view (FOV)

are imaged (1376 X 1024 pixels) representing approximately 10 mm² of the binding surface. Typical imaging density is 100- 1200 counted reporters per field of view depending on the degree of multiplexing, the amount of sample input, and overall target abundance. Data is output in simple spreadsheet format listing the number of counts per target, per sample. This
5 system can be used along with nanoreporters. Additional disclosure regarding nanoreporters can be found in International Publication No. WO 07/076129 and WO07/076132, and US Patent Publication No. 2010/0015607 and 2010/0261026, the contents of which are incorporated herein in their entireties. Further, the term nucleic acid probes and nanoreporters can include the rationally designed (e.g. synthetic sequences) described in International
10 Publication No. WO 2010/019826 and US Patent Publication No.2010/0047924, incorporated herein by reference in its entirety.

Level of a gene may be expressed as absolute level or normalized level. Typically, levels are normalized by correcting the absolute level of a gene by comparing its expression to the expression of a gene that is not a relevant for determining the risk. This normalization allows
15 the comparison of the level in one sample, e.g., a subject sample, to another sample, or between samples from different sources.

In some embodiments, a score which is a composite of the expression levels of the different genes is determined and compared to a predetermined reference value wherein a difference between said score and said predetermined reference value is indicative whether the
20 subject will have a long or short survival time.

The score can be calculated in any appropriate manner, such as principal components analysis, support vector machines, or other techniques known to the person of ordinary skill in the art having the benefit of the present disclosure.

In some embodiments, the predetermined reference value is a threshold value or a cut-off value. Typically, a "threshold value" or "cut-off value" can be determined experimentally, empirically, or theoretically. A threshold value can also be arbitrarily selected based upon the existing experimental and/or clinical conditions, as would be recognized by a person of ordinary skilled in the art. For example, retrospective measurement of the score in properly banked historical subject samples may be used in establishing the predetermined reference value. The
25 threshold value has to be determined in order to obtain the optimal sensitivity and specificity according to the function of the test and the benefit/risk balance (clinical consequences of false positive and false negative). Typically, the optimal sensitivity and specificity (and so the threshold value) can be determined using a Receiver Operating Characteristic (ROC) curve based on experimental data. For example, after determining the score in a group of reference,
30

one can use algorithmic analysis for the statistic treatment of the measured expression levels of the gene(s) in samples to be tested, and thus obtain a classification standard having significance for sample classification. The full name of ROC curve is receiver operator characteristic curve, which is also known as receiver operation characteristic curve. It is mainly used for clinical biochemical diagnostic tests. ROC curve is a comprehensive indicator that reflects the continuous variables of true positive rate (sensitivity) and false positive rate (1-specificity). It reveals the relationship between sensitivity and specificity with the image composition method. A series of different cut-off values (thresholds or critical values, boundary values between normal and abnormal results of diagnostic test) are set as continuous variables to calculate a series of sensitivity and specificity values. Then sensitivity is used as the vertical coordinate and specificity is used as the horizontal coordinate to draw a curve. The higher the area under the curve (AUC), the higher the accuracy of diagnosis. On the ROC curve, the point closest to the far upper left of the coordinate diagram is a critical point having both high sensitivity and high specificity values. The AUC value of the ROC curve is between 1.0 and 0.5. When AUC>0.5, the diagnostic result gets better and better as AUC approaches 1. When AUC is between 0.5 and 0.7, the accuracy is low. When AUC is between 0.7 and 0.9, the accuracy is moderate. When AUC is higher than 0.9, the accuracy is quite high. This algorithmic method is preferably done with a computer. Existing software or systems in the art may be used for the drawing of the ROC curve, such as: MedCalc 9.2.0.1 medical statistical software, SPSS 9.0, ROCPOWER.SAS, DESIGNROC.FOR, MULTIREADER POWER.SAS, CREATE-ROC.SAS, GB STAT VI0.0 (Dynamic Microsystems, Inc. Silver Spring, Md., USA), etc.

In some embodiments, the predetermined reference value is determined by carrying out a method comprising the steps of a) providing a collection of samples; b) providing, for each sample provided at step a), information relating to the actual clinical outcome for the corresponding subject (i.e. the duration of the survival); c) providing a serial of arbitrary quantification values; d) determining the expression levels of different genes for each sample contained in the collection provided at step a) so as to calculate the score as described above; e) classifying said samples in two groups for one specific arbitrary quantification value provided at step c), respectively: (i) a first group comprising samples that exhibit a quantification value for the score that is lower than the said arbitrary quantification value contained in the said serial of quantification values; (ii) a second group comprising samples that exhibit a quantification value for said score that is higher than the said arbitrary quantification value contained in the said serial of quantification values; whereby two groups of samples are obtained for the said specific quantification value, wherein the samples of each group are

separately enumerated; f) calculating the statistical significance between (i) the quantification value obtained at step e) and (ii) the actual clinical outcome of the patients from which samples contained in the first and second groups defined at step f) derive; g) reiterating steps f) and g) until every arbitrary quantification value provided at step d) is tested; h) setting the said
5 predetermined reference value as consisting of the arbitrary quantification value for which the highest statistical significance (most significant) has been calculated at step g).

For example the score has been assessed for 100 samples of 100 patients. The 100 samples are ranked according to the determined score. Sample 1 has the highest score and sample 100 has the lowest score. A first grouping provides two subsets: on one side sample Nr
10 1 and on the other side the 99 other samples. The next grouping provides on one side samples 1 and 2 and on the other side the 98 remaining samples etc., until the last grouping: on one side samples 1 to 99 and on the other side sample Nr 100. According to the information relating to the actual clinical outcome for the corresponding subject, Kaplan Meier curves are prepared for each of the 99 groups of two subsets. Also for each of the 99 groups, the p value between both
15 subsets was calculated. The predetermined reference value is then selected such as the discrimination based on the criterion of the minimum p value is the strongest. In other terms, the score corresponding to the boundary between both subsets for which the p value is minimum is considered as the predetermined reference value.

Typically, a score that is higher than the predetermined reference value indicates that
20 the patient will have a short survival time and a score that is lower than the predetermined reference value indicates that the patient will have a long survival time.

In some embodiments, the predetermined reference value thus allows discrimination between a poor and a good prognosis for a patient. Practically, high statistical significance values (e.g. low P values) are generally obtained for a range of successive arbitrary
25 quantification values, and not only for a single arbitrary quantification value. Thus, in one alternative embodiment of the invention, instead of using a definite predetermined reference value, a range of values is provided. Therefore, a minimal statistical significance value (minimal threshold of significance, e.g. maximal threshold P value) is arbitrarily set and a range of a plurality of arbitrary quantification values for which the statistical significance value calculated
30 at step g) is higher (more significant, e.g. lower P value) are retained, so that a range of quantification values is provided. This range of quantification values includes a "cut-off" value as described above. For example, according to this specific embodiment of a "cut-off" value, the outcome can be determined by comparing the calculated score with the range of values which are identified. In some embodiments, a cut-off value thus consists of a range of

quantification values, e.g. centered on the quantification value for which the highest statistical significance value is found (e.g. generally the minimum p value which is found). For example, on a hypothetical scale of 1 to 10, if the ideal cut-off value (the value with the highest statistical significance) is 5, a suitable (exemplary) range may be from 4-6. For example, a patient may be assessed by comparing values obtained by measuring the calculated score, where values higher than 5 reveal a poor prognosis and values less than 5 reveal a good prognosis. In some embodiments, a patient may be assessed by comparing values obtained by measuring the calculated score and comparing the values on a scale, where values above the range of 4-6 indicate a poor prognosis and values below the range of 4-6 indicate a good prognosis, with values falling within the range of 4-6 indicating an intermediate occurrence (or prognosis).

Typically, overexpression of ATP1A3, AVIL, CCL8, CFB, COL6A3, CRABP2, DNAJC12, F3, FAIM2, IL2RG, IL34, KRT13, KRT17, LRG1, MAGEB1, MAGEB2, MAGEC2, MAGEC3, MMP3, PLAC, PPEF1, PTGIR, , RAC2, SAA1, SAA2, SAMSN1, SERPINA3, SLC11A1, SLC2A3, TPRG1, TSPAN32, UCK2, WAS and WT1 correlates with a poor prognosis.

Typically, down expression of CES2, NAT8, PPARGC1A, and PTPN14 correlates with a good prognosis.

In some embodiments, the method of the invention comprises the use of a classification algorithm typically selected from Linear Discriminant Analysis (LDA), Topological Data Analysis (TDA), Neural Networks, Support Vector Machine (SVM) algorithm and Random Forests algorithm (RF) such as described in the Example. In some embodiments, the method of the invention comprises the step of determining the patient's survival using a classification algorithm.

As used herein, the term "classification algorithm" has its general meaning in the art and refers to classification and regression tree methods and multivariate classification well known in the art such as described in US 8,126,690; WO2008/156617. As used herein, the term "support vector machine (SVM)" is a universal learning machine useful for pattern recognition, whose decision surface is parameterized by a set of support vectors and a set of corresponding weights, refers to a method of not separately processing, but simultaneously processing a plurality of variables. Thus, the support vector machine is useful as a statistical tool for classification. The support vector machine non-linearly maps its n-dimensional input space into a high dimensional feature space, and presents an optimal interface (optimal parting plane) between features. The support vector machine comprises two phases: a training phase and a testing phase. In the training phase, support vectors are produced, while estimation is performed

according to a specific rule in the testing phase. In general, SVMs provide a model for use in classifying each of n subjects to two or more disease categories based on one k -dimensional vector (called a k -tuple) of biomarker measurements per subject. An SVM first transforms the k -tuples using a kernel function into a space of equal or higher dimension. The kernel function projects the data into a space where the categories can be better separated using hyperplanes than would be possible in the original data space. To determine the hyperplanes with which to discriminate between categories, a set of support vectors, which lie closest to the boundary between the disease categories, may be chosen. A hyperplane is then selected by known SVM techniques such that the distance between the support vectors and the hyperplane is maximal within the bounds of a cost function that penalizes incorrect predictions. This hyperplane is the one which optimally separates the data in terms of prediction (Vapnik, 1998 Statistical Learning Theory. New York: Wiley). Any new observation is then classified as belonging to any one of the categories of interest, based where the observation lies in relation to the hyperplane. When more than two categories are considered, the process is carried out pairwise for all of the categories and those results combined to create a rule to discriminate between all the categories. As used herein, the term "Random Forests algorithm" or "RF" has its general meaning in the art and refers to classification algorithm such as described in US 8,126,690; WO2008/156617. Random Forest is a decision-tree-based classifier that is constructed using an algorithm originally developed by Leo Breiman (Breiman L, "Random forests," Machine Learning 2001, 45:5-32). The classifier uses a large number of individual decision trees and decides the class by choosing the mode of the classes as determined by the individual trees. The individual trees are constructed using the following algorithm: (1) Assume that the number of cases in the training set is N , and that the number of variables in the classifier is M ; (2) Select the number of input variables that will be used to determine the decision at a node of the tree; this number, m should be much less than M ; (3) Choose a training set by choosing N samples from the training set with replacement; (4) For each node of the tree randomly select m of the M variables on which to base the decision at that node; (5) Calculate the best split based on these m variables in the training set. In some embodiments, the score is generated by a computer program.

In some embodiments, the method of the present invention comprises a) quantifying the level of a plurality of genes in the sample; b) implementing a classification algorithm on data comprising the quantified plurality of genes so as to obtain an algorithm output; c) determining the survival time from the algorithm output of step b).

The algorithm of the present invention can be performed by one or more programmable processors executing one or more computer programs to perform functions by operating on

input data and generating output. The algorithm can also be performed by, and apparatus can also be implemented as, special purpose logic circuitry, e.g., an FPGA (field programmable gate array) or an ASIC (application-specific integrated circuit). Processors suitable for the execution of a computer program include, by way of example, both general and special purpose

5 microprocessors, and any one or more processors of any kind of digital computer. Generally, a processor will receive instructions and data from a read-only memory or a random access memory or both. The essential elements of a computer are a processor for performing instructions and one or more memory devices for storing instructions and data. Generally, a computer will also include, or be operatively coupled to receive data from or transfer data to,

10 or both, one or more mass storage devices for storing data, e.g., magnetic, magneto-optical disks, or optical disks. However, a computer need not have such devices. Moreover, a computer can be embedded in another device. Computer-readable media suitable for storing computer program instructions and data include all forms of non-volatile memory, media and memory devices, including by way of example semiconductor memory devices, e.g., EPROM,

15 EEPROM, and flash memory devices; magnetic disks, e.g., internal hard disks or removable disks; magneto-optical disks; and CD-ROM and DVD-ROM disks. The processor and the memory can be supplemented by, or incorporated in, special purpose logic circuitry. To provide for interaction with a user, embodiments of the invention can be implemented on a computer having a display device, e.g., in non-limiting examples, a CRT (cathode ray tube) or LCD

20 (liquid crystal display) monitor, for displaying information to the user and a keyboard and a pointing device, e.g., a mouse or a trackball, by which the user can provide input to the computer. Other kinds of devices can be used to provide for interaction with a user as well; for example, feedback provided to the user can be any form of sensory feedback, e.g., visual feedback, auditory feedback, or tactile feedback; and input from the user can be received in any

25 form, including acoustic, speech, or tactile input. Accordingly, in some embodiments, the algorithm can be implemented in a computing system that includes a back-end component, e.g., as a data server, or that includes a middleware component, e.g., an application server, or that includes a front-end component, e.g., a client computer having a graphical user interface or a Web browser through which a user can interact with an implementation of the invention, or any

30 combination of one or more such back-end, middleware, or front-end components. The components of the system can be interconnected by any form or medium of digital data communication, e.g., a communication network. Examples of communication networks include a local area network ("LAN") and a wide area network ("WAN"), e.g., the Internet. The computing system can include clients and servers. A client and server are generally remote from

each other and typically interact through a communication network. The relationship of client and server arises by virtue of computer programs running on the respective computers and having a client-server relationship to each other.

5 In some embodiment, in view of the currently limited options for RCC management, the group of biomarkers as disclosed herein is useful for identifying patients with poor-prognosis, in particular patients with localized RCCs that are likely to relapse and metastasize. Accordingly, subject identified with a poor prognosis can be administered therapy, for example systematic therapy. In some embodiments, the method of the present invention be used to identify patients in need of frequent follow-up by a physician or clinician to monitor RCC
10 disease progression. Screening patients for identifying patients having a poor prognosis using the group of The biomarkers as disclosed herein is also useful to identify patients most suitable or amenable to be enrolled in clinical trial for assessing a therapy for RCC, which will permit more effective subgroup analyses and follow-up studies. Furthermore, the expression of the group of biomarkers as disclosed herein can be monitored in patients enrolled in a clinical trial
15 to provide a quantitative measure for the therapeutic efficacy of the therapy which is subject to the clinical trial.

This invention also provides a method for selecting a therapeutic regimen or determining if a certain therapeutic regimen is more appropriate for a patient identified as having a poor prognosis as identified by the methods as disclosed herein. For example, an
20 aggressive anti-cancer therapeutic regime can be perused in which a patient having a poor prognosis, where the patient is administered a therapeutically effective amount of an anti-cancer agent to treat the RCC. In some embodiments, a patient can be monitored for RCC using the methods and biomarkers as disclosed herein, and if on a first (i.e. initial) testing the patient is identified as having a poor prognosis, the patient can be administered an anti-cancer therapy,
25 and on a second (i.e. follow-up testing), the patient is identified as having a good prognosis, the patient can be administered an anti-cancer therapy at a maintenance dose. The method of the present invention is particularly suited to determining which patients will be responsive or experience a positive treatment outcome to a treatment.

In general, a therapy is considered to “treat” RCC if it provides one or more of the
30 following treatment outcomes: reduce or delay recurrence of the RCC after the initial therapy; increase median survival time or decrease metastases. In some embodiments, an anti-cancer therapy is, for example but not limited to administration of a chemotherapeutic agent, radiotherapy etc. Such anti-cancer therapies are disclosed herein, as well as others that are well known by persons of ordinary skill in the art and are encompassed for use in the present

invention. The term “anti-cancer agent” or “anti-cancer drug” is any agent, compound or entity that would be capable of negatively affecting the cancer in the patient, for example killing cancer cells, inducing apoptosis in cancer cells, reducing the growth rate of cancer cells, reducing the number of metastatic cells, reducing tumor size, inhibiting tumor growth, reducing blood supply to a tumor or cancer cells, promoting an immune response against cancer cells or a tumor, preventing or inhibiting the progression of cancer, or increasing the lifespan of the patient with cancer. Anti-cancer therapy includes biological agents (biotherapy), chemotherapy agents, and radiotherapy agents. In some embodiments, the anti-cancer therapy includes a chemotherapeutic regimen further comprises radiation therapy. In some embodiments, the anti-cancer treatment comprises the administration of a chemotherapeutic drug, alone or in combination with surgical resection of the tumor. In some embodiments, the treatment comprises radiation therapy and/or surgical resection of the tumor masses.

The term “chemotherapeutic agent” or “chemotherapy agent” are used interchangeably herein and refers to an agent that can be used in the treatment of RCC. In some embodiments, a chemotherapeutic agent can be in the form of a prodrug which can be activated to a cytotoxic form. Chemotherapeutic agents are commonly known by persons of ordinary skill in the art and are encompassed for use in the present invention. For example, chemotherapeutic drugs include, but are not limited to: temozolomide (Temodar), procarbazine (Matulane), and lomustine (CCNU). Chemotherapy given intravenously (by IV, via needle inserted into a vein) includes vincristine (Oncovin or Vincasar PFS), cisplatin (Platinol), carmustine (BCNU, BiCNU), and carboplatin (Paraplatin), Mexotrexate (Rheumatrex or Trexall), irinotecan (CPT-11); erlotinib; oxaliplatin; anthracyclins-idarubicin and daunorubicin; doxorubicin; alkylating agents such as melphalan and chlorambucil; cis-platinum, methotrexate, and alkaloids such as vindesine and vinblastine.

In some embodiments, the patients are administered with anti-VEGF agents. As used herein the term “anti-VEGF agent” refers to any compound or agent that produces a direct effect on the signaling pathways that promote growth, proliferation and survival of a cell by inhibiting the function of the VEGF protein, including inhibiting the function of VEGF receptor proteins. The term “agent” or “compound” as used herein means any organic or inorganic molecule, including modified and unmodified nucleic acids such as antisense nucleic acids, RNAi agents such as siRNA or shRNA, peptides, peptidomimetics, receptors, ligands, and antibodies. Preferred VEGF inhibitors, include for example, AVASTIN® (bevacizumab), an anti-VEGF monoclonal antibody of Genentech, Inc. of South San Francisco, Calif., VEGF Trap (Regeneron/Aventis). Additional VEGF inhibitors include CP-547,632 (3-(4-Bromo-2,6-

5 difluoro-benzyloxy)-5-[3-(4-pyrrolidin-1-yl-butyl)-ureido]-isothiazole-4-carboxylic acid
amide hydrochloride; Pfizer Inc., NY), AG13736, AG28262 (Pfizer Inc.), SU5416, SU11248,
& SU6668 (formerly Sugen Inc., now Pfizer, New York, N.Y.), ZD-6474 (AstraZeneca),
ZD4190 which inhibits VEGF-R2 and -R1 (AstraZeneca), CEP-7055 (Cephalon Inc., Frazer,
10 Pa.), PKC 412 (Novartis), AEE788 (Novartis), AZD-2171), NEXAVAR® (BAY 43-9006,
sorafenib; Bayer Pharmaceuticals and Onyx Pharmaceuticals), vatalanib (also known as PTK-
787, ZK-222584: Novartis & Schering: AG), MACUGEN® (pegaptanib octasodium, NX-
1838, EYE-001, Pfizer Inc./Gilead/Eyetech), IM862 (glufanide disodium, Cytran Inc. of
Kirkland, Wash., USA), VEGFR2-selective monoclonal antibody DC101 (ImClone Systems,
10 Inc.), angiozyme, a synthetic ribozyme from Ribozyme (Boulder, Colo.) and Chiron
(Emeryville, Calif.), Sirna-027 (an siRNA-based VEGFR1 inhibitor, Sirna Therapeutics, San
Francisco, Calif.) Caplostatin, soluble ectodomains of the VEGF receptors, Neovastat (AEterna
Zentaris Inc; Quebec City, Calif.) and combinations thereof.

15 The compounds used in connection with the treatment methods of the present invention
are administered and dosed in accordance with good medical practice, taking into account the
clinical condition of the individual subject, the site and method of administration, scheduling
of administration, patient age, sex, body weight and other factors known to medical
practitioners. The pharmaceutically “effective amount” for purposes herein is thus determined
by such considerations as are known in the art. The amount must be effective to achieve
20 improvement including, but not limited to, improved survival rate or more rapid recovery, or
improvement or elimination of symptoms and other indicators as are selected as appropriate
measures by those skilled in the art.

The invention will be further illustrated by the following figures and examples.
However, these examples and figures should not be interpreted in any way as limiting the scope
25 of the present invention.

FIGURES:

Figure 1 summarizes steps of the methodology.

30 **Figure 2.** Kaplan-Meier survival analysis stratified by score index for MAGEB2,
MAGEB1, PTPN14 (-), TSPAN32 gene signature. Censored patients (alive at last follow-up)
are indicated on the curves. RIG = low: patient group with low risk index; RIG = medium:
patient group with medium risk index; RIG = high: patient group with high risk index.

Figure 3. Kaplan-Meier survival analysis stratified by score index for ATP1A3 gene
signature. Censored patients (alive at last follow-up) are indicated on the curves. RIG = low:

patient group with low risk index; RIG = medium: patient group with medium risk index; RIG = high: patient group with high risk index.

Figure 4. Kaplan-Meier survival analysis stratified by score index for NAT8 (-), PPARGC1A (-), TPRG1, WAS gene signature. Censored patients (alive at last follow-up) are indicated on the curves. RIG = low: patient group with low risk index; RIG = medium: patient group with medium risk index; RIG = high: patient group with high risk index.

Figure 5. Kaplan-Meier survival analysis stratified by score index for MAGEC2, MAGEC3, CCL8, CFB, COL6A3, DNAJC12, IL2RG, IL34, PLAC8 gene signature. Censored patients (alive at last follow-up) are indicated on the curves. RIG = low: patient group with low risk index; RIG = medium: patient group with medium risk index; RIG = high: patient group with high risk index.

Figure 6. Kaplan-Meier survival analysis stratified by score index for KRT13, LRG1, SAA2, SAA1, SAMSNI, SERPINA3, SLC2A3, UCK2, WT1 gene signature. Censored patients (alive at last follow-up) are indicated on the curves. RIG = low: patient group with low risk index; RIG = medium: patient group with medium risk index; RIG = high: patient group with high risk index.

Figure 7. Kaplan-Meier survival analysis stratified by score index for CRABP2, F3, MMP3, PPEF1, SLC11A1 gene signature. Censored patients (alive at last follow-up) are indicated on the curves. RIG = low: patient group with low risk index; RIG = medium: patient group with medium risk index; RIG = high: patient group with high risk index.

Figure 8. Kaplan-Meier survival analysis stratified by score index for AVIL, CES2 (-), FAIM2, KRT17, PTGIR, RAC2 gene signature. Censored patients (alive at last follow-up) are indicated on the curves. RIG = low: patient group with low risk index; RIG = medium: patient group with medium risk index; RIG = high: patient group with high risk index.

Figure 9. Validation of the "Lung" signature. (A and B) Kaplan-Meier for overall (OS) and disease-free (DFS) survival analysis stratified in 3 groups of equivalent size. Signature "low": patient group with low score; signature "medium": patient group with a medium score; signature "high": patient group with high score. (C and D) Density plot of p values (log-rank test) from 1000 random signatures of equal size for OS and DFS. p = empirical p-value.

Figure 10. Summary table of up and downregulated genes from the first biomarker discovery step.

EXAMPLE:

Material & Methods

The results are depicted in Figure 1-8.

Example 1:

Bioinformatics analyses

Biomarker discovery

1. Differential expression between parental cell line and passage 6 experiments.

To select genes that present the strongest difference in expression between parental cell line and passage 6 experiments, we used a z-score approach. First, we compute the log fold change for each gene as $\log FC = \log_2(\text{expression in p6}) - \log_2(\text{expression in parental cell line})$. Second, we compute the mean and standard deviation of the logFC and a z-score for all genes. A gene will be considered as differentially expressed if the absolute value of its z-score is ≥ 2.58 and if its logFC is ≤ -1 (down expressed) or ≥ 2 (up expressed).

2. Progressive expression pattern through passages.

To capture genes that present a progressive expression through passages, we compute mean values of expression for each passage. A gene will be considered progressive if its mean values are strictly increasing or decreasing through passages 3 to 5. Passages 1 and 2 were excluded because of their low number of replicates. To ensure the global progressive trend, we conserve genes only if the t.test between passages 1, 2 and 3 versus 6 is significant ($\text{fdr} \leq 0.01$).

3. Differentially expressed and progressive genes.

A gene will be considered differentially progressively expressed through passages if it is in the intersection of the gene sets selected in 1. and 2. steps with the constraint that the expression should be in the same direction (up or down).

Biomarker validation

1. Selection of clinically relevant genes.

In order to conserve genes that are relevant in human kidney cancer, we used the well known TCGA KIRC cohort. For each gene, we fit a Cox proportional hazard regression model based on overall survival time. A gene is conserved if the fdr adjusted p-value of its log-rank test is ≤ 0.01 and if its hazard ratio (HR) is in accord with the differential expression ($\text{HR} \geq 1$ for up regulated genes and $\text{HR} \leq 1$ for down regulated genes).

2. Relevance of the association of selected genes.

To measure the clinical relevance of the resulting signature, we used the Leave-one-out cross validation : for each patient, the risk index is defined as the linear combination of gene

expression values weighted by their estimated Cox model regression coefficients fitted from all the cohort but this patient. Patients are then classified in accord to their risk index and split in 3 groups to perform a log rank test.

All these steps have been performed for:

- specific experiment types: kidney (K), tail (T) and lung (L)
- composed experiment types: kidney-tail (KT), kidney-lung (KL), tail-lung (TL)
- global experiment type: kidney-tail-lung (KTL)

Each gene is specific to an experiment type. Indeed, if a gene is part of KTL signature, it will not be included in the 6 other experiment type signatures even if it was considered as differentially and progressively expressed in associated experiments.

Results

Renal Cell Carcinoma (RCC) encompasses a heterogeneous group of cancers derived from renal tubular epithelial cells and has a worldwide mortality. However, mortality rates have barely improved over the last 20 years. Novel biomarkers and biomarkers are thus urgently required for this cancer. The inventors have devised a strategy to produce mouse cancer cell lines of progressively enhanced aggressiveness and specialization. The mouse renal cancer cell line RENCA was serially passaged in vivo using multiple implantation strategies designed to replicate different aspects of primary tumour growth and metastasis. Transcriptomic and epigenomic data has been acquired for the derived cell lines and primary analyses have been performed. The inventors then selected plurality of genes with no reported role in RCC, and checked their relevance in patient data samples. This approach contributes to identify several gene signatures that are suitable for predicting survival time in patients suffering from RCC. Final signatures are:

- Kidney: MAGEB2, MAGEB1, PTPN14 (-), TSPAN32
- Tail: CRABP2, F3, MMP3, PPEF1, SLC11A1
- Lung: KRT13, LRG1, SAA2, SAA1, SAMSN1, SERPINA3, SLC2A3, UCK2, WT1
- Kidney-tail: NAT8 (-), PPARGC1A (-), TPRG1, WAS
- Kidney-lung: ATP1A3
- Tail-lung: AVIL, CES2 (-), FAIM2, KRT17, PTGIR, RAC2
- Kidney-tail: MAGEC2, MAGEC3, CCL8, CFB, COL6A3, DNAJC12, IL2RG, IL34, PLAC8

For all genes, the overexpression is predictive of a bad prognosis except for those that are followed by the (-) sign.

Example 2:

Transcriptomic analysis

We performed full genome transcriptomic analysis of the 67 cell lines. The P1 cell line was excluded from the analysis because of insufficient number of animals (< 3). Therefore, data acquired from the different lines were labeled S0 to S5 (S0 being the parental cells and S1 to S5 representing P2 to P6). We used Principal Component Analysis (PCA) in order to summarize the information contained in our data sets for sample series S2 and S5. Principal Component Analyses showed that, while the subtypes had similar expression profiles in the early passages, the transcriptional profiles became different between “Kidney” (K), “Tail” (T), and “Lung” (L) groups in later passages and, thus, clustered into distinct groups. Tight clustering of the biological replicates (one mouse per data point) also showed that the profiles were stable. When all genes that were major contributors to the Principal Component 1 (PC1) and 2 (PC2) were pooled, the heat map of the transcriptomic profiles revealed a gradual change in expression. Enrichment analysis of these most contributing genes demonstrated several highly enriched categories. These include the following GO terms: extracellular matrix, angiogenesis, cell proliferation, cell adhesion, cell migration, immune process and inflammation and apoptosis.

Next, we investigated whether transcriptional signatures derived from the differentially expressed genes in the K, T and L groups could predict outcome for patients using the Clear Cell Renal Cell Carcinoma dataset (KIRC) from The Cancer Genome Atlas (TCGA). We compared genes that changed their expression between the parental and S5 cell line and named these genes as progressively regulated (up or down) (**Figure 10**). We included in our analysis only genes having their expression consistently increasing in the different series. The analysis has been performed by comparison to the parental cell line for K, T and L groups alone. The additional group “Kidney-Tail-Lung” (KTL) is represented by pooling the different groups in a unique group. The KTL group is also compared to the parental cell line.

A total of 131 genes for all groups satisfied the inclusion criteria. Among these, 26 were in the K, 45 in the T, 59 in the L and 31 in the KTL group (**Figure 10**). Among those, 123 genes were upregulated and only a small fraction was downregulated. We next investigated the predictive value of these 4 gene signatures using TCGA (KIRC cohort). Since only 8 genes were downregulated, we chose only to include the upregulated genes in the validation of our

signatures. Supplementary Table 2 depicts the list of genes for each group. For each gene, we fitted a Cox-proportional hazard regression model based on overall survival (OS) or disease-free survival (DFS). A gene was conserved if its falsediscovery rate (FDR) adjusted p-value of its log-rank test was lower than 0.01 and if the hazard ratio was in agreement with the differential expression. We have further created a 5th “All Merged” signature composed of genes that are members of all of the previously defined 4 signatures.

Signature	Gene
K	IL34, MAGEB2, TSPAN32, WAS
T	BASP1, CRABP2, DNAJC12, DPYSL3, F3, GLIPR2, HMGA2, MAGEC2, MAGEC3, PLAC8, PPEF1, RAB15, SLC11A1, SNAI2
L	ATP1A3, CCL8, CFB, IL2RG, KRT13, KRT17, PCBP3, PLAC8, RAC2, SAA1, SAA2, SAMSN1, SERPINA3, SLC2A3, UCK2, WT1
KTL	CCL8, CFB, DNAJC12, IL2RG, IL34, MAGEC2, MAGEC3, PLAC8, SERPINA3
allMerged	ATP1A3, BASP1, CCL8, CFB, CRABP2, DNAJC12, DPYSL3, F3, GLIPR2, HMGA2, IL2RG, IL34, KRT13, KRT17, MAGEB2, MAGEC2, MAGEC3, PCBP3, PLAC8, PPEF1, RAB15, RAC2, SAA1, SAA2, SAMSN1, SERPINA3, SLC11A1, SLC2A3, SNAI2, TSPAN32, UCK2, WAS, WT1

Table 1: Summary table of the signatures and their predictive value in the KIRC TCGA cohort.

10

Table 1 depicts the signatures for each group (K, T, L, KTL) and their significance for OS and DFS. We analyzed OS (**Figure 9A**) of all patients and DFS (**Figure 9B**) of M0 patients. We validated our signatures by computing an empirical p-value and by testing our signature against 1000 random signatures of equivalent size (**Figure 9C and 9D**). Furthermore, we performed multivariate Cox regression analysis of our signature (**Data not shown**). After adjusting for clinical variables (TNM stage and Fuhrman grade), the Lung signature remained an independent prognostic factor for predicting both OS and DFS. The Lung signature remains the most significant for OS and DFS when compared to the others, except for the “All Merged” signature. The performances of the signatures were the following for the different groups when ranked according to the hazard ratio: K<T<KTL<L

20

The identification of signatures predictive of patient outcome also validates our experimental approach and shows that the strategy of generation of increasingly specialized mouse cell lines revealed novel genes and signatures with relevance to human RCC.

REFERENCES:

5

Throughout this application, various references describe the state of the art to which this invention pertains. The disclosures of these references are hereby incorporated by reference into the present disclosure.

CLAIMS:

1. A method for predicting the survival time of a patient suffering from a renal cell carcinoma (RCC) comprising determining the expression level of at least one gene in a sample obtained from the patient wherein said gene is selected from the group consisting
5 of ATP1A3, AVIL, CCL8, CES2, CFB, COL6A3, CRABP2, DNAJC12, F3, FAIM2, IL2RG, IL34, KRT13, KRT17, LRG1, MAGEB1, MAGEB2, MAGEC2, MAGEC3, MMP3, NAT8, PLAC, PPARGC1A, PPEF1, PTGIR, PTPN14, RAC2, SAA1, SAA2, SAMS1, SERPINA3, SLC11A1, SLC2A3, TPRG1, TSPAN32, UCK2, WAS, WT1 , BASP1, DPYSL3, GLIPR2, HMGA2, RAB15, SNAI2 and PCBP3.
- 10 2. The method of claim 1 comprising determining the expression levels of KRT13, LRG1, SAA2, SAA1, SAMS1, SERPINA3, SLC2A3, UCK2, and WT1 in the sample obtained from the patient.
3. The method of claim 1 comprising determining the expression levels of NAT8, PPARGC1A, TPRG1, and WAS in the sample obtained from the patient.
- 15 4. The method of claim 1 comprising determining the expression level of ATP1A3 in the sample obtained from the patient.
5. The method of claim 1 comprising determining the expression levels of AVIL, CES2, FAIM2, KRT17, PTGIR, and RAC2 in the sample obtained from the patient.
- 20 6. The method of claim 1 comprising determining the expression levels of MAGEC2, MAGEC3, CCL8, CFB, COL6A3, DNAJC12, IL2RG, IL34, and PLAC8 in the sample obtained from the patient.
7. The method of claim 1 wherein the sample is a tissue tumor sample.
8. The method of claim 1 wherein a score which is a composite of the expression levels of the different biomarkers is determined and compared to the predetermined reference
25 value wherein a difference between said score and said predetermined reference value is indicative whether the patient will have a long or short survival time.

9. Use of the method of claim 1 for selecting a therapeutic regimen or determining if a certain therapeutic regimen is more appropriate for a patient identified as having a poor prognosis.
- 5 10. Use of the method of claim 1 for monitoring of RCC in a patient wherein if on a first (i.e. initial) testing the patient is identified as having a poor prognosis, the patient can be administered an anti-cancer therapy, and on a second (i.e. follow-up testing), the patient is identified as having a good prognosis, the patient can be administered an anti-cancer therapy at a maintenance dose.
- 10 11. Use of the method of claim 1 for determining whether the patient will be responsive or experience a positive treatment outcome to a treatment.

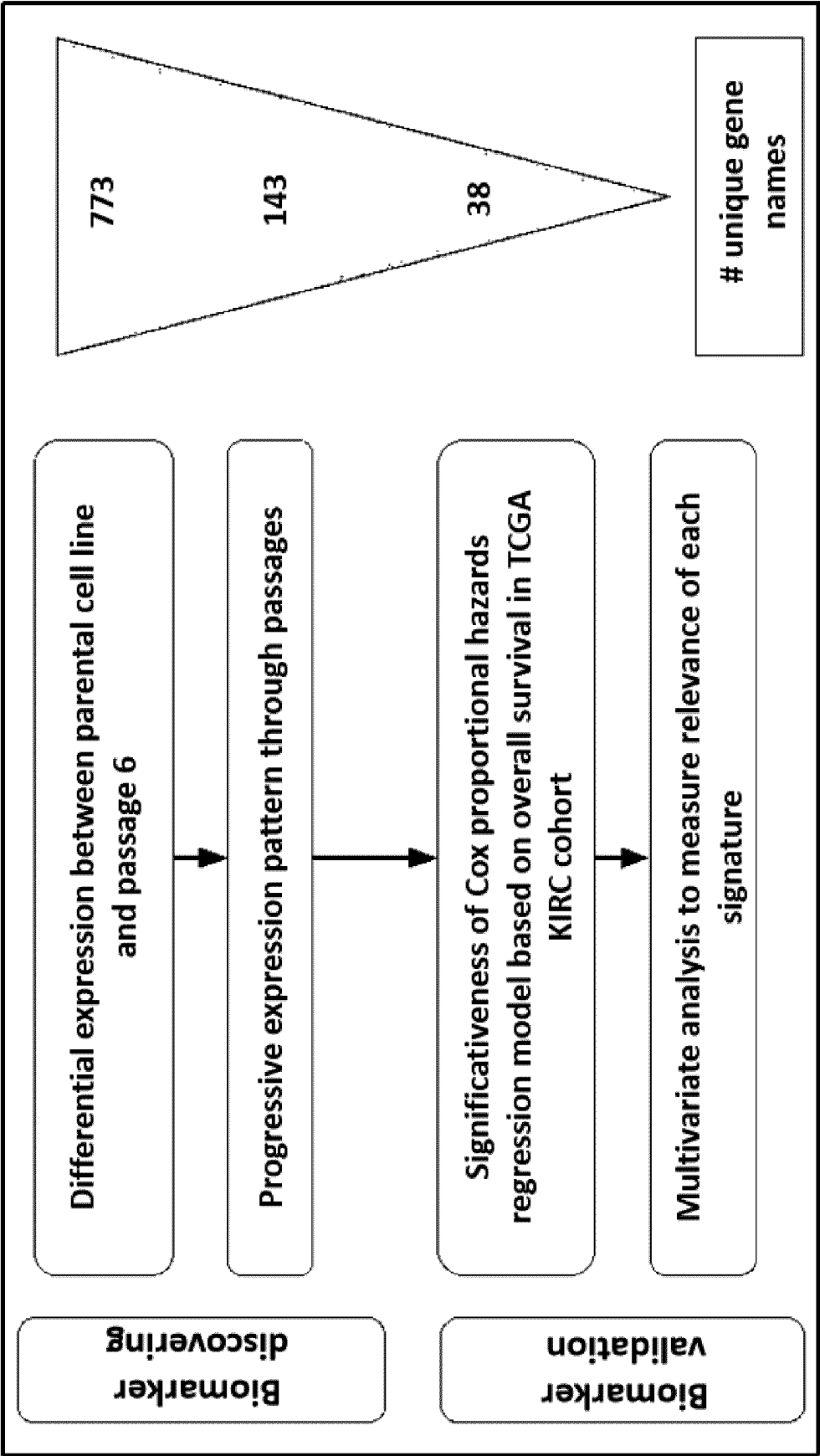


Figure 1

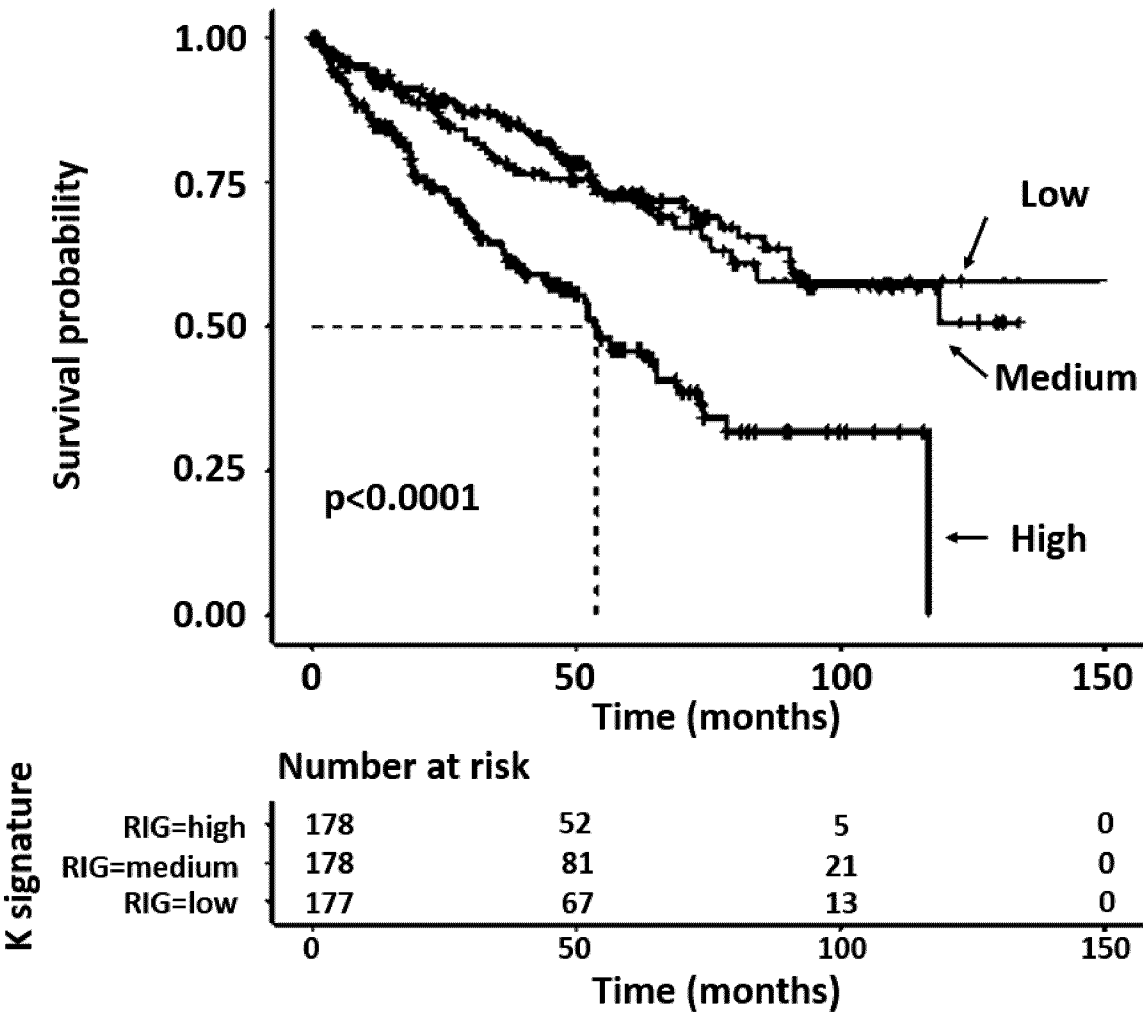


Figure 2

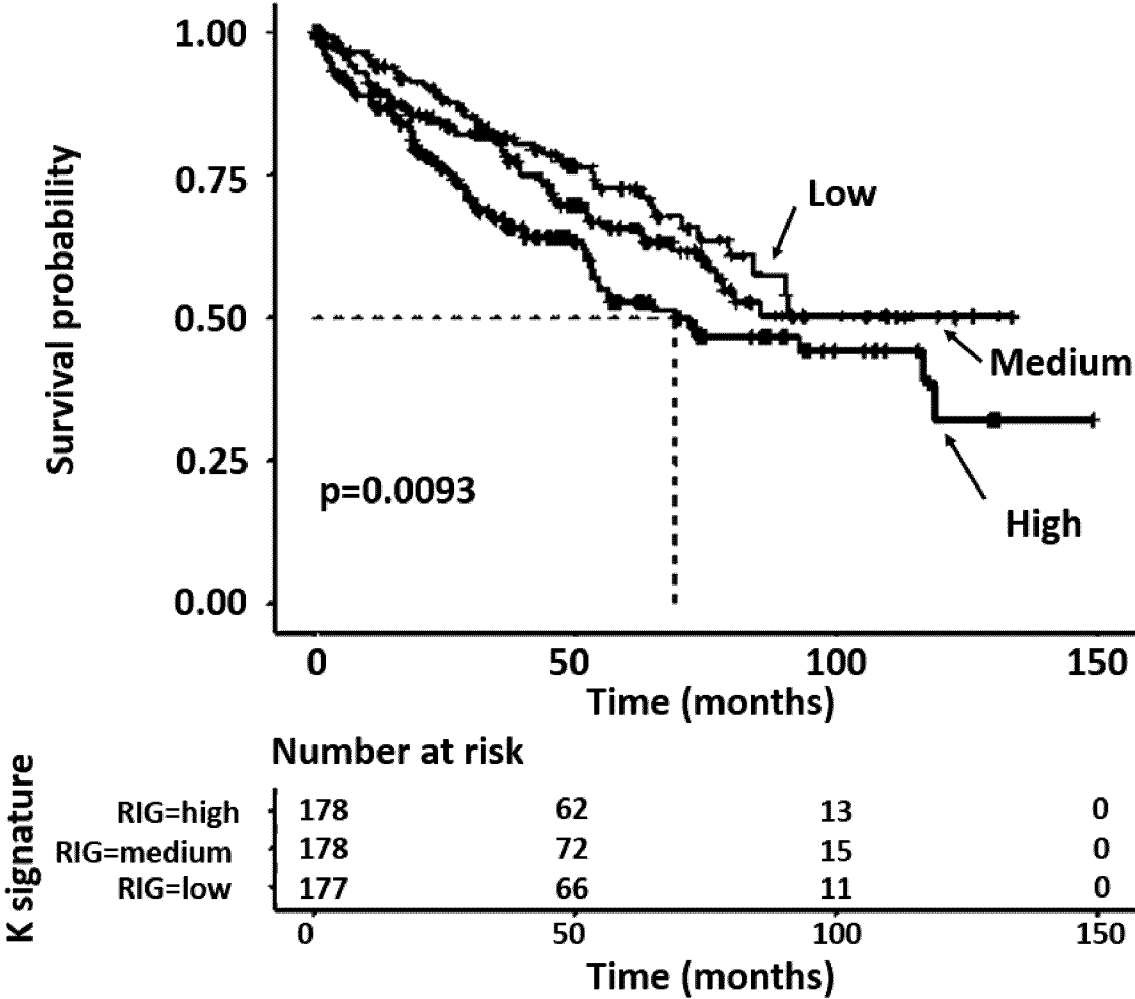


Figure 3

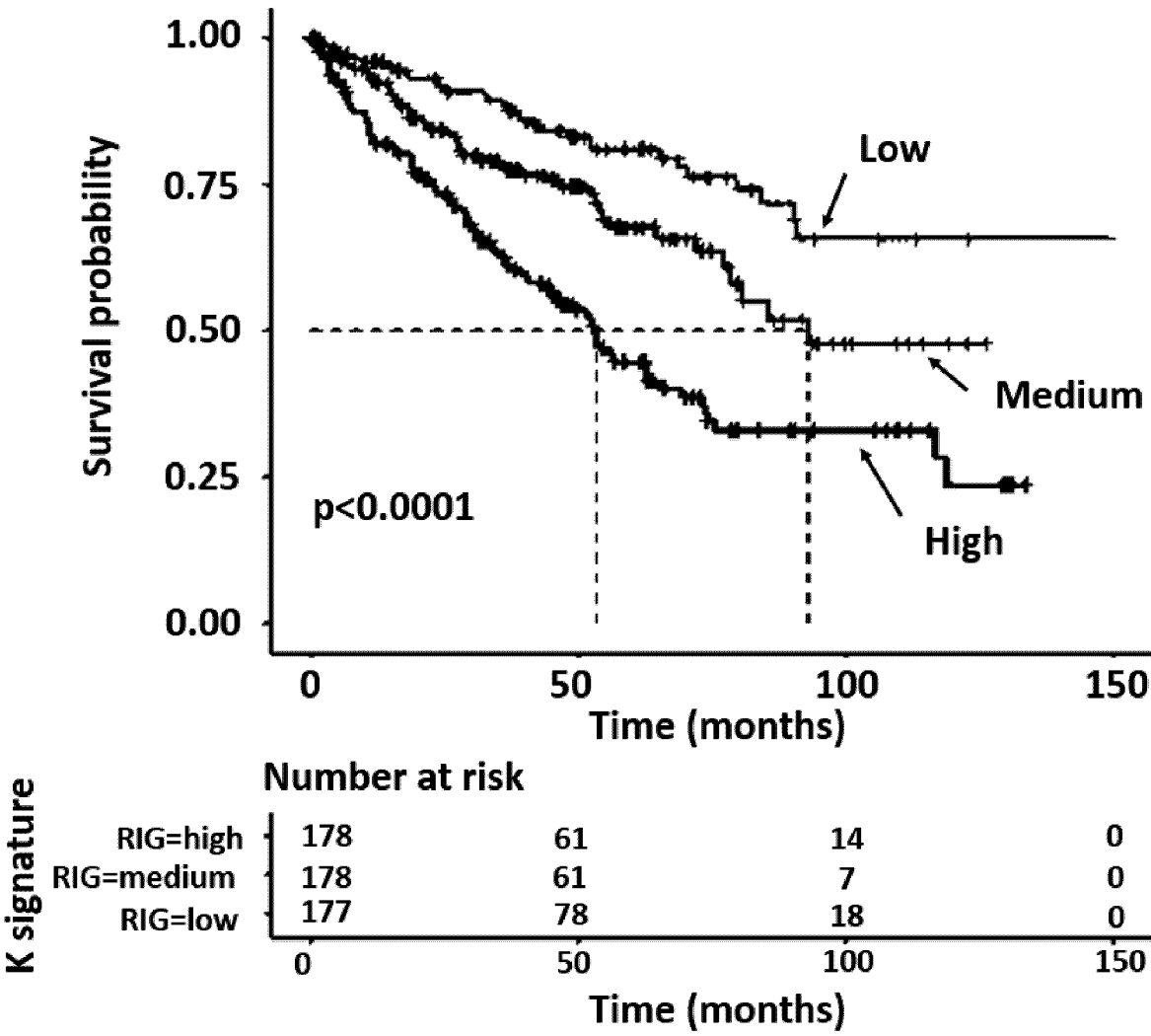


Figure 4

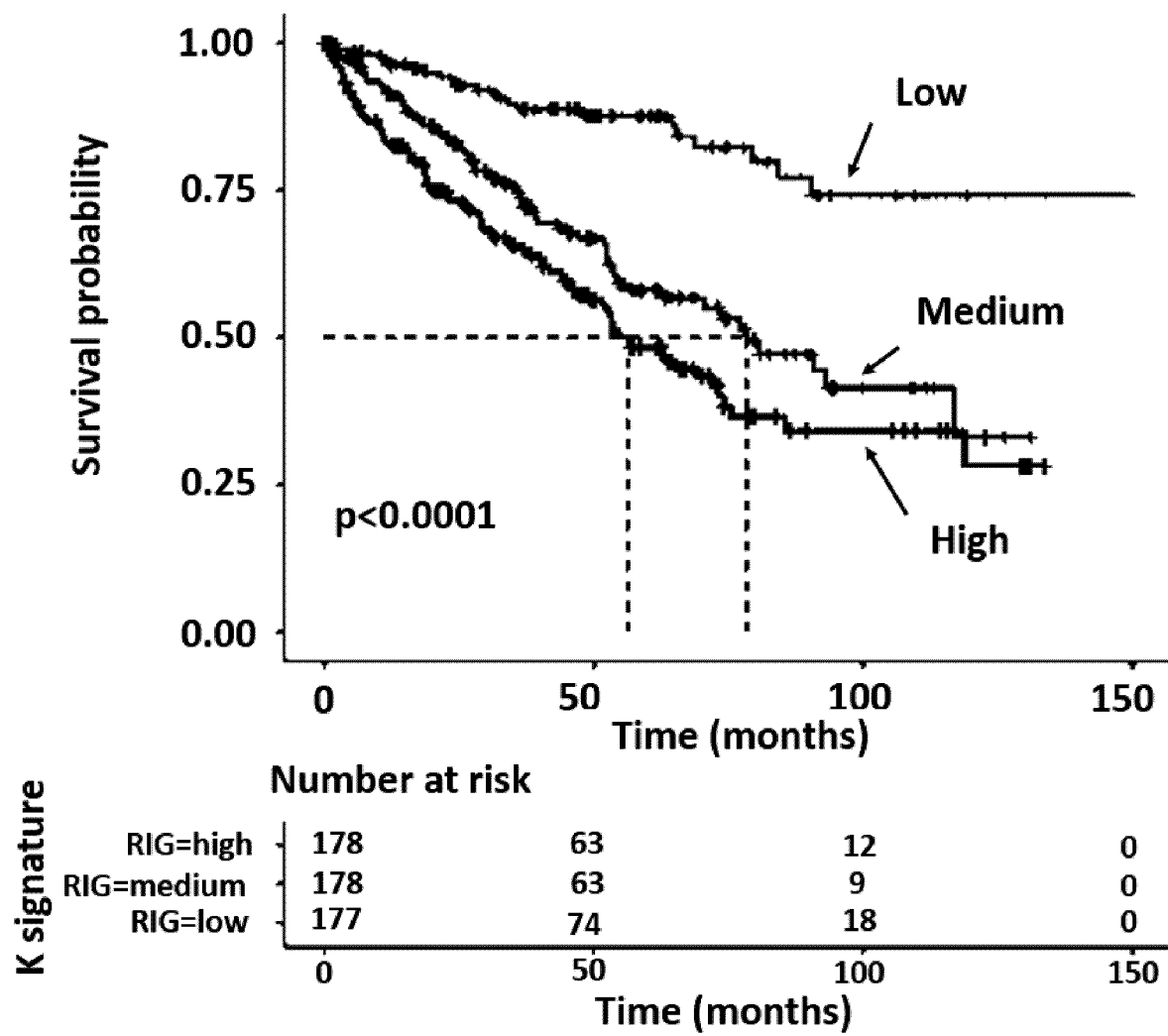


Figure 5

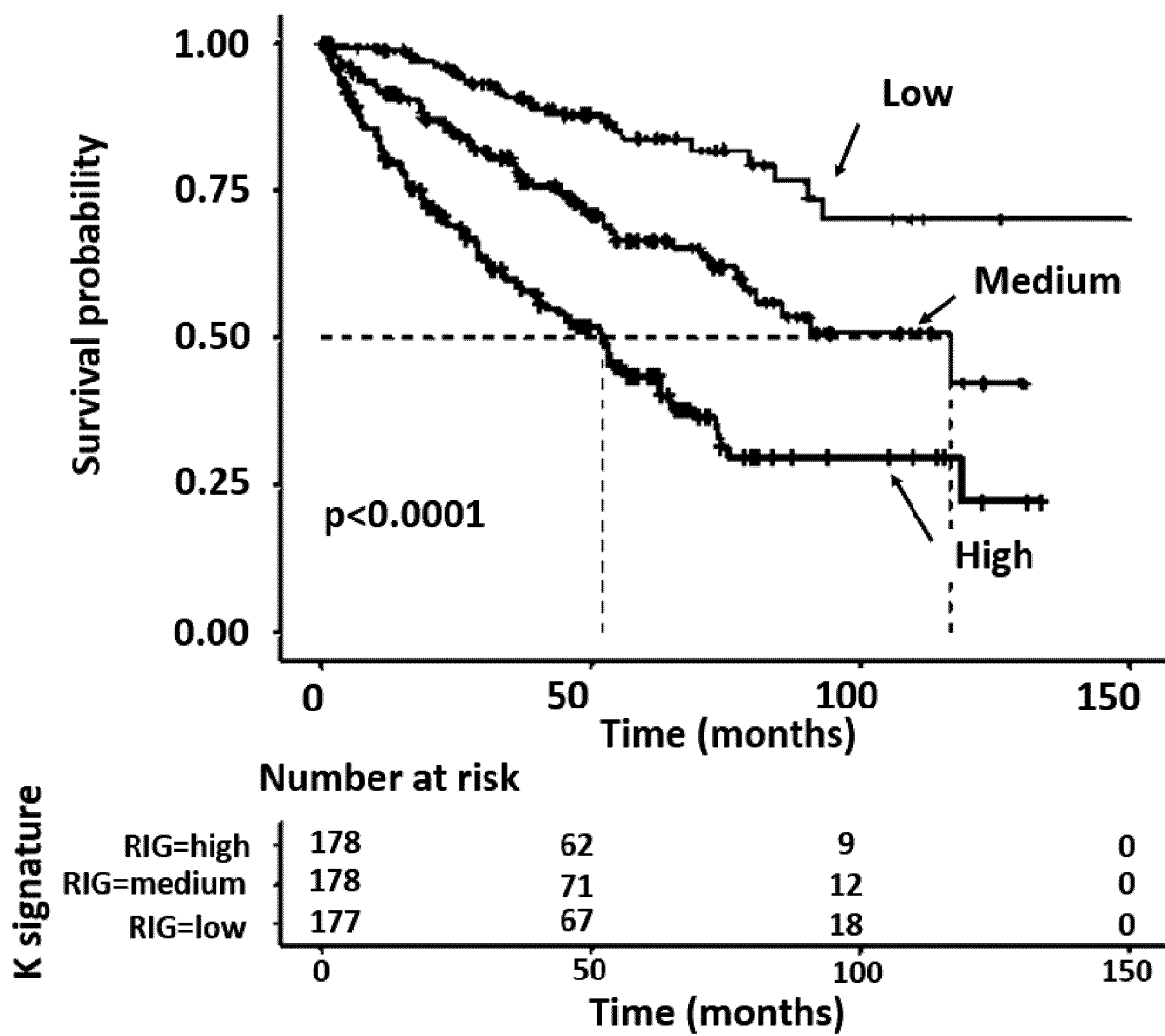


Figure 6

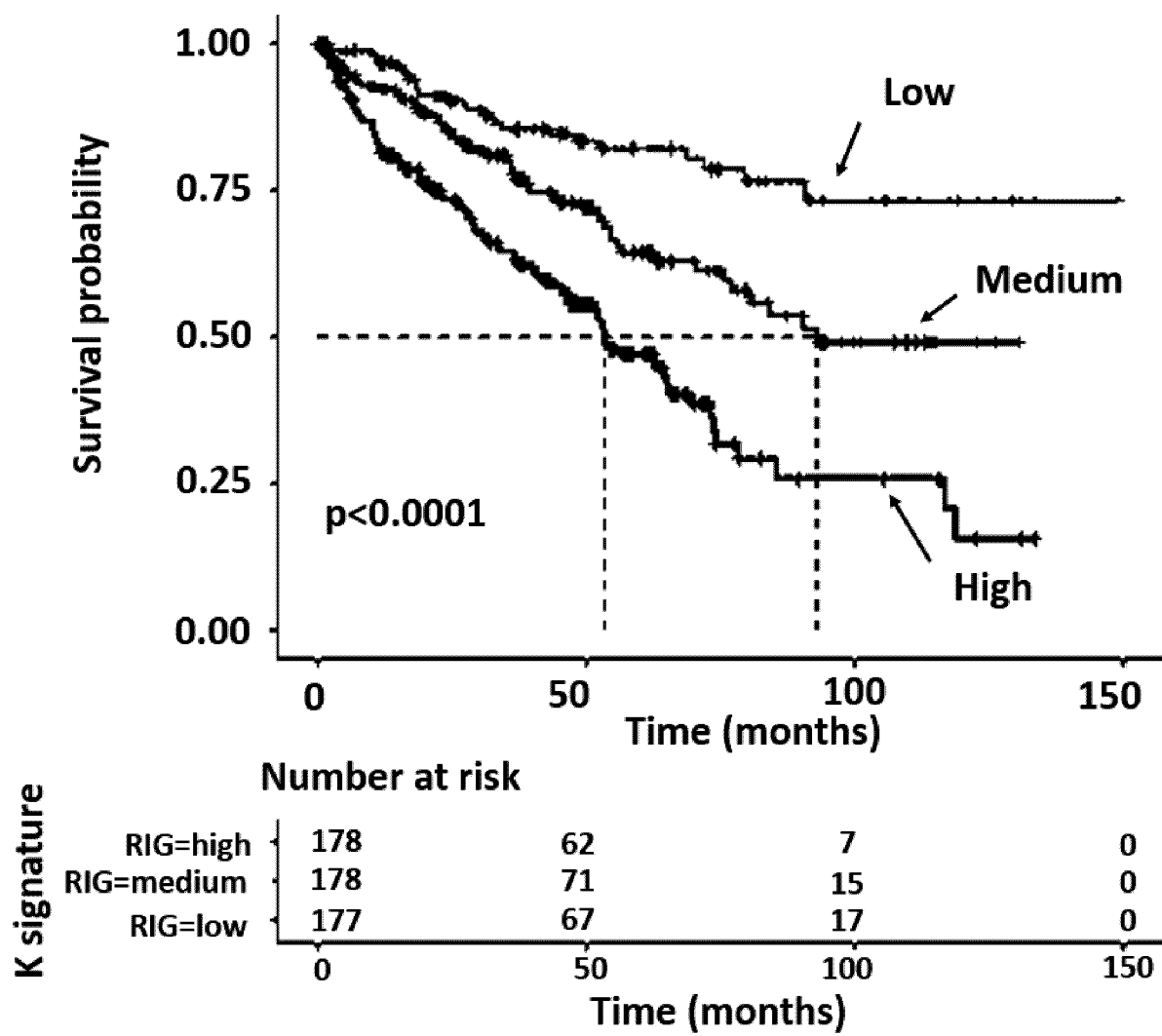


Figure 7

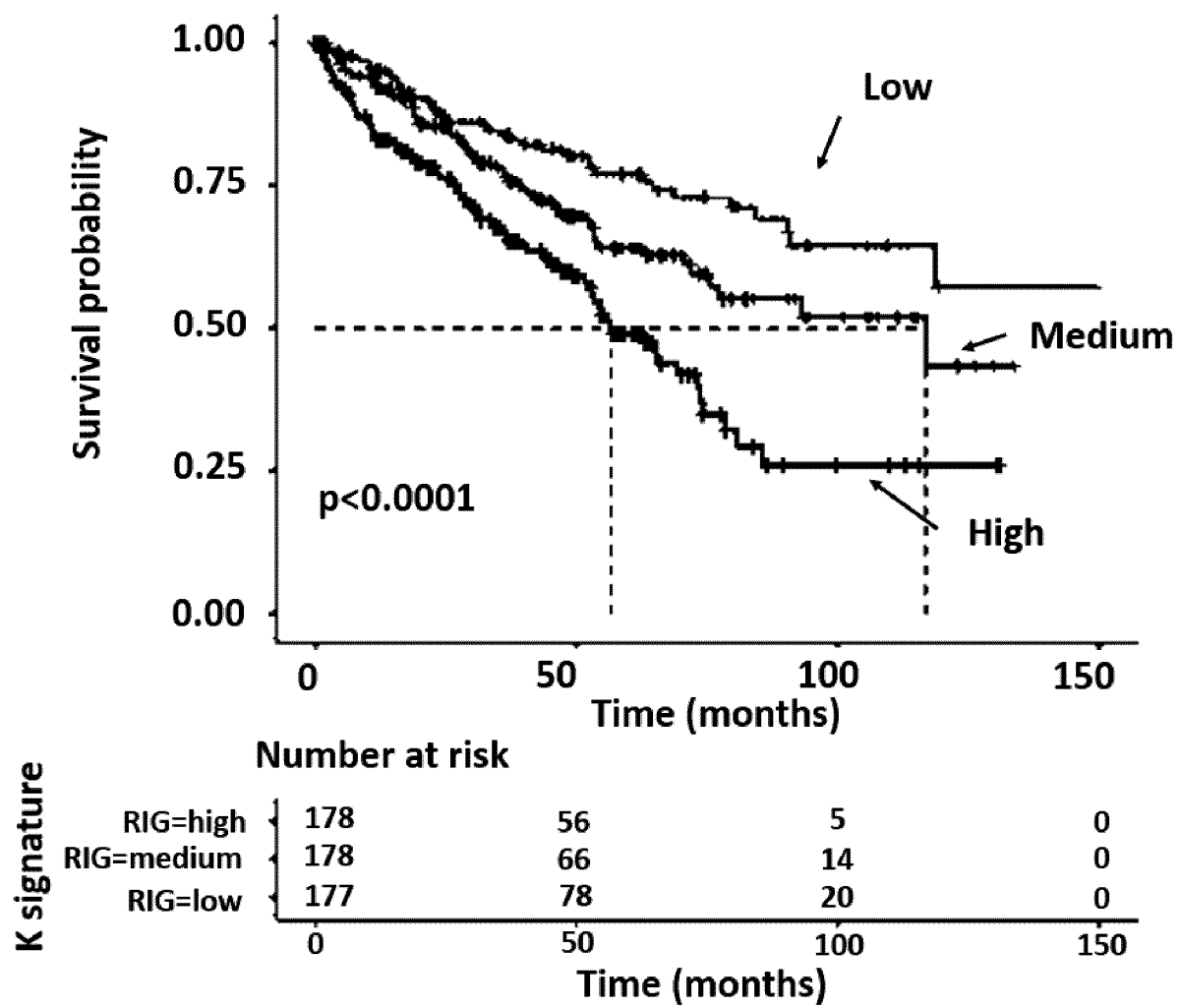


Figure 8

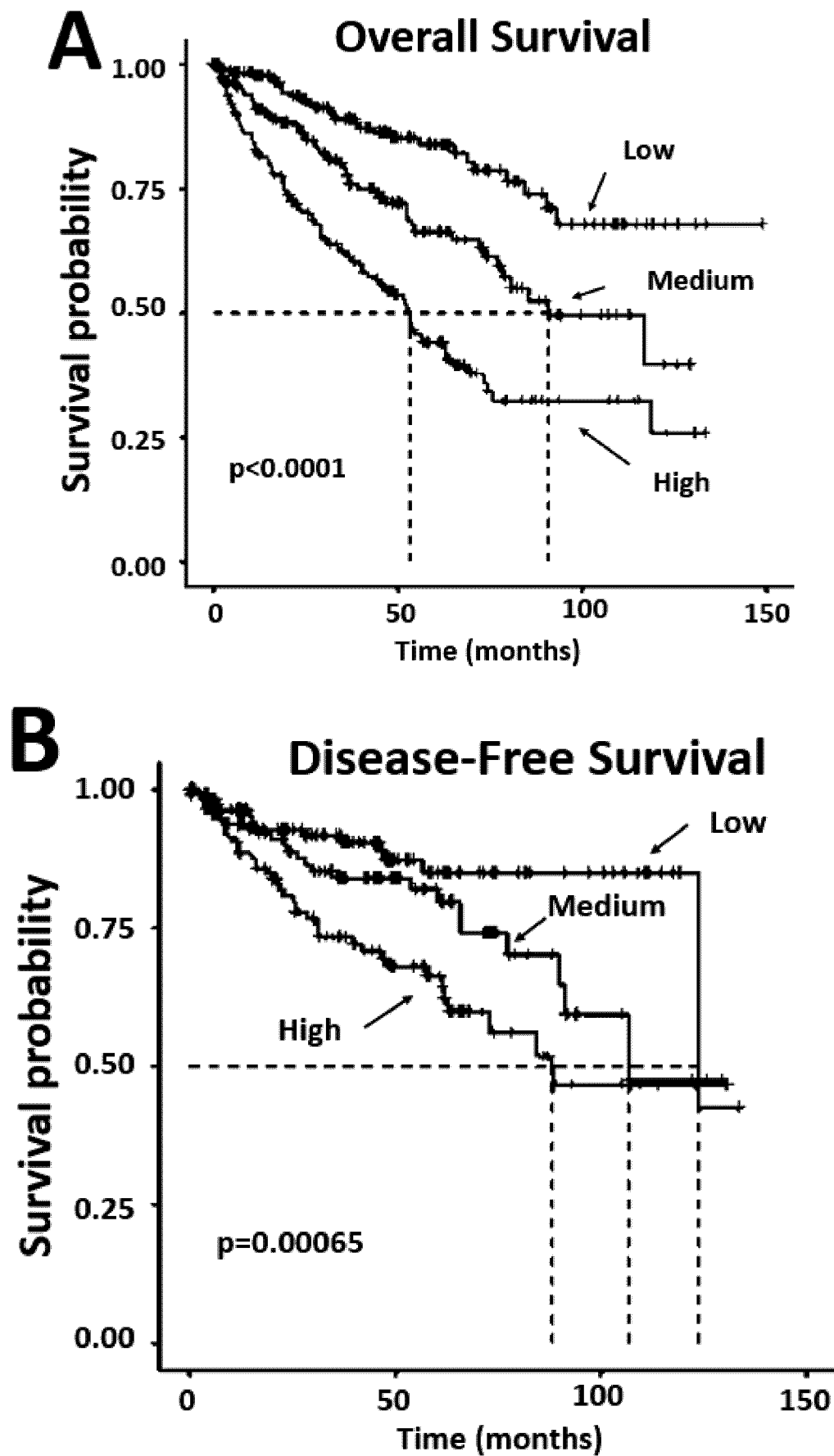


Figure 9A and 9B

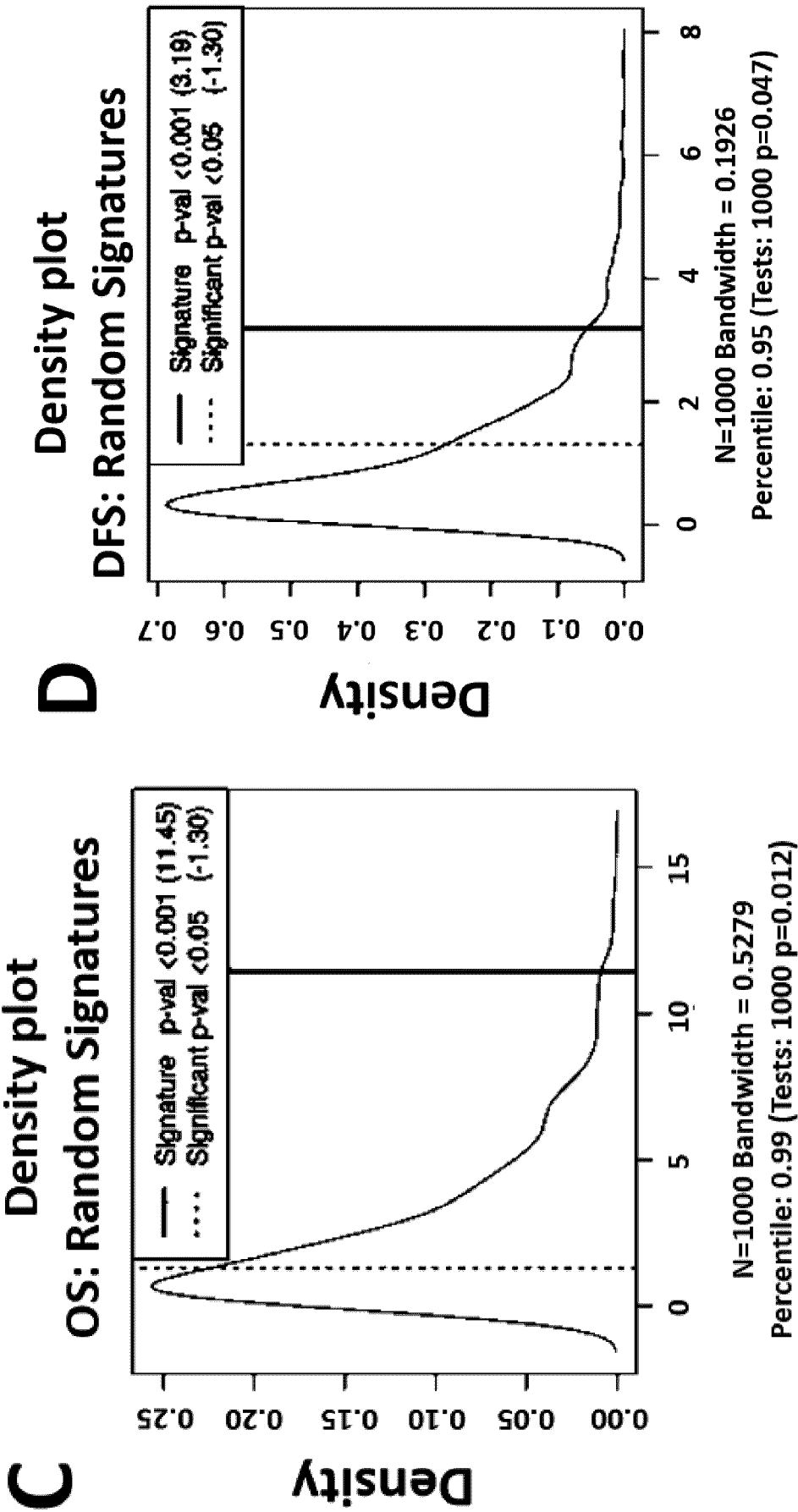


Figure 9C and 9D

Signature	Direction	Count	Genes
K	up	25	2010300C02Rik, Abcb1a, Abcb1b, Abcb4, Arhgap30, Arhgef40, Bmp4, Chnra1, D4Bwg0951e, Fam167b, Gja1, Gm14625, Gm2012, Gm5168, Icam2, Il18r1, Il1rl1, Il34, Mageb4, P2ry14, Slc7a3, Tnfrsf26, Tspan32, Was, Wdfy4
K	down	1	Ddit4
T	up	43	1700019B21Rik, 4930550L24Rik, 4930583H14Rik, Acta1, Armcx2, B3galt1, B4galnt2, Basp1, Cck, Clec10a, Crabp2, Cryaa, Dapk1, Dnajc12, Dpysl3, F3, Fbln5, Glipr2, Gm6880, Gm6890, Gng11, H2-M10.1, Hmga2, Hvcn1, Khdrbs3, Klhl13, Krt20, Ldhd, Nsg1, Pdgrfb, Pdlim1, Plac8, Postn, Ppef1, Prss23, Rab15, Raet1c, Sdpr, Slc11a1, Snai2, Thbd, Tspan6, Wipf1
T	down	2	B230114P17Rik, Lipn
L	up	54	2310043M15Rik, 9030619P08Rik, Abcb1a, Abcb1b, Adam22, Apbb1p, Arap3, Atp1a3, B4galnt2, Ccl8, Cdhr5, Cfb, Cldn1, Fam5c, Fcrlb, Gcnt2, Gm2022, Gpr116, Hvcn1, Il2rg, Itgb3, Krt13, Krt17, LOC641136, Ms4a4d, Oas2, Orm1, Orm2, Orm3, P2ry14, Pak1, Pcbp3, Plac8, Rac2, Robo2, Saa2, Samsn1, Serinc2, Serpina3b, Serpina3g, Serpina3h, Serpina3i, Slc10a6, Slc2a3, Slc7a4, Steap4, Syt13, Tgfb2, Tnfrsf26, Tusc5, Uck2, Uox, Wscd2, Wt1
L	down	5	4930548J01Rik, Camsap3, Ces2g, Serpinb9d, Syce1l
KTL	up	31	2010300C02Rik, 4930550L24Rik, Abcb1b, Abcb4, Acsgb1, Ccl8, Cfb, Chd7, Cryaa, D4Bwg0951e, Dhrr3, Dnajc12, Hvcn1, Il2rg, Il34, Lcp1, Mbnl3, Mustn1, Nt5e, Pamr1, Pdgrfb, Plac8, Robo2, Sdpr, Serinc2, Serpina3g, Slc7a3, Tnfrsf26, Tusc5, Uox, Wdfy4

Figure 10

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2020/056605

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12Q1/6886
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, 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	HONGJUAN ZHAO ET AL: "Gene Expression Profiling Predicts Survival in Conventional Renal Cell Carcinoma", PLOS MEDICINE, vol. 3, no. 1, 6 December 2005 (2005-12-06), page e13, XP055621558, DOI: 10.1371/journal.pmed.0030013 the whole document ----- -/--	1-11



Further documents are listed in the continuation of Box C.



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

4 May 2020

Date of mailing of the international search report

14/05/2020

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Bruma, Anja

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2020/056605

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	PARET CLAUDIA ET AL: "Inflammatory Protein Serum Amyloid A1 Marks a Subset of Conventional Renal Cell Carcinomas with Fatal Outcome", EUROPEAN UROLOGY, vol. 57, no. 5, 1 May 2010 (2010-05-01), - 27 August 2009 (2009-08-27), pages 859-866, XP029877126, ISSN: 0302-2838, DOI: 10.1016/J.EURURO.2009.08.014 the whole document -----	1-11
X	EGBERT OOSTERWIJK ET AL: "Basic Research in Kidney Cancer", EUROPEAN UROLOGY, ELSEVIER, AMSTERDAM, NL, vol. 60, no. 4, 24 June 2011 (2011-06-24), pages 622-633, XP028276639, ISSN: 0302-2838, DOI: 10.1016/J.EURURO.2011.06.048 [retrieved on 2011-06-28] the whole document -----	1-11
X	KALLAKURY B V S ET AL: "INCREASED EXPRESSION OF MATRILYSIN (MMP7) AND STROMELYSIN (MMP3, MMP10, MMP11) PROTEINS AS PREDICTORS OF SURVIVAL IN RCCS", MODERN PATHOLOGY, NATURE PUBLISHING GROUP, GB, vol. 16, no. 1, 1 January 2003 (2003-01-01), XP008054954, ISSN: 0893-3952 707 -----	1-11
X	WO 2016/126883 A1 (CEDARS SINAI MEDICAL CENTER [US]; UNIV DUKE [US]) 11 August 2016 (2016-08-11) claims 1-53 examples 1-3 -----	1-11
X	CHRISTOPHER J. RICKETTS ET AL: "Gender Specific Mutation Incidence and Survival Associations in Clear Cell Renal Cell Carcinoma (CCRCC)", PLOS ONE, vol. 10, no. 10, 20 October 2015 (2015-10-20), pages 1-10, XP055498875, DOI: 10.1371/journal.pone.0140257 the whole document ----- -/--	1-11

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2020/056605

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	"Abstracts from 38th Congress of the Soci��t�� Internationale d'Urologie Seoul Dragon City, October 4-7, 2018 ED - G��mez Reynaldo G; Elliott Sean P", WORLD JOURNAL OF UROLOGY, SPRINGER INTERNATIONAL, DE, vol. 36, no. 1, 4 October 2018 (2018-10-04), pages 1-380, XP036605872, ISSN: 0724-4983, DOI: 10.1007/S00345-018-2492-3 [retrieved on 2018-10-04] MP-12.02 -----	1-11

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2020/056605

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2016126883 A1	11-08-2016	US 2018044736 A1 WO 2016126883 A1	15-02-2018 11-08-2016
