

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
15 December 2011 (15.12.2011)

PCT

(10) International Publication Number
WO 2011/153684 A1

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

(21) International Application Number:
PCT/CN2010/073659

(22) International Filing Date:
8 June 2010 (08.06.2010)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant (for all designated States except US):
BIOMERIEUX [FR/FR]; Chemin de l'Orme, F-69280
Marcy L'etoile (FR).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **YE, Xun** [CN/CN];
Rm 506, Building 2, Xietu Road 2200, Xuhui District,
Shanghai 200032 (CN). **WU, Fei** [CN/CN]; Rm 801,
Building 25, East Guo Quan Road 99, Shanghai 200433
(CN). **XU, Qinghua** [CN/CN]; Room 302, Module 4,
Building 3, Hanggang Nanyuan, Hangzhou, Zhejiang
310022 (CN). **LIU, Fang** [CN/CN]; 306 Room, 351 Al-
ley, DongAn Road, Shanghai 200032 (CN). **MENG, Xia**
[CN/CN]; No 3, Apple Court, 88 Huanzhen South Road,
Shanghai 201103 (CN). **MOUGIN, Bruno** [FR/FR]; 29
rue Lamartine, F-69003 Lyon (FR).

(74) Agent: **NTD PATENT & TRADEMARK AGENCY
LIMITED BEIJING OFFICE**; 10th Floor, Block A, In-
vestment Plaza, 27 Jinrongdajie, Beijing 100033 (CN).

(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, 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, IS, JP, KE, KG, KM, KN, KP,
KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD,
ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI,
NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, 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, SD, SL, SZ, TZ, UG,
ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, 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, SE, SI, SK,
SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ,
GW, ML, MR, NE, SN, TD, TG).

Published:

- with international search report (Art. 21(3))
- with sequence listing part of description (Rule 5.2(a))

(54) Title: METHOD AND KIT FOR THE PROGNOSIS OF COLORECTAL CANCER

(57) Abstract: The present invention discloses a method and kit for the prognosis of colorectal cancer, especially a method comprising: a) obtaining the peripheral blood sample and extracting total RNA from the blood sample, b) contacting the total RNA with at least one reagent that is specific for at least one NK cell gene and no more than 25 specific reagents for 25 NK cell genes, c) determining the expression level of the at least one NK cell gene and of the most 25 NK cell genes to obtain an expression profile for the patient, and d) performing analysis of the expression profile of the patient.



WO 2011/153684 A1

Method and kit for the prognosis of colorectal cancer

Field of the invention

5 The present invention relates to the prognosis of a colorectal cancer, especially to a method and kit for prognosis such a cancer.

Background

10 Colorectal cancer (CRC), also called colon cancer or large bowel cancer is the fifth most common form of cancer in the United States, the fourth common cancer in China and the third leading cause of cancer-related death in Europe. The early detection of CRC remains a major public health challenge.
15 Indeed, CRC is often curable particularly when diagnosed at early stages. Several screening strategies are already in place in various countries. Conventional CRC screening tests include fecal occult blood test (FOBT), sigmoidoscopy, colonoscopy, double contrast barium enema, or digital rectal
20 examination. All of them have advantages and limitations, but compliance remains less than expected mainly due to logistics or discomfort for the patients.

 Search for blood biomarkers aimed at early detection of CRC became a focus since several years, especially for its
25 convenience. Meantime, blood-based test feasibility was supported by very few studies, which have shown that gene biomarkers in blood could differentiate CRC patients from controls. These studies were based on the flow cytometry that is a technique for counting and examining microscopic
30 particles, such as cells by suspending them in a stream of fluid and passing them by an electronic detection apparatus.

 The present inventors have found that differentially expressed genes were mostly associated to immune cell

activation and trafficking. Especially, they have shown that Natural Killer cells (NK cells) represent important biomarkers in peripheral blood samples. They did not used classical technique of flow cytometry but the determination of differential expression of genes from whole blood. It is non usual to determine an expression level of genes via the analysis of transcripts in whole blood, because it is commonly admitted by the persons skilled in the art that it is very difficult to retrieve a specific information when it is diluted in a complex mixture of RNAs (total RNA) without a step of specific purification. An advantage of the present method is also to avoid the step of purification of RNA.

Accordingly, the present invention relates to a method for determining the prognosis of a colorectal cancer in a peripheral blood sample from a patient, the method comprising:

a) obtaining the peripheral blood sample and extracting total RNA from the blood sample,

b) contacting the total RNA with at least one reagent that is specific for at least one NK cell gene and no more than 25 specific reagents for 25 NK cell genes,

c) determining the expression level of the at least one NK cell gene and of the most 25 NK cell genes to obtain an expression profile for the patient,

d) performing analysis of the expression profile of the patient with expression profiles of NK cell genes from patients previously clinically classified as a good prognosis and expression profiles of NK cell genes from patients previously classified as a poor prognosis, wherein

- if the expression profile for the patient is clustered with the expression profiles from patients previously clinically classified as a poor prognosis, then the patient is determined to have a poor prognosis, and

- if the expression profile for the patient is clustered with the expression profiles from patients previously clinically classified as a good prognosis, then the patient is determined to have a good prognosis.

5

Especially in the above step b) the total RNA is brought into contact with at least one reagent is specific for at least one NK cell gene and no more than 25 specific reagents for 25 NK cell genes, said NK cell gene comprising the nucleic acid sequences set forth in SEQ ID NOs: 1 to 12, wherein the at least one reagent is specific for at least one NK cell gene selected from the group consisting of:

- (i) KLRB1 gene comprising a full length sequence such as identified in SEQ ID NO: 1,
 - 15 (ii) KLRC2 gene comprising a full length sequence such as identified in SEQ ID NOs: 2, 3 or 4,
 - (iii) KLRC3 gene comprising a full length sequence such as identified in SEQ ID NOs: 5, 6 or 7,
 - (iv) KLRD1 gene comprising a full length sequence such as identified in SEQ ID NOs: 8, 9, 10, 11, or 12, and
 - 20 (v) KLRK1 gene comprising a full length sequence such as identified in SEQ ID NO: 13, and
- the expression level of the at least one NK cell gene is determined in step c) to obtain the expression profile for the patient.

25

The expression level of at least one of the above genes is a sufficient information for predicting a risk of CRC, as detailed in the experimental data.

30

In one embodiment in step b) the total RNA is brought into contact with reagents specific for a combination of least 5 NK cell genes and no more than 25 NK cell genes, wherein the reagents include at least reagents specific for the NK cell genes consisting of:

35

(i) KLRB1 gene comprising a full length sequence such as identified in SEQ ID NO: 1,

(ii) KLRC2 gene comprising a full length sequence such as identified in SEQ ID NOs: 2, 3 or 4,

5 (iii) KLRC3 gene comprising a full length sequence such as identified in SEQ ID NOs: 5, 6 or 7,

(iv) KLRD1 gene comprising a full length sequence such as identified in SEQ ID NOs: 8, 9, 10, 11, or 12, and

10 (v) KLRK1 comprising a full length sequence such as identified in SEQ ID NO: 13,

the expression level of at least said 4 NK cell genes is determined in step c) to obtain the expression profile for the patient.

15 Furthermore, in step b) the total RNA can be brought into contact with at least one reagent specific for at least one target cell gene and no more than 5 specific reagents for 5 target cell genes, said target cell gene comprising the nucleic acid sequences set forth in SEQ ID NOs: 12 to 24,
20 wherein the at least one reagent is specific for at least one target cell gene selected from the group consisting of:

(i) GZMB gene comprising a full length sequence such as identified in SEQ ID NO: 14, 15, 16 or 17,

25 (ii) CD247 gene comprising a full length sequence such as identified in SEQ ID NO: 18, 19 or 20,

(iii) RRAS2 gene comprising a full length sequence such as identified in SEQ ID NO: 21 or 22, and

(iv) SH2D1B gene comprising a full length sequence such as identified in SEQ ID NO: 23 or 24, and

30 (v) LCK gene comprising a full length sequence such as identified in SEQ ID NO: 25, 26, 27, 28, 29 or 30, and

the expression level of the at least one cell gene is determined in step c) to obtain the expression profile for the patient.; and in one embodiment the total RNA is brought into

contact with reagents specific for a combination 5 target cell genes, wherein the reagents are specific for the target cell genes consisting of:

- (i) GZMB gene comprising a full length sequence such as
5 identified in SEQ ID NO: 14, 15, 16 or 17,
 - (ii) CD247 gene comprising a full length sequence such as identified in SEQ ID NO: 18, 19 or 20,
 - (iii) RRAS2 gene comprising a full length sequence such as identified in SEQ ID NO: 21 or 22, and
 - 10 (iv) SH2D1B gene comprising a full length sequence such as identified in SEQ ID NO: 23 or 24, and
 - (v) LCK gene comprising a full length sequence such as identified in SEQ ID NO: 25, 26, 27, 28, 29 or 30, and
- the expression level of the at least 5 cell gene is determined
15 in step c) to obtain the expression profile for the patient.

In another embodiment, in step b) the total RNA is further brought into contact with at least one reagent specific for at least one target cell gene and no more than
20 100 specific reagents for 100 target cell genes, said target cell gene comprising the nucleic acid sequences set forth in SEQ ID NOs: 25 to 59, wherein the at least one reagent is specific for at least one target cell gene selected from the group consisting of:

- 25 (i) MRPS6 gene comprising a full length sequence such as identified in SEQ ID NO: 31, 32, or 33,
- (ii) SPRY4 gene comprising a full length sequence such as identified in SEQ ID NO: 34,
- (iii) NEAT1 gene comprising a full length sequence such as
30 identified in SEQ ID NO: 35,
- (iv) CYBB gene comprising a full length sequence such as identified in SEQ ID NO: 36,
- (v) DUSP2 gene comprising a full length sequence such as identified in SEQ ID NO: 37,.

- (vi) PDEAD gene comprising a full length sequence such as identified in SEQ ID NO: 38 or 39,
- (vii) SH2D2A gene comprising a full length sequence such as identified in SEQ ID NO: 40, 41 or 42,
- 5 (viii) INSR gene comprising a full length sequence such as identified in SEQ ID NO: 43 or 44,
- (ix) ITGAM gene comprising a full length sequence such as identified in SEQ ID NO: 45,
- (x) VCAN gene comprising a full length sequence such as
- 10 identified in SEQ ID NO: 46, 47, 48 or 49,
- (xi) CD 163 gene comprising a full length sequence such as identified in SEQ ID NO: 50, or 51.
- (xii) P2RY10 gene comprising a full length sequence such as identified in SEQ ID NO: 52 or 53,
- 15 (xii) CD226 gene comprising a full length sequence such as identified in SEQ ID NO: 54,
- (xiii) MRPL10 gene comprising a full length sequence such as identified in SEQ ID NO: 55 or 56,
- (xiv) ITPRIPL2 gene comprising a full length sequence such as
- 20 identified in SEQ ID NO: 57,
- (xv) CD2 gene comprising a full length sequence such as identified in SEQ ID NO: 58, and
- (xvi) NUDT16 gene comprising a full length sequence such as identified in SEQ ID NO: 59, and.
- 25 the expression level of the at least one cell gene is determined in step c) to obtain the expression profile for the patient.

Especially, in step b) the total RNA is brought into

30 contact with reagents specific for a combination of least 17 target cell genes and no more than 100 target cell genes, wherein the reagents include at least reagents specific for the target cell genes consisting of:

- (i) MRPS6 gene comprising a full length sequence such as identified in SEQ ID NO: 31, 32, or 33,
- (ii) SPRY4 gene comprising a full length sequence such as identified in SEQ ID NO: 34,
- 5 (iii) NEAT1 gene comprising a full length sequence such as identified in SEQ ID NO: 35,
- (iv) CYBB gene comprising a full length sequence such as identified in SEQ ID NO: 36,
- (v) DUSP2 gene comprising a full length sequence such as
10 identified in SEQ ID NO: 37,.
- (vi) PDEAD gene comprising a full length sequence such as identified in SEQ ID NO: 38 or 39,
- (vii) SH2D2A gene comprising a full length sequence such as identified in SEQ ID NO: 40, 41 or 42,
- 15 (viii) INSR gene comprising a full length sequence such as identified in SEQ ID NO: 43 or 44,
- (ix) ITGAM gene comprising a full length sequence such as identified in SEQ ID NO: 45,
- (x) VCAN gene comprising a full length sequence such as
20 identified in SEQ ID NO: 46, 47, 48 or 49,
- (xi) CD 163 gene comprising a full length sequence such as identified in SEQ ID NO: 50, or 51.
- (xii) P2RY10 gene comprising a full length sequence such as identified in SEQ ID NO: 52 or 53,
- 25 (xiii) CD226 gene comprising a full length sequence such as identified in SEQ ID NO: 54,
- (xiv) MRPL10 gene comprising a full length sequence such as identified in SEQ ID NO: 55 or 56,
- (xv) ITPRIPL2 gene comprising a full length sequence such as
30 identified in SEQ ID NO: 57,
- (xvi) CD2 gene comprising a full length sequence such as identified in SEQ ID NO: 58, and
- (xvi) NUDT16 gene comprising a full length sequence such as identified in SEQ ID NO: 59, and

the expression level of the at least 17 cell genes is determined in step c) to obtain the expression profile for the patient.

5 More precisely, in the methods described above the at least one specific reagent of step b) comprises at least one hybridization probe, in particular at least one hybridization probe and at least one primer and more particularly at least one hybridization probe and two primers.

10 Total RNA comprises transfer RNAs (tRNA), messenger RNAs (mRNAs), such as the mRNAs transcribed from the target gene, but also transcribed from any other gene, and ribosomal RNAs.

15 By way of indication, the extraction of total RNA can be carried out by: a step consisting of lysis of the cells present in the blood sample, in order to release the nucleic acids contained in the cells of the patient. By way of example, use may be made of the methods of lysis as described in patent
20 applications: WO 00/05338 regarding mixed magnetic and mechanical lysis, WO 99/53304 regarding electrical lysis, WO 99/15321 regarding mechanical lysis. Those skilled in the art may use other well-known methods of lysis, such as thermal or osmotic shocks or chemical lyses using chaotropic agents such
25 as guanidinium salts (U.S. Pat. No. 5,234,809). It is also possible to provide an additional step for separating the nucleic acids from the other cellular constituents released in the lysis step. This generally makes it possible to concentrate the nucleic acids. By way of example, use may be
30 made of magnetic particles optionally coated with oligonucleotides, by adsorption or covalence (in this respect, see U.S. Pat. No. 4,672,040 and U.S. Pat. No. 5,750,338), and the nucleic acids which are bound to these magnetic particles can thus be purified by means of a washing step. This nucleic

acid purification step is particularly advantageous if it is desired to subsequently amplify said nucleic acids. A particularly advantageous embodiment of these magnetic particles is described in patent applications: WO-A-97/45202 and WO-A-99/35500.

The term "specific reagent" is intended to mean a reagent which, when it is brought into contact with biological material as defined above, binds with the material specific for said target gene. By way of indication, when the specific reagent and the biological material are of nucleic origin, bringing the specific reagent into contact with the biological material allows the specific reagent to hybridize with the material specific for the target gene. The term "hybridization" is intended to mean the process during which, under appropriate conditions, two nucleotide fragments bind with stable and specific hydrogen bonds so as to form a double-stranded complex. These hydrogen bonds form between the complementary adenine (A) and thymine (T) (or uracile (U)) bases (this is referred to as an A-T bond) or between the complementary guanine (G) and cytosine (C) bases (this is referred to as a G-C bond). The hybridization of two nucleotide fragments may be complete (reference is then made to complementary nucleotide fragments or sequences), i.e. the double-stranded complex obtained during this hybridization comprises only A-T bonds and C-G bonds. This hybridization may be partial (reference is then made to sufficiently complementary nucleotide fragments or sequences), i.e. the double-stranded complex obtained comprises A-T bonds and C-G bonds that make it possible to form the double-stranded complex, but also bases not bound to a complementary base. The hybridization between two nucleotide fragments depends on the working conditions that are used, and in particular on the stringency. The stringency is defined in particular as a

function of the base composition of the two nucleotide fragments, and also by the degree of mismatching between two nucleotide fragments. The stringency can also depend on the reaction parameters, such as the concentration and the type of ionic species present in the hybridization solution, the nature and the concentration of denaturing agents and/or the hybridization temperature. All these data are well known and the appropriate conditions can be determined by those skilled in the art. In general, depending on the length of the nucleotide fragments that it is intended to hybridize, the hybridization temperature is between approximately 20 and 70.degree. C., in particular between 35 and 65.degree. C. in a saline solution at a concentration of approximately 0.5 to 1 M. A sequence, or nucleotide fragment, or oligonucleotide, or polynucleotide, is a series of nucleotide motifs assembled together by phosphoric ester bonds, characterized by the informational sequence of the natural nucleic acids, capable of hybridizing to a nucleotide fragment, it being possible for the series to contain monomers having different structures and to be obtained from a natural nucleic acid molecule and/or by genetic recombination and/or by chemical synthesis. A motif is a derivative of a monomer which may be a natural nucleotide of nucleic acid, the constitutive elements of which are a sugar, a phosphate group and a nitrogenous base; in DNA, the sugar is deoxy-2-ribose, in RNA, the sugar is ribose; depending on whether DNA or RNA is involved, the nitrogenous base is selected from adenine, guanine, uracile, cytosine and thymine; alternatively the monomer is a nucleotide that is modified in at least one of the three constitutive elements; by way of example, the modification may occur either at the level of the bases, with modified bases such as inosine, methyl-5-deoxycytidine, deoxyuridine, dimethylamino-5-deoxyuridine, diamino-2,6-purine, bromo-5-deoxyuridine or any other modified base capable of hybridization, or at the level of the sugar,

for example the replacement of at least one deoxyribose with a polyamide (P. E. Nielsen et al, Science, 254, 1497-1500 (1991)[3]), or else at the level of the phosphate group, for example its replacement with esters in particular selected from diphosphates, alkyl- and arylphosphonates and phosphorothioates.

According to a specific embodiment of the invention, the specific reagent comprises at least one hybridization probe or at least one hybridization probe and at least one primer which is specific for the target gene or at least one hybridization probe and two primers specific for the target genes.

For the purpose of the present invention, the term "amplification primer" is intended to mean a nucleotide fragment comprising from 5 to 100 nucleotides, preferably from 15 to 30 nucleotides that allow the initiation of an enzymatic polymerization, for instance an enzymatic amplification reaction. The term "enzymatic amplification reaction" is intended to mean a process which generates multiple copies of a nucleotide fragment through the action of at least one enzyme. Such amplification reactions are well known to those skilled in the art and mention may in particular be made of the following techniques: PCR (polymerase chain reaction), as described in U.S. Pat. No. 4,683,195, U.S. Pat. No. 4,683,202 and U.S. Pat. No. 4,800,159, LCR (ligase chain reaction), disclosed, for, example, in patent application EP 0 201 184, RCR (repair chain reaction), described in patent application WO 90/01069, 3SR (self sustained sequence replication) with patent application WO 90/06995, NASBA (nucleic acid sequence-based amplification) with patent application WO 91/02818, TMA (transcription mediated amplification) with U.S. Pat. No. 5,399,491 and RT-PCR.

When the enzymatic amplification is a PCR, the specific reagent comprises at least two amplification primers, specific for a target gene, that allow the amplification of the material specific for the target gene. The material specific for the target gene then preferably comprises a complementary DNA obtained by reverse transcription of messenger RNA derived from the target gene (reference is then made to target-gene-specific cDNA) or a complementary RNA obtained by transcription of the cDNAs specific for a target gene (reference is then made to target-gene-specific cRNA). When the enzymatic amplification is a PCR carried out after a reverse transcription reaction, reference is made to RT-PCR.

The term "hybridization probe" is intended to mean a nucleotide fragment comprising at least 5 nucleotides, such as from 5 to 100 nucleotides, in particular from 10 to 75 nucleotides, such as 15-35 nucleotides and 60-70 nucleotides, having a hybridization specificity under given conditions so as to form a hybridization complex with the material specific for a target gene. In the present invention, the material specific for the target gene may be a nucleotide sequence included in a messenger RNA derived from the target gene (reference is then made to target-gene-specific mRNA), a nucleotide sequence included in a complementary DNA obtained by reverse transcription of said messenger RNA (reference is then made to target-gene-specific cDNA), or else a nucleotide sequence included in a complementary RNA obtained by transcription of said cDNA as described above (reference will then be made to target-gene-specific cRNA). The hybridization probe may include a label for its detection. The term "detection" is intended to mean either a direct detection such as a counting method, or an indirect detection by a method of detection using a label. Many methods of detection exist for detecting nucleic acids (see, for example, Kricka et al.,

Clinical Chemistry, 1999, no 45 (4), p. 453-458 or Keller G. H. et al., DNA Probes, 2nd Ed., Stockton Press, 1993, sections 5 and 6, p. 173-249 . The term "label" is intended to mean a tracer capable of generating a signal that can be detected. A
5 non limiting list of these tracers includes enzymes which produce a signal that can be detected, for example, by colorimetry, fluorescence or luminescence, such as horseradish peroxidase, alkaline phosphatase, beta-galactosidase, glucose-6-phosphate dehydrogenase; chromophores such as fluorescent,
10 luminescent or dye compounds; electron dense groups detectable by electron microscopy or by virtue of their electrical properties such as conductivity, by amperometry or voltametry methods, or by impedance measurement; groups that can be detected by optical methods such as diffraction, surface
15 plasmon resonance, or contact angle variation, or by physical methods such as atomic force spectroscopy, tunnel effect, etc.; radioactive molecules such as ^{32}P , ^{35}S or ^{125}I .

For the purpose of the present invention, the
20 hybridization probe may be a "detection" probe. In this case, the "detection" probe is labeled by means of a label. The detection probe may in particular be a "molecular beacon" detection probe as described by Tyagi & Kramer (Nature biotech, 1996, 14:303-308). These "molecular beacons" become
25 fluorescent during the hybridization. They have a stem-loop-type structure and contain a fluorophore and a "quencher" group. The binding of the specific loop sequence with its complementary target nucleic acid sequence causes the stem to unroll and the emission of a fluorescent signal during
30 excitation at the appropriate wavelength. The detection probe in particular may be a "reporter probe" comprising a "color-coded barcode" according to NanoStringTM's technology.

For the detection of the hybridization reaction, use may be made of target sequences that have been labeled, directly (in particular by the incorporation of a label within the target sequence) or indirectly (in particular using a detection probe as defined above). It is in particular possible to carry out, before the hybridization step, a step consisting in labeling and/or cleaving the target sequence, for example using a labeled deoxy-ribonucleotide triphosphate during the enzymatic amplification reaction. The cleavage may be carried out in particular by the action of imidazole or of manganese chloride. The target sequence may also be labeled after the amplification step, for example by hybridizing a detection probe according to the sandwich hybridization technique described in document WO 91/19812. Another specific preferred method of labeling nucleic acids is described in application FR 2780059.

According to a preferred embodiment of the invention, the detection probe comprises a fluorophore and a quencher. According to an even more preferred embodiment of the invention, the hybridization probe comprises an FAM (6-carboxy-fluorescein) or ROX (6-carboxy-X-rhodamine) fluorophore at its 5' end and a quencher (Dabsyl) at its 3' end.

The hybridization probe may also be a "capture" probe. In this case, the "capture" probe is immobilized or can be immobilized on a solid substrate by any appropriate means, i.e. directly or indirectly, for example by covalence or adsorption. As solid substrate, use may be made of synthetic materials or natural materials, optionally chemically modified, in particular polysaccharides such as cellulose-based materials, for example paper, cellulose derivatives such as cellulose acetate and nitrocellulose or dextran, polymers, copolymers,

in particular based on styrene-type monomers, natural fibers such as cotton, and synthetic fibers such as nylon; inorganic materials such as silica, quartz, glasses or ceramics; lattices; magnetic particles; metal derivatives, gels, etc. The solid substrate may be in the form of a microtitration plate, of a membrane as described in application WO-A-94/12670 or of a particle. It is also possible to immobilize on the substrate several different capture probes, each being specific for a target gene. In particular, a biochip on which a large number of probes can be immobilized may be used as substrate. The term "biochip" is intended to mean a solid substrate that is small in size, to which a multitude of capture probes are attached at predetermined positions. The biochip, or DNA chip, concept dates from the beginning of the 1990s. It is based on a multidisciplinary technology that integrates microelectronics, nucleic acid chemistry, image analysis and information technology. The operating principle is based on a foundation of molecular biology: the hybridization phenomenon, i.e. the pairing, by complementarity, of the bases of two DNA and/or RNA sequences. The biochip method is based on the use of capture probes attached to a solid substrate, on which probes a sample of target nucleotide fragments directly or indirectly labeled with fluorochromes is made to act. The capture probes are positioned specifically on the substrate or chip and each hybridization gives a specific piece of information, in relation to the target nucleotide fragment. The pieces of information obtained are cumulative, and make it possible, for example, to quantify the level of expression of one or more target genes. In order to analyze the expression of a target gene, a substrate comprising a multitude of probes, which correspond to all or part of the target gene, which is transcribed to mRNA, can then be prepared. For the purpose of the present invention, the term "low-density substrate" is intended to mean a substrate comprising fewer than 50 probes.

For the purpose of the present invention, the term "medium-density substrate" is intended to mean a substrate comprising from 50 probes to 10 000 probes. For the purpose of the present invention, the term "high-density substrate" is intended to mean a substrate comprising more than 10 000 probes.

The cDNAs or cRNAs specific for a target gene that it is desired to analyze are then hybridized, for example, to specific capture probes. After hybridization, the substrate or chip is washed and the labeled cDNA or cRNA/capture probe complexes are revealed by means of a high-affinity ligand bound, for example, to a fluorochrome-type label. The fluorescence is read, for example, with a scanner and the analysis of the fluorescence is processed by information technology. By way of indication, mention may be made of the DNA chips developed by the company Affymetrix ("Accessing Genetic Information with High-Density DNA arrays", M. Chee et al., Science, 1996, 274, 610-614. "Light-generated oligonucleotide arrays for rapid DNA sequence analysis", A. Caviani Pease et al., Proc. Natl. Acad. Sci. USA, 1994, 91, 5022-5026), for molecular diagnoses. In this technology, the capture probes are generally small in size, around 25 nucleotides. Other examples of biochips are given in the publications by G. Ramsay, Nature Biotechnology, 1998, No. 16, p. 40-44; F. Ginot, Human Mutation, 1997, No. 10, p. 1-10 ; J. Cheng et al, Molecular diagnosis, 1996, No. 1 (3), p. 183-200 ; T. Livache et al, Nucleic Acids Research, 1994, No. 22 (15), p. 2915-2921 J. Cheng et al, Nature Biotechnology, 1998, No. 16, p. 541-546 or in U.S. Pat. No. 4,981,783, U.S. Pat. No. 5,700,637, U.S. Pat. No. 5,445,934, U.S. Pat. No. 5,744,305 and U.S. Pat. No. 5,807,522. The main characteristic of the solid substrate should be to conserve the hybridization characteristics of the capture probes on the target nucleotide

fragments while at the same time generating a minimum background noise for the method of detection. Three main types of fabrication can be distinguished for immobilizing the probes on the substrate.

5

First of all, there is a first technique which consists in depositing pre-synthesized probes. The attachment of the probes is carried out by direct transfer, by means of micropipettes or of microdots or by means of an inkjet device.

10 This technique allows the attachment of probes having a size ranging from a few bases (5 to 10) up to relatively large sizes of 60 bases (printing) to a few hundred bases (microdeposition).

15 Printing is an adaptation of the method used by inkjet printers. It is based on the propulsion of very small spheres of fluid (volume $<1 \text{ nl}$) at a rate that may reach 4000 drops/second. The printing does not involve any contact between the system releasing the fluid and the surface on
20 which it is deposited.

Microdeposition consists in attaching long probes of a few tens to several hundred bases to the surface of a glass slide. These probes are generally extracted from databases and are in
25 the form of amplified and purified products. This technique makes it possible to produce chips called microarrays that carry approximately ten thousand spots, called recognition zones, of DNA on a surface area of a little less than 4 cm^2 . The use of nylon membranes, referred to as
30 "macroarrays", which carry products that have been amplified, generally by PCR, with a diameter of 0.5 to 1 mm and the maximum density of which is 25 spots/cm^2 , should not however be forgotten. This very flexible technique is used by many laboratories. In the present invention, the latter

technique is considered to be included among biochips. A certain volume of sample can, however, be deposited at the bottom of a microtitration plate, in each well, as in the case in patent applications WO-A-00/71750 and FR 00/14896, or a certain number of drops that are separate from one another can be deposited at the bottom of one and the same Petri dish, according to another patent application, FR 00/14691.

The second technique for attaching the probes to the substrate or chip is called in situ synthesis. This technique results in the production of short probes directly at the surface of the chip. It is based on in situ oligonucleotide synthesis (see, in particular, patent applications WO 89/10977 and WO 90/03382) and is based on the oligonucleotide synthesizer process. It consists in moving a reaction chamber, in which the oligonucleotide extension reaction takes place, along the glass surface.

Finally, the third technique is called photolithography, which is a process that is responsible for the biochips developed by Affymetrix. It is also an in situ synthesis. Photolithography is derived from microprocessor techniques. The surface of the chip is modified by the attachment of photolabile chemical groups that can be light-activated. Once illuminated, these groups are capable of reacting with the 3' end of an oligonucleotide. By protecting this surface with masks of defined shapes, it is possible to selectively illuminate and therefore activate areas of the chip where it is desired to attach one or other of the four nucleotides. The successive use of different masks makes it possible to alternate cycles of protection/reaction and therefore to produce the oligonucleotide probes on spots of approximately a few tens of square micrometers (μm^2). This resolution makes it possible to create up to several hundred thousand spots on a

surface area of a few square centimeters (cm^2). Photolithography has advantages: in bulk in parallel, it makes it possible to create a chip of N-mers in only 4.times.N cycles. All these techniques can be used with the present invention. According to a preferred embodiment of the invention, the at least one specific reagent of step b) defined above comprises at least one hybridization probe which is preferably immobilized on a substrate. This substrate is preferably a low-, high- or medium-density substrate as defined above.

These hybridization steps on a substrate comprising a multitude of probes may be preceded by an enzymatic amplification reaction step, as defined above, in order to increase the amount of target genetic material.

In step c), the determination of the expression level of a target gene can be carried out by any of the protocols known to those skilled in the art. In general, the expression of a target gene can be analyzed by detecting the mRNAs (messenger RNAs) that are transcribed from the target gene at a given moment.

The invention preferably relates to the determination of the expression level of a target gene by detection of the mRNAs derived from this target gene according to any of the protocols well known to those skilled in the art. According to a specific embodiment of the invention, the expression level of several target genes is determined simultaneously, by detection of several different mRNAs, each mRNA being derived from a target gene.

When the specific reagent comprises at least one amplification primer, it is possible, to determine the expression level of the target gene in the following way: 1)

After having extracted the total RNA (comprising the transfer RNAs (tRNAs), the ribosomal RNAs (rRNAs) and the messenger RNAs (mRNAs)) from the whole blood, a reverse transcription step is carried out in order to obtain the complementary DNAs (or cDNAs) of said mRNAs. By way of indication, this reverse transcription reaction can be carried out using a reverse transcriptase enzyme which makes it possible to obtain, from an RNA fragment, a complementary DNA fragment. The reverse transcriptase enzyme from AMV (Avian Myoblastosis Virus) or from MMLV (Moloney Murine Leukaemia Virus) can in particular be used. When it is more particularly desired to obtain only the cDNAs of the mRNAs, this reverse transcription step is carried out in the presence of nucleotide fragments comprising only thymine bases (polyT), which hybridize by complementarity to the polyA sequence of the mRNAs so as to form a polyT-polyA complex which then serves as a starting point for the reverse transcription reaction carried out by the reverse transcriptase enzyme. cDNAs complementary to the mRNAs derived from a target gene (target-gene-specific cDNA) and cDNAs complementary to the mRNAs derived from genes other than the target gene (cDNAs not specific for the target gene) are then obtained. 2) The amplification primer(s) specific for a target gene is (are) brought into contact with the target-gene-specific cDNAs and the cDNAs not specific for the target gene. The amplification primer(s) specific for a target gene hybridize(s) with the target-gene-specific cDNAs and a predetermined region, of known length, of the cDNAs originating from the mRNAs derived from the target gene is specifically amplified. The cDNAs not specific for the target gene are not amplified, whereas a large amount of target-gene-specific cDNAs is then obtained. For the purpose of the present invention, reference is made, without distinction, to "target-gene-specific cDNAs" or to "cDNAs originating from the mRNAs derived from the target gene". This step can be carried

out in particular by means of a PCR-type amplification reaction or by any other amplification technique as defined above. By PCR, it is also possible to simultaneously amplify several different cDNAs, each one being specific for different target genes, by using several pairs of different amplification primers, each one being specific for a target gene: reference is then made to multiplex amplification. 3) The expression of the target gene is determined by detecting and quantifying the target-gene-specific cDNAs obtained in step 2) above. This detection can be carried out after electrophoretic migration of the target-gene-specific cDNAs according to their size. The gel and the medium for the migration can include ethidium bromide so as to allow direct detection of the target-gene-specific cDNAs when the gel is placed, after a given migration period, on a UV (ultraviolet)-ray light table, through the emission of a light signal. The greater the amount of target-gene-specific cDNAs, the brighter this light signal. These electrophoresis techniques are well known to those skilled in the art. The target-gene-specific cDNAs can also be detected and quantified using a quantification range obtained by means of an amplification reaction carried out until saturation. In order to take into account the variability in enzymatic efficiency that may be observed during the various steps (reverse transcription, PCR, etc.), the expression of a target gene of various groups of patients can be normalized by simultaneously determining the expression of a "housekeeping" gene, the expression of which is similar in the various groups of patients. By realizing a ratio of the expression of the target gene to the expression of the housekeeping gene, i.e. by realizing a ratio of the amount of target-gene-specific cDNAs to the amount of housekeeping-gene-specific cDNAs, any variability between the various experiments is thus corrected. Those skilled in the art may refer in particular to the following publications:

Bustin S A, J Mol Endocrinol, 2002, 29: 23-39; Giulietti A Methods, 2001, 25: 386-401.

When the specific reagent comprises at least one
5 hybridization probe, the expression of a target gene can be
determined in the following way: 1) After having extracted the
total RNA from the whole blood, a reverse transcription step
is carried out as described above in order to obtain cDNAs
complementary to the mRNAs derived from a target gene (target-
10 gene-specific cDNA) and cDNAs complementary to the mRNAs
derived from genes other than the target gene (cDNA not
specific for the target gene). 2) All the cDNAs are brought
into contact with a substrate, on which are immobilized
capture probes specific for the target gene whose expression
15 it is desired to analyze, in order to carry out a
hybridization reaction between the target-gene-specific cDNAs
and the capture probes, the cDNAs not specific for the target
gene not hybridizing to the capture probes. The hybridization
reaction can be carried out on a solid substrate which
20 includes all the materials as indicated above. According to a
preferred embodiment, the hybridization probe is immobilized
on a substrate. Preferably, the substrate is a low-, high- or
medium-density substrate as defined above. The hybridization
reaction may be preceded by a step consisting of enzymatic
25 amplification of the target-gene-specific cDNAs as described
above, so as to obtain a large amount of target-gene-specific
cDNAs and to increase the probability of a target-gene-
specific cDNA hybridizing to a capture probe specific for the
target gene. The hybridization reaction may also be preceded
30 by a step consisting in labeling and/or cleaving the target-
gene-specific cDNAs as described above, for example using a
labeled deoxyribonucleotide triphosphate for the amplification
reaction. The cleavage can be carried out in particular by the
action of imidazole and manganese chloride. The target-gene-

specific cDNA can also be labeled after the amplification step, for example by hybridizing a labeled probe according to the sandwich hybridization technique described in document WO-A-91/19812. Other preferred specific methods for labeling and/or

5 cleaving nucleic acids are described in applications WO 99/65926, WO 01/44507, WO 01/44506, WO 02/090584, WO 02/090319.

3) A step consisting of detection of the hybridization reaction is subsequently carried out. The detection can be carried out by bringing the substrate on which the capture

10 probes specific for the target gene are hybridized with the target-gene-specific cDNAs into contact with a "detection" probe labeled with a label, and detecting the signal emitted by the label. When the target-gene-specific cDNA has been labeled beforehand with a label, the signal emitted by the

15 label is detected directly.

When the at least one specific reagent is brought into contact in step b) comprises at least one hybridization probe, the expression of a target gene can also be determined in the

20 following way: 1) After having extracted the total RNA from the whole blood, a reverse transcription step is carried out as described above in order to obtain the cDNAs of the mRNAs of the biological material. The polymerization of the complementary RNA of the cDNA is subsequently carried out

25 using a T7 polymerase enzyme which functions under the control of a promoter and which makes it possible to obtain, from a DNA template, the complementary RNA. The cRNAs of the cDNAs of the mRNAs specific for the target gene (reference is then made to target-gene-specific cRNA) and the cRNAs of the cDNAs of

30 the mRNAs not specific for the target gene are then obtained.

2) All the cRNAs are brought into contact with a substrate on which are immobilized capture probes specific for the target gene whose expression it is desired to analyze, in order to carry out a hybridization reaction between the target-gene-

specific cRNAs and the capture probes, the cRNAs not specific for the target gene not hybridizing to the capture probes. When it is desired to simultaneously analyze the expression of several target genes, several different capture probes can be immobilized on the substrate, each one being specific for a target gene. The hybridization reaction may also be preceded by a step consisting in labeling and/or cleaving the target-gene-specific cRNAs as described above. 3) A step consisting of detection of the hybridization reaction is subsequently carried out. The detection can be carried out by bringing the substrate on which the capture probes specific for the target gene are hybridized with the target-gene-specific cRNA into contact with a "detection" probe labeled with a label, and detecting the signal emitted by the label. When the target-gene-specific cRNA has been labeled beforehand with a label, the signal emitted by the label is detected directly. The use of cRNA is particularly advantageous when a substrate of biochip type on which a large number of probes are hybridized is used.

The present invention also includes a kit for the prognosis of a colorectal cancer in a peripheral blood sample from a patient comprising at least one specific reagent for at least one NK cell gene and no more than 25 specific reagents for 25 NK cell genes comprising at least the nucleic acid sequences set forth in SEQ NOs 1 to 13, wherein the at least one reagent is specific for at least one NK cell gene selected from the group consisting of:

- (i) KLRB1 gene comprising a full length sequence such as identified in SEQ ID NO: 1,
- (ii) KLRC2 gene comprising a full length sequence such as identified in SEQ ID NOs: 2, 3 or 4,
- (iii) KLRC3 gene comprising a full length sequence such as identified in SEQ ID NOs: 5, 6 or 7,

- (iv) KLRD1 gene comprising a full length sequence such as identified in SEQ ID NOs: 8, 9, 10, 11, or 12, and
- (v) KLRK1 gene comprising a full length sequence such as identified in SEQ ID NO: 13.

5

In one embodiment, the kit comprises a combination of reagents that are specific for the NK cell genes consisting of:

- (i) KLRB1 gene comprising a full length sequence such as identified in SEQ ID NO: 1,
- 10 (ii) KLRC2 gene comprising a full length sequence such as identified in SEQ ID NOs: 2, 3 or 4,
- (iii) KLRC3 gene comprising a full length sequence such as identified in SEQ ID NOs: 5, 6 or 7,
- (iv) KLRD1 gene comprising a full length sequence such as
- 15 identified in SEQ ID NOs: 8, 9, 10, 11, or 12, and
- (v) KLRK1 gene comprising a full length sequence such as identified in SEQ ID NO: 13.

20 In such an embodiment, the specific reagents can targeted a combination of several NK cell genes but no more than 25 NK genes.

Furthermore, the kit can comprise at least one reagent that is specific for at least one target cell gene and no more

25 than 5 target cell genes, said at least one target cell gene being selected from the group consisting of:

- (i) GZMB gene comprising a full length sequence such as identified in SEQ ID NO: 14, 15, 16 or 17,
- (ii) CD247 gene comprising a full length sequence such as
- 30 identified in SEQ ID NO: 18, 19 or 20,
- (iii) RRAS2 gene comprising a full length sequence such as identified in SEQ ID NO: 21 or 22, and
- (iv) SH2D1B gene comprising a full length sequence such as identified in SEQ ID NO: 23 or 24, and

(v) LCK gene comprising a full length sequence such as identified in SEQ ID NO: 25, 26, 27, 28, 29 or 30.

In particular, it comprises 5 reagents that are specific for the target cell genes consisting of:

(i) GZMB gene comprising a full length sequence such as identified in SEQ ID NO: 14, 15, 16 or 17,

(ii) CD247 gene comprising a full length sequence such as identified in SEQ ID NO: 18, 19 or 20,

(iii) RRAS2 gene comprising a full length sequence such as identified in SEQ ID NO: 21 or 22, and

(iv) SH2D1B gene comprising a full length sequence such as identified in SEQ ID NO: 23 or 24, and

(v) LCK gene comprising a full length sequence such as identified in SEQ ID NO: 25, 26, 27, 28, 29 or 30.

In such an embodiment, the specific reagents can targeted a combination of several target cell genes, such as described above but no more than 5 target cell genes.

In another embodiment, the kit such as defined above can comprise at least one reagent that is specific for at least one target cell gene and at the most 100 reagents that are specific for 100 target cell genes, said at least target cell gene being selected from the group consisting of:

(i) MRPS6 gene comprising a full length sequence such as identified in SEQ ID NO: 31, 32, or 33,

(ii) SPRY4 gene comprising a full length sequence such as identified in SEQ ID NO: 34,

(iii) NEAT1 gene comprising a full length sequence such as identified in SEQ ID NO: 35,

(iv) CYBB gene comprising a full length sequence such as identified in SEQ ID NO: 36,

- (v) DUSP2 gene comprising a full length sequence such as identified in SEQ ID NO: 37,.
- (vi) PDEAD gene comprising a full length sequence such as identified in SEQ ID NO: 38 or 39,
- 5 (vii) SH2D2A gene comprising a full length sequence such as identified in SEQ ID NO: 40, 41 or 42,
- (viii) INSR gene comprising a full length sequence such as identified in SEQ ID NO: 43 or 44,
- (ix) ITGAM gene comprising a full length sequence such as
10 identified in SEQ ID NO: 45,
- (x) VCAN gene comprising a full length sequence such as identified in SEQ ID NO: 46, 47, 48 or 49,
- (xi) CD 163 gene comprising a full length sequence such as identified in SEQ ID NO: 50, or 51.
- 15 (xii) P2RY10 gene comprising a full length sequence such as identified in SEQ ID NO: 52 or 53,
- (xii) CD226 gene comprising a full length sequence such as identified in SEQ ID NO: 54,
- (xiii) MRPL10 gene comprising a full length sequence such as
20 identified in SEQ ID NO: 55 or 56,
- (xiv) ITPRIPL2 gene comprising a full length sequence such as identified in SEQ ID NO: 57,
- (xv) CD2 gene comprising a full length sequence such as identified in SEQ ID NO: 58, and
- 25 (xvi) NUDT16 gene comprising a full length sequence such as identified in SEQ ID NO: 59.

And especially, it comprises 17 reagents that are specific for 17 target cell genes consisting of:

- 30 (i) MRPS6 gene comprising a full length sequence such as identified in SEQ ID NO: 31, 32, or 33,
- (ii) SPRY4 gene comprising a full length sequence such as identified in SEQ ID NO: 34,

- (iii) NEAT1 gene comprising a full length sequence such as identified in SEQ ID NO: 35,
- (iv) CYBB gene comprising a full length sequence such as identified in SEQ ID NO: 36,
- 5 (v) DUSP2 gene comprising a full length sequence such as identified in SEQ ID NO: 37, .
- (vi) PDEAD gene comprising a full length sequence such as identified in SEQ ID NO: 38 or 39,
- (vii) SH2D2A gene comprising a full length sequence such as
10 identified in SEQ ID NO: 40, 41 or 42,
- (viii) INSR gene comprising a full length sequence such as identified in SEQ ID NO: 43 or 44,
- (ix) ITGAM gene comprising a full length sequence such as identified in SEQ ID NO: 45,
- 15 (x) VCAN gene comprising a full length sequence such as identified in SEQ ID NO: 46, 47, 48 or 49,
- (xi) CD 163 gene comprising a full length sequence such as identified in SEQ ID NO: 50, or 51.
- (xii) P2RY10 gene comprising a full length sequence such as
20 identified in SEQ ID NO: 52 or 53,
- (xii) CD226 gene comprising a full length sequence such as identified in SEQ ID NO: 54,
- (xiii) MRPL10 gene comprising a full length sequence such as identified in SEQ ID NO: 55 or 56,
- 25 (xiv) ITPRIPL2 gene comprising a full length sequence such as identified in SEQ ID NO: 57,
- (xv) CD2 gene comprising a full length sequence such as identified in SEQ ID NO: 58, and
- (xvi) NUDT16 gene comprising a full length sequence such as
30 identified in SEQ ID NO: 59.

In such an embodiment, the specific reagents can targeted a combination of several target cell genes, such as described above but no more than 100 target cell genes.

As explained above the at least one specific reagent comprises at least one hybridization probe, in particular at least one hybridization probe and at least one primer and more particularly at least one hybridization probe and two primers.

5

Finally, the invention concerns the use of at least one specific reagent for at least one NK cell genes and no more than 25 specific reagents for 25 NK cell genes comprising the nucleic acid sequences set forth in SEQ ID NOs 1 to 12 in the manufacture of a composition for the prognosis of colorectal cancer in a biological sample from a patient, wherein the at least one reagent is specific for at least one NK cell gene comprising a nucleic acid sequence selected from the nucleic acid sequences set forth in anyone of SEQ ID NOs: 1 to 12;

10

especially the use of reagents specific for a combination of at least 5 NK cell genes and no more than 25 NK cell genes in the manufacture of a composition for the prognosis of colorectal cancer in a biological sample from a patient, wherein the reagents are specific for at least 5 NK cell genes comprising a nucleic acid sequence selected from the nucleic acid sequences set forth in SEQ ID NOs: 1, 2 to 4, 5 to 7 and 8 to 12, respectively;

15

20

in particular, the use of reagents specific for a combination of 10 target cell genes in the manufacture of a composition for the prognosis of colorectal cancer in a biological sample from a patient, wherein the reagents are specific for target cell genes comprising a nucleic acid sequence selected from the nucleic acid sequences set forth in SEQ ID NOs: 1, 2 to 4, 5 to 7, 8 to 12, 13, 14 to 17, 18 to 20, 21-22, 23-24, and 25 to 30, respectively; and

25

30

more particularly the use of reagents specific for a combination of 10 target cell genes and no more than 100 target genes in the manufacture of a composition for the prognosis of colorectal cancer in a biological sample from a

patient, wherein the reagents are specific for target cell genes comprising a nucleic acid sequence selected from the nucleic acid sequences set forth in SEQ ID NOs: 1, 2 to 4, 5 to 7, 8 to 12, 13, 14 to 17, 18 to 20, 21-22, 23-24, 25 to 30, 31 to 33, 34, 35, 36, 37, 38-39, 40 to 42, 43-44, 45, 46-49, 50-51, 52-53, 54, 55-56, 57, 58 and 59 respectively;

wherein the at least one specific reagent comprises at least one hybridization probe, at least one hybridization probe and at least one primer or at least one hybridization probe and two primers.

Figure

NK Cell Score in colonoscopy negative control (CNC) and colorectal cancer (CRC) patient blood samples, with distribution for CRC samples according to the cancer stage. Circles represent CNC; squares, up triangles, down triangle and lozenges represent CRC, Stage I, II, III and IV, respectively.

Exemples:

I) Materials and Methods

1. Patients and Sample Collection

The study was approved by the local Ethical Committee for Clinical Research. Written informed consent was obtained for all participants.

For the CRC group, 119 colorectal patients were consecutively recruited for the study, between July 2006 and March 2008 at the Department of Colorectal Surgery, Fudan University Cancer Hospital (FUCH), China. The tumors were staged according to the International Union Against Cancer (UICC) recommended tumor-node-metastasis (TNM) system. No patient received preoperative radiotherapy or chemotherapy. Patients suffering from hereditary colorectal cancer or inflammatory bowel

disease (Crohn's disease or ulcerative colitis) were excluded from this study. For each patient, 2.5 ml of peripheral blood were collected into PAXgene™ Blood RNA tubes (PreAnalytiX GmbH, Hombrechtikon, CH) at least one week after colonoscopy, before surgery, and processed according to manufacturer's guidelines. For the control group, 101 FOBT test-positive participants without carrying any symptom of polyps or colorectal cancer, which had been confirmed by colonoscopy, were enrolled from the Community Hospital in Shanghai area. The peripheral blood samples were collected into PAXgene tubes one week before colonoscopy examination. A detailed characterization of all participants included in this study is given in Table 1.

Table 1: Characteristics of the Patients

Category	Colorectal cancer (CRC) n=119	Colonoscopy negative controls (CNC) n=101
Age (y)		
Average	57.6	54.9
Max	82	71
Min	27	38
Gender		
Male	59	42
Female	60	59
Site		
Colon	60	-
Rectal	59	-
Cancer UICC Stage		
Stage I	17	-
Stage II	44	-
Stage III	26	-
Stage IV	32	-

2. RNA Extraction and microarray experiments

Total RNA was extracted with the PAXgene™ Blood RNA System (PreAnalytix) following manufacturer's instructions. The quantity of total RNA was measured by spectrophotometer at

optical density 260 nanometers and the quality was assessed using the RNA 6000 Nano LabChip® Kit on a BioAnalyzer Agilent 2100 (Agilent Technologies, Palo Alto, CA, U.S.A.). Only samples with RNA Integrity Number between 7 and 10 were
5 analyzed. 50 nanograms of total RNA was then reversely transcribed and linearly amplified to single strand cDNA using Ribo-SPIA™ technology with WT-Ovation™ RNA Amplification System (NuGEN Technologies Inc., San Carlos, CA, U.S.A.) according to the manufacturer's standard protocol, and the
10 products were purified with QIAquick™ PCR purification kit (QIAGEN GmbH, Hilden, Germany). 2 micro grams of amplified and purified cDNA were subsequently fragmented with RQ1 RNase-Free DNase (Promega Corp., Fitchburg, WI, U.S.A.) and labeled with biotinylated deoxynucleoside triphosphates by Terminal
15 Transferase (Roche Diagnostics Corp., Indianapolis, IN, U.S.A.) and GeneChip® DNA Labeling Reagent (Affymetrix Inc., Santa Clara, CA, U.S.A.). The labeled cDNA was hybridized onto HG U133 Plus 2.0 Array (Affymetrix) in a Hybridization Oven 640 (Agilent Technologies) at 60 rotations per minute, 50 °C for
20 18 hours. The HG U133 Plus 2.0 Array contains 54,675 probe sets representing approximately 39,000 best-characterized human genes. After hybridization, the arrays were washed and stained according to the Affymetrix protocol EukGE-WS2v4 using a GeneChip® Fluidics Station 450 (Affymetrix). The arrays were
25 scanned with the GeneChip® Scanner 3000 (Affymetrix).

3. Microarray Data Analysis

Quality control analyses were done according to the suggestions of standard Affymetrix quality control parameters.
30 Based on the evaluation criteria, all our experiments fulfilled the minimal quality requirements. The Affymetrix expression arrays were preprocessed by RMA (Robust Multi-chip Average) with background correction, quantile normalization and median polish summarization [1]. The probe sets with

extreme signal intensity (lower than 50 or higher than 2×10^{14}) were filtered out. To reduce the likelihood of batch effect, a normalization algorithm, Combat was applied to the filtered expression data¹¹. The ComBat method (<http://statistics.byu.edu/johnson/ComBat/>) applies either parametric or nonparametric empirical Bayes framework for adjusting batch effects in a given data set. Differential expressed genes (DEG) were identified by Significance Analysis of Microarrays (SAM) at False Discovery Rate (FDR) equals 0.05¹². The preprocessing and statistical steps were executed using R-environment with Bioconductor libraries^{13, 14}. Gene Ontology and Canonical Pathways analysis were conducted by using Ingenuity Pathway Analysis software version 8.5 (Ingenuity Systems, Redwood City, CA, U.S.A).

II) Results

1. Characteristics of the colorectal cancer and control patient populations

Clinical and demographic variables for the 119 colorectal cancer (CRC) patients and the 101 colonoscopy-negative controls (CNC) are summarized in Table 1. For the CRC, the diagnosis of colorectal cancer has been confirmed by the pathologist following the colonoscopy. The controls have been selected among FOBT positive patients enrolled in the Community Hospital, for whom the colonoscopy performed at Fudan University Cancer Hospital (FUCH) was finally negative. The age and the gender were well balanced between the CRC and the CNC groups.

2. Identification of genes whose expression in peripheral blood is different for colorectal cancer patients and colonoscopy-negative controls

The inventors looked for differentially expressed genes (DEG) between the 119 CRC and the 101 CNC, with the highest

differences between the two groups, considering the CRC group as a whole (Stage I, II, III and IV). After appropriate preprocessing, 20,169 probe sets were retained to perform DEG analysis. Using SAM, 327 DEGs were identified at FDR equal to 0.05, with fold change (FC) higher than 1.2.

Among these 327 DEG, 195 (59.6%) and 132 (40.36%) were found to be expressed at higher and lower levels in CRC samples respectively. The t-test p-values ranged from 1.43×10^{-25} to 1.51×10^{-01} , with 18 DEG having t-test p-values lower than 6.27×10^{-15} and all corresponding to well-annotated genes: MRPS6, SPRY4, NEAT1, CYBB, DUSP2, PDE4D, SH2D2A, G(1-2)NSR, ITGAM, VCAN, CD163, P2RY10, CD226, MRPL10, ITPRIPL2, CD2 and NUDT16 (Table 2). The highest fold change (FC) values were 1.83 (NEAT1 with higher level for CRC) and 1.71 (HBG2 with lower level for CRC), while 26 (8%) out of the 327 DEG have a FC value higher than 1.40.

As illustration, the results observed for SPRY4 (first ranked with higher expression level in CRC, t-test p-value 4.04×10^{-23} , FC 1.79) and MRPS6 (first ranked with lower expression level in CRC, t-test p-value 1.43×10^{-25} , FC 1.27). Such examples illustrate genes significantly differentially expressed between CRC and CNC patients. For SPRY4, rather homogenous hybridization signal values were observed for the 101 CNC, while the values for the CRC were more heterogeneous but with a mean value significantly (p -value 4.04×10^{-23}) increased compared to CNC (FC 1.78). For MRPS6, both populations presented a similar dispersion, with a significant (p -value 1.43×10^{-25}) mean decrease for CRC (FC 1.27).

Among the Top 18 DEG, four membrane leukocyte markers were observed, indicating different levels of expression in the peripheral blood of CRC patients compared to CNC: lower levels for CD2 and CD226 expressed by T cells and mainly NK cells

respectively; higher levels for CD163 and CD11B (ITGAM) expressed mainly expressed by monocytes and in many leukocytes involved in the innate immune system, respectively. Also interesting is the lower expression of granzyme B encoded by the GZMB gene in cytotoxic T lymphocytes and Natural Killer (NK) cells, in CRC samples. The other genes like INSR, SPRY4, DUSP2, PDE4D, and ITPRIPL2 are reported to be part of various signaling pathways, SH2D2A reported to be T-cell specific. VCAN has been reported to be expressed in monocytes, and its higher expression levels in CRC samples, together with CD163 and ITGAM, would be associated with some activation of circulating monocytes in the peripheral blood of these patients compared to CNC.

Analysis of the 327 DEG has been performed by using Ingenuity Pathway Analysis (IPA), which returned 321 mapped IDs suitable for interpretation of associated Bio Functions and Canonical Pathways. For Physiological System Development and Function, a high score was observed for Immune Cell Trafficking (p-value from 1.44×10^{-12} to 1.57×10^{-02} , with 50 molecules), covering activation, migration, accumulation, influx, chemotaxis, cell spreading, cell movement, chemoattraction, priming and adhesion of various immune cells. Interestingly for Canonical Pathways, Natural Killer Cell Signaling was the one with the lowest p-value (2.55×10^{-05}), with 10 genes: CD247, KLRB1, KLRC2, KLRC3, KLRD1, KLRK1, LCK, PRKCH, RRAS2 and SH2D1D. The implication of five membrane receptors specific to NK cells (KLRB1, KLRC2, KLRC3, KLRD1, KLRK1), very strongly suggests a particular NK cells component in the differences at the gene expression level in the peripheral blood of CRC patients. All NK cell genes are down-expressed in CRC. The results are summarized in the following tables 2 and 3.

Table 2: TOP 18 differentially expressed genes (DEGs) between colorectal cancer (CRC) and colonoscopy negative control (CNC) patient samples; Gene description, T-test p-value and fold change-related information

Affymetrix Probeset ID	SEQ ID NOs:	Gene Name	Gene Description *	T-test p- value	Fold Change	Direction (in CRC)
224919_at	31, 32, 33	MRPS6	Mitochondrial ribosomal protein S6	1.43×10^{-25}	1.27	Down
220983_s_at	34	SPRY4	Sprouty homolog 4	4.04×10^{-23}	1.79	Up
227062_at	35	NEAT1	Nuclear paraspeckle assembly transcript 1	7.62×10^{-22}	1.83	Up
203922_s_at	36	CYBB	Cytochrome b- 245, beta polypeptide	4.36×10^{-21}	1.31	Up
204794_at	37	DUSP2	Dual specificity phosphatase 2	1.44×10^{-20}	1.49	Down
204491_at	38, 39	PDE4D	Phosphodiesterase 4D, cAMP- specific	7.83×10^{-20}	1.49	Down
207351_s_at	40, 41, 42	SH2D2A	SH2 domain protein 2A	1.28×10^{-19}	1.47	Down
210164_at	14, 15, 16, 17	GZMB	Granzyme B	3.75×10^{-18}	1.62	Down
213792_s_at	43, 44	INSR	Insulin receptor	4.24×10^{-18}	1.35	Up
205785_at	45	ITGAM	Integrin alpha M	5.43×10^{-18}	1.32	Up
215646_s_at	46, 47, 48, 49	VCAN	Versican	6.03×10^{-18}	1.49	Up
203645_s_at	50, 51	CD163	CD163	3.78×10^{-17}	1.44	Up
1553856_s_at	52, 53	P2RY10	Purinergic receptor P2Y, G- protein coupled, 10	4.19×10^{-17}	1.26	Down
207315_at	54	CD226	CD226	1.14×10^{-16}	1.29	Down
224671_at	55, 56	MRPL10	Mitochondrial ribosomal protein L10	1.68×10^{-16}	1.21	Down
227954_at	57	ITPRIPL2	Inositol 1,4,5- triphosphate receptor interacting protein-like 2	8.71×10^{-16}	1.26	Up
205831_at	58	CD2	CD2	5.96×10^{-15}	1.28	Down
235002_at	59	NUDT16	Nudix (nucleoside diphosphate linked moiety X)-type motif 16	6.27×10^{-15}	1.21	Up

5

* Gene description from NetAffx™ and from Ingenuity Pathway Analysis® version 8.5

Table 3: NK cell score: Selected genes, T-test p-value and fold change-related information

Affymetrix Probeset ID	SEQ ID Nos:	Gene Name	Gene Description *	T-test p-value	Fold change CNC/CRC
214470_at	1	KLRB1	Killer cell lectin-like receptor subfamily B, member 1	2.67×10^{-04}	1.23
206785_s_at	2, 3, 4	KLRC2	Killer cell lectin-like receptor subfamily C, member 2	3.02×10^{-05}	1.40
207723_s_at	5, 6, 7	KLRC3 (NKG2E)	Killer cell lectin-like receptor subfamily C, member 3	4.42×10^{-05}	1.36
210606_x_at	8, 9, 10, 11, 12	KLRD1	Killer cell lectin-like receptor subfamily D, member 1	1.57×10^{-05}	1.23
205821_at	13	KLRK1 (NKG2D)	Killer cell lectin-like receptor subfamily K, member 1	5.22×10^{-06}	1.22
210164_at	14, 15, 16, 17	GZMB	Granzyme B	3.75×10^{-18}	1.62
210031_at	18, 19, 20	CD247 (CD3-zeta)	CD247 molecule	2.82×10^{-10}	1.27
212589_at	21, 22	RRAS2	Related RAS viral(r-ras) oncogene homolog 2	7.17×10^{-04}	1.20
1553176_at	23, 24	SH2D1B	SH2 domain containing 1B	2.11×10^{-10}	1.47
204891_s_at	25, 26, 27, 28, 29, 30	LCK	Lymphocyte-specific protein tyrosine kinase	1.93×10^{-11}	1.22

* Gene description from NetAffx™ and from Ingenuity Pathway Analysis® version 8.5

5 For these 10 NK cell-related genes, lower expression levels have been observed in the CRC group, suggesting either a decrease in the number of circulating NK cells, or an efflux of such cells towards other organ/tissue compartments and particularly the tumor sites. The lower expression levels
10 observed for GZMB is also remarkable, evocative of a major event occurring at the level of cellular cytotoxicity in CRC patients.

The top canonical pathways were related to T Cell Receptor
15 Signaling, Communication between Innate and Adaptive Immune Cells, and iCOS-iCOSL Signaling in T Helper Cells, with p-values equal to 9.08×10^{-05} , 2.85×10^{-04} and 5.78×10^{-04} respectively.

Interestingly, a low NK Cell Score under the first quarter, was observed for 51 out of the 119 CRC patients samples, and in only 4 out of the 101 CNC patients samples. Using such a straightforward cut-off, the performance of this discrimination can be expressed as 43% sensitivity and 96% specificity. Furthermore, when stratifying the CRC patients samples according to their tumor TNM staging (Stage I, II, III or IV), we observed that this NK Cell Score gradually decreased in CRC patients from Stage I to Stage IV (Figure 1). Statistically significant differences were mainly observed between CNC and CRC Stage II, III and IV, and between CRC Stage I and CRC Stage II-III and IV.

This study shows the potential of transcriptomics in peripheral blood, to discover biomarkers, and provide new insight on immune response in colorectal cancer. In addition to prepare possible alternative/complement to current screening modalities, these results also show that the expression analysis of genes like those related to NK cells should allow to stratify patients with colorectal cancer, opening the door to personalized medicine.

References:

1. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP. Exploration, normalization,
5 and summaries of high density oligonucleotide array probe level data. Biostatistics 2003;4:249-64)
2. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 2007;8:118-27.
- 10 3. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A 2001;98:5116-21.
4. Team RDC. R: A Language and Environment for Statistical Computing. Vienna, Austria, 2009.

CLAIMS

1. A method for determining the prognosis of a colorectal cancer in a peripheral blood sample from a patient, the method comprising:

a) obtaining the peripheral blood sample and extracting total RNA from the blood sample,

b) contacting the total RNA with at least one reagent that is specific for at least one NK cell gene and no more than 25 specific reagents for 25 NK cell genes,

c) determining the expression level of the at least one NK cell gene and of the most 25 NK cell genes to obtain an expression profile for the patient,

d) performing analysis of the expression profile of the patient with expression profiles of NK cell genes from patients previously clinically classified as a good prognosis and expression profiles of NK cell genes from patients previously classified as a poor prognosis, wherein

- if the expression profile for the patient is clustered with the expression profiles from patients previously clinically classified as a poor prognosis, then the patient is determined to have a poor prognosis, and

- if the expression profile for the patient is clustered with the expression profiles from patients previously clinically classified as a good prognosis, then the patient is determined to have a good prognosis.

2. The method as claimed in claim 1, wherein in step b) the total RNA is brought into contact with at least one reagent is specific for at least one NK cell gene and no more than 25 specific reagents for 25 NK cell genes, said NK cell gene comprising the nucleic acid sequences set forth in SEQ ID NOs: 1 to 12, wherein the at least one reagent is specific for at least one NK cell gene selected from the group consisting of:

(i) KLRB1 gene comprising a full length sequence such as identified in SEQ ID NO: 1,

(ii) KLRC2 gene comprising a full length sequence such as identified in SEQ ID NOs: 2, 3 or 4,

5 (iii) KLRC3 gene comprising a full length sequence such as identified in SEQ ID NOs: 5, 6 or 7,

(iv) KLRD1 gene comprising a full length sequence such as identified in SEQ ID NOs: 8, 9, 10, 11, or 12, and

10 (v) KLRK1 gene comprising a full length sequence such as identified in SEQ ID NO: 13, and

the expression level of the at least one NK cell gene is determined in step c) to obtain the expression profile for the patient.

15 3. The method as claimed in claim 2, wherein in step b) the total RNA is brought into contact with reagents specific for a combination of least 5 NK cell genes and no more than 25 NK cell genes, wherein the reagents include at least reagents specific for the NK cell genes consisting of:

20 (i) KLRB1 gene comprising a full length sequence such as identified in SEQ ID NO: 1,

(ii) KLRC2 gene comprising a full length sequence such as identified in SEQ ID NOs: 2, 3 or 4,

25 (iii) KLRC3 gene comprising a full length sequence such as identified in SEQ ID NOs: 5, 6 or 7,

(iv) KLRD1 gene comprising a full length sequence such as identified in SEQ ID NOs: 8, 9, 10, 11, or 12, and

(v) KLRK1 comprising a full length sequence such as identified in SEQ ID NO: 13,

30 the expression level of at least said 4 NK cell genes is determined in step c) to obtain the expression profile for the patient.

4. The method as claimed in claim 3, wherein in step b) the total RNA is further brought into contact with at least one

reagent specific for at least one target cell gene and no more than 5 specific reagents for 5 target cell genes, said target cell gene comprising the nucleic acid sequences set forth in SEQ ID NOs: 12 to 24, wherein the at least one reagent is
5 specific for at least one target cell gene selected from the group consisting of:

(i) GZMB gene comprising a full length sequence such as identified in SEQ ID NO: 14, 15, 16 or 17,

10 (ii) CD247 gene comprising a full length sequence such as identified in SEQ ID NO: 18, 19 or 20,

(iii) RRAS2 gene comprising a full length sequence such as identified in SEQ ID NO: 21 or 22, and

(iv) SH2D1B gene comprising a full length sequence such as identified in SEQ ID NO: 23 or 24, and

15 (v) LCK gene comprising a full length sequence such as identified in SEQ ID NO: 25, 26, 27, 28, 29 or 30, and

the expression level of the at least one cell gene is determined in step c) to obtain the expression profile for the patient.

20

5. The method as claimed in claim 4, wherein in step b) the total RNA is brought into contact with reagents specific for a combination 5 target cell genes, wherein the reagents are specific for the target cell genes consisting of:

25 (i) GZMB gene comprising a full length sequence such as identified in SEQ ID NO: 14, 15, 16 or 17,

(ii) CD247 gene comprising a full length sequence such as identified in SEQ ID NO: 18, 19 or 20,

30 (iii) RRAS2 gene comprising a full length sequence such as identified in SEQ ID NO: 21 or 22, and

(iv) SH2D1B gene comprising a full length sequence such as identified in SEQ ID NO: 23 or 24, and

(v) LCK gene comprising a full length sequence such as identified in SEQ ID NO: 25, 26, 27, 28, 29 or 30, and

the expression level of the at least 5 cell gene is determined in step c) to obtain the expression profile for the patient.

6. The method as claimed in claim 5 wherein in step b) the total RNA is further brought into contact with at least one reagent specific for at least one target cell gene and no more than 100 specific reagents for 100 target cell genes, said target cell gene comprising the nucleic acid sequences set forth in SEQ ID NOs: 25 to 59, wherein the at least one reagent is specific for at least one target cell gene selected from the group consisting of:
- (i) MRPS6 gene comprising a full length sequence such as identified in SEQ ID NO: 31, 32, or 33,
 - (ii) SPRY4 gene comprising a full length sequence such as identified in SEQ ID NO: 34,
 - (iii) NEAT1 gene comprising a full length sequence such as identified in SEQ ID NO: 35,
 - (iv) CYBB gene comprising a full length sequence such as identified in SEQ ID NO: 36,
 - (v) DUSP2 gene comprising a full length sequence such as identified in SEQ ID NO: 37, .
 - (vi) PDEAD gene comprising a full length sequence such as identified in SEQ ID NO: 38 or 39,
 - (vii) SH2D2A gene comprising a full length sequence such as identified in SEQ ID NO: 40, 41 or 42,
 - (viii) INSR gene comprising a full length sequence such as identified in SEQ ID NO: 43 or 44,
 - (ix) ITGAM gene comprising a full length sequence such as identified in SEQ ID NO: 45,
 - (x) VCAN gene comprising a full length sequence such as identified in SEQ ID NO: 46, 47, 48 or 49,
 - (xi) CD 163 gene comprising a full length sequence such as identified in SEQ ID NO: 50, or 51.

(xii) P2RY10 gene comprising a full length sequence such as identified in SEQ ID NO: 52 or 53,

(xii) CD226 gene comprising a full length sequence such as identified in SEQ ID NO: 54,

5 (xiii) MRPL10 gene comprising a full length sequence such as identified in SEQ ID NO: 55 or 56,

(xiv) ITPRIPL2 gene comprising a full length sequence such as identified in SEQ ID NO: 57,

10 (xv) CD2 gene comprising a full length sequence such as identified in SEQ ID NO: 58, and

(xvi) NUDT16 gene comprising a full length sequence such as identified in SEQ ID NO: 59, and.

the expression level of the at least one cell gene is determined in step c) to obtain the expression profile for the
15 patient.

7. The method as claimed in claim 6, wherein in step b) the total RNA is brought into contact with reagents specific for a combination of least 17 target cell genes and no more than 100
20 target cell genes, wherein the reagents include at least reagents specific for the target cell genes consisting of:

(i) MRPS6 gene comprising a full length sequence such as identified in SEQ ID NO: 31, 32, or 33,

25 (ii) SPRY4 gene comprising a full length sequence such as identified in SEQ ID NO: 34,

(iii) NEAT1 gene comprising a full length sequence such as identified in SEQ ID NO: 35,

(iv) CYBB gene comprising a full length sequence such as identified in SEQ ID NO: 36,

30 (v) DUSP2 gene comprising a full length sequence such as identified in SEQ ID NO: 37, .

(vi) PDEAD gene comprising a full length sequence such as identified in SEQ ID NO: 38 or 39,

(vii) SH2D2A gene comprising a full length sequence such as identified in SEQ ID NO: 40, 41 or 42,

(viii) INSR gene comprising a full length sequence such as identified in SEQ ID NO: 43 or 44,

5 (ix) ITGAM gene comprising a full length sequence such as identified in SEQ ID NO: 45,

(x) VCAN gene comprising a full length sequence such as identified in SEQ ID NO: 46, 47, 48 or 49,

10 (xi) CD 163 gene comprising a full length sequence such as identified in SEQ ID NO: 50, or 51.

(xii) P2RY10 gene comprising a full length sequence such as identified in SEQ ID NO: 52 or 53,

(xiii) CD226 gene comprising a full length sequence such as identified in SEQ ID NO: 54,

15 (xiv) MRPL10 gene comprising a full length sequence such as identified in SEQ ID NO: 55 or 56,

(xv) ITPRIPL2 gene comprising a full length sequence such as identified in SEQ ID NO: 57,

20 (xvi) CD2 gene comprising a full length sequence such as identified in SEQ ID NO: 58, and

(xvii) NUDT16 gene comprising a full length sequence such as identified in SEQ ID NO: 59, and

the expression level of the at least 17 cell genes is determined in step c) to obtain the expression profile for the
25 patient.

8. The method as claimed in anyone of the preceding claims wherein the at least one specific reagent of step b) comprises at least one hybridization probe.

30

9. The method as claimed in claim 8, wherein the specific reagents of step b) comprises at least one hybridization probe and at least one primer.

10. The method as claimed in claim 7, wherein the specific reagents of step b) comprises at least one hybridization probe and two primers

5 11. A kit for the prognosis of a colorectal cancer in a peripheral blood sample from a patient comprising at least one specific reagent for at least one NK cell gene and no more than 25 specific reagents for 25 NK cell genes comprising at least the nucleic acid sequences set forth in SEQ NOs 1 to 13,
10 wherein the at least one reagent is specific for at least one NK cell gene selected from the group consisting of:

- (i) KLRB1 gene comprising a full length sequence such as identified in SEQ ID NO: 1,
- (ii) KLRC2 gene comprising a full length sequence such as
15 identified in SEQ ID NOs: 2, 3 or 4,
- (iii) KLRC3 gene comprising a full length sequence such as identified in SEQ ID NOs: 5, 6 or 7,
- (iv) KLRD1 gene comprising a full length sequence such as identified in SEQ ID NOs: 8, 9, 10, 11, or 12, and
20 (v) KLRK1 gene comprising a full length sequence such as identified in SEQ ID NO: 13.

12. The kit as claimed in claim 11 comprising reagents that are specific for the NK cell genes consisting of:

- 25 (i) KLRB1 gene comprising a full length sequence such as identified in SEQ ID NO: 1,
- (ii) KLRC2 gene comprising a full length sequence such as identified in SEQ ID NOs: 2, 3 or 4,
- (iii) KLRC3 gene comprising a full length sequence such as
30 identified in SEQ ID NOs: 5, 6 or 7,
- (iv) KLRD1 gene comprising a full length sequence such as identified in SEQ ID NOs: 8, 9, 10, 11, or 12, and
- (v) KLRK1 gene comprising a full length sequence such as identified in SEQ ID NO: 13.

13. The kit as claimed in claim 12 comprising at least one reagent that is specific for at least one target cell gene and no more than 5 target cell genes, said at least one target cell gene being selected from the group consisting of:

(i) GZMB gene comprising a full length sequence such as identified in SEQ ID NO: 14, 15, 16 or 17,

(ii) CD247 gene comprising a full length sequence such as identified in SEQ ID NO: 18, 19 or 20,

(iii) RRAS2 gene comprising a full length sequence such as identified in SEQ ID NO: 21 or 22, and

(iv) SH2D1B gene comprising a full length sequence such as identified in SEQ ID NO: 23 or 24, and

(v) LCK gene comprising a full length sequence such as identified in SEQ ID NO: 25, 26, 27, 28, 29 or 30.

14. The kit as claimed in claim 13 comprising 5 reagents that are specific for the target cell genes consisting of:

(i) GZMB gene comprising a full length sequence such as identified in SEQ ID NO: 14, 15, 16 or 17,

(ii) CD247 gene comprising a full length sequence such as identified in SEQ ID NO: 18, 19 or 20,

(iii) RRAS2 gene comprising a full length sequence such as identified in SEQ ID NO: 21 or 22, and

(iv) SH2D1B gene comprising a full length sequence such as identified in SEQ ID NO: 23 or 24, and

(v) LCK gene comprising a full length sequence such as identified in SEQ ID NO: 25, 26, 27, 28, 29 or 30.

15. The kit as claimed in claim 14 comprising at least one reagent that is specific for at least one target cell gene and at the most 100 reagents that are specific for 100 target cell genes, said at least target cell gene being selected from the group consisting of:

- (i) MRPS6 gene comprising a full length sequence such as identified in SEQ ID NO: 31, 32, or 33,
- (ii) SPRY4 gene comprising a full length sequence such as identified in SEQ ID NO: 34,
- 5 (iii) NEAT1 gene comprising a full length sequence such as identified in SEQ ID NO: 35,
- (iv) CYBB gene comprising a full length sequence such as identified in SEQ ID NO: 36,
- (v) DUSP2 gene comprising a full length sequence such as
10 identified in SEQ ID NO: 37,.
- (vi) PDEAD gene comprising a full length sequence such as identified in SEQ ID NO: 38 or 39,
- (vii) SH2D2A gene comprising a full length sequence such as identified in SEQ ID NO: 40, 41 or 42,
- 15 (viii) INSR gene comprising a full length sequence such as identified in SEQ ID NO: 43 or 44,
- (ix) ITGAM gene comprising a full length sequence such as identified in SEQ ID NO: 45,
- (x) VCAN gene comprising a full length sequence such as
20 identified in SEQ ID NO: 46, 47, 48 or 49,
- (xi) CD 163 gene comprising a full length sequence such as identified in SEQ ID NO: 50, or 51.
- (xii) P2RY10 gene comprising a full length sequence such as identified in SEQ ID NO: 52 or 53,
- 25 (xiii) CD226 gene comprising a full length sequence such as identified in SEQ ID NO: 54,
- (xiv) MRPL10 gene comprising a full length sequence such as identified in SEQ ID NO: 55 or 56,
- (xv) ITPRIPL2 gene comprising a full length sequence such as
30 identified in SEQ ID NO: 57,
- (xvi) CD2 gene comprising a full length sequence such as identified in SEQ ID NO: 58, and
- (xvi) NUDT16 gene comprising a full length sequence such as identified in SEQ ID NO: 59.

16. The kit as claimed in claim 15, comprising 17 reagents that are specific for 17 target cell genes consisting of:

- (i) MRPS6 gene comprising a full length sequence such as
5 identified in SEQ ID NO: 31, 32, or 33,
- (ii) SPRY4 gene comprising a full length sequence such as
identified in SEQ ID NO: 34,
- (iii) NEAT1 gene comprising a full length sequence such as
identified in SEQ ID NO: 35,
- 10 (iv) CYBB gene comprising a full length sequence such as
identified in SEQ ID NO: 36,
- (v) DUSP2 gene comprising a full length sequence such as
identified in SEQ ID NO: 37, .
- (vi) PDEAD gene comprising a full length sequence such as
15 identified in SEQ ID NO: 38 or 39,
- (vii) SH2D2A gene comprising a full length sequence such as
identified in SEQ ID NO: 40, 41 or 42,
- (viii) INSR gene comprising a full length sequence such as
identified in SEQ ID NO: 43 or 44,
- 20 (ix) ITGAM gene comprising a full length sequence such as
identified in SEQ ID NO: 45,
- (x) VCAN gene comprising a full length sequence such as
identified in SEQ ID NO: 46, 47, 48 or 49,
- (xi) CD 163 gene comprising a full length sequence such as
25 identified in SEQ ID NO: 50, or 51.
- (xii) P2RY10 gene comprising a full length sequence such as
identified in SEQ ID NO: 52 or 53,
- (xii) CD226 gene comprising a full length sequence such as
identified in SEQ ID NO: 54,
- 30 (xiii) MRPL10 gene comprising a full length sequence such as
identified in SEQ ID NO: 55 or 56,
- (xiv) ITPRIPL2 gene comprising a full length sequence such as
identified in SEQ ID NO: 57,

(xv) CD2 gene comprising a full length sequence such as identified in SEQ ID NO: 58, and

(xvi) NUDT16 gene comprising a full length sequence such as identified in SEQ ID NO: 59.

5

17. Use of at least one specific reagent for at least one NK cell genes and no more than 25 specific reagents for 25 NK cell genes comprising the nucleic acid sequences set forth in SEQ ID NOs 1 to 12 in the manufacture of a composition for the prognosis of colorectal cancer in a biological sample from a patient, wherein the at least one reagent is specific for at least one NK cell gene comprising a nucleic acid sequence selected from the nucleic acid sequences set forth in anyone of SEQ ID NOs: 1 to 12.

15

18. Use of reagents specific for a combination of at least 5 NK cell genes and no more than 25 NK cell genes in the manufacture of a composition for the prognosis of colorectal cancer in a biological sample from a patient, wherein the reagents are specific for at least 5 NK cell genes comprising a nucleic acid sequence selected from the nucleic acid sequences set forth in SEQ ID NOs: 1, 2 to 4, 5 to 7 and 8 to 12, respectively.

19. Use of reagents specific for a combination of 10 target cell genes in the manufacture of a composition for the prognosis of colorectal cancer in a biological sample from a patient, wherein the reagents are specific for target cell genes comprising a nucleic acid sequence selected from the nucleic acid sequences set forth in SEQ ID NOs: 1, 2 to 4, 5 to 7, 8 to 12, 13, 14 to 17, 18 to 20, 21-22, 23-24, and 25 to 30, respectively.

20. Use of reagents specific for a combination of 10 target cell genes and no more than 100 target genes in the manufacture of a composition for the prognosis of colorectal cancer in a biological sample from a patient, wherein the reagents are specific for target cell genes comprising a nucleic acid sequence selected from the nucleic acid sequences set forth in SEQ ID NOs: 1, 2 to 4, 5 to 7, 8 to 12, 13, 14 to 17, 18 to 20, 21-22, 23-24, 25 to 30, 31 to 33, 34, 35, 36, 37, 38-39, 40 to 42, 43-44, 45, 46-49, 50-51, 52-53, 54, 55-56, 57, 58 and 59 respectively.

21. The use as claimed in anyone of claims 17-20, wherein the at least one specific reagent comprises at least one hybridization probe.

22. The use as claimed in anyone of claims 17-20, wherein the at least one specific reagent comprises at least one hybridization probe and at least one primer.

23. The use as claimed in anyone of claims 17-20, wherein the at least one specific reagent comprises at least one hybridization probe and two primers.

1/1

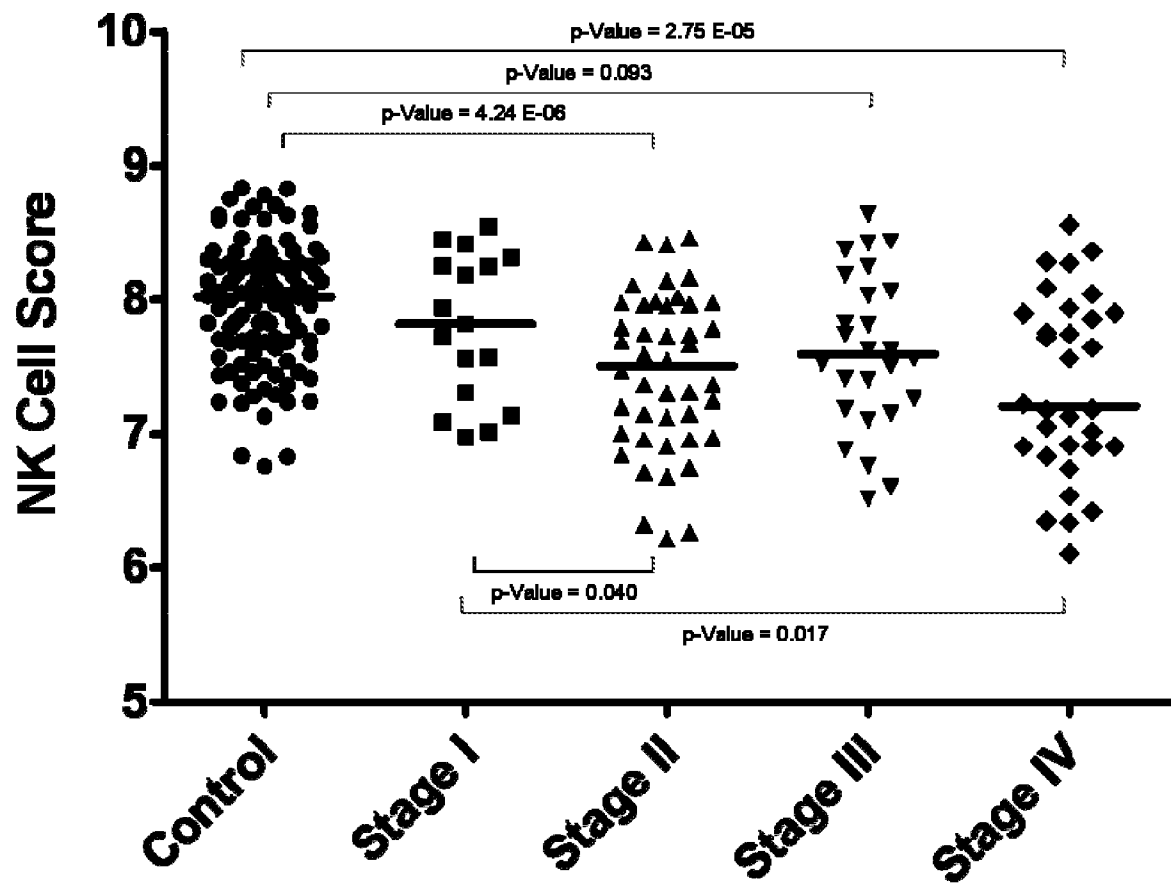


Figure 1

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2010/073659

A. CLASSIFICATION OF SUBJECT MATTER

C12Q 1/68 (2006.01) i

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)

IPC: C12Q1

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)

WPI; EPODOC; CNPAT; CNKI; Elsevier SD; CA: 肠, 癌, 肿瘤, 天然杀伤细胞, 自然杀伤细胞, 表达, 预后, colorectal cancer, CR, CRC, Colorectal carcinoma, NK cell, natural killer cell(s), CD56, expression, prognosis

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN101068936 A(WYETH CORP), 7 NOV. 2007(7.11.2007), SEE DESCRIPTION P.42 PARA.7 - P.43 PARA.2 AND P.39 PARA.3	1-23
Y	WO2009126804 A(UNIV COLORADO), 15 OCT. 2009(15.10.2009), SEE CLAIMS 1,13, DESCRIPTION P.3 PARA.3 - P.4 PARA.1	1-23
A	CN101111604 A(AGENCY SCIENCE TECH & RES), 23 JAN. 2008(23.1.2008), SEE DESCRIPTION P.18 PARA.2 - P.21 PARA.1	1-23

☐ Further documents are listed in the continuation of Box C.☒ See patent family annex.

* Special categories of cited documents:	"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
"A" document defining the general state of the art which is not considered to be of particular relevance	"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
"E" earlier application or patent but published on or after the international filing date	"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
"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)	"&" document member of the same patent family
"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	

Date of the actual completion of the international search 24 Feb. 2011(24.2.2011)	Date of mailing of the international search report 17 Mar. 2011 (17.03.2011)
Name and mailing address of the ISA/CN The State Intellectual Property Office, the P.R.China 6 Xitucheng Rd., Jimen Bridge, Haidian District, Beijing, China 100088 Facsimile No. 86-10-62019451	Authorized officer ZHAO, Yanhao Telephone No. (86-10)62411043

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2010/073659

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN101068936 A	07-11-2007	WO2006060265	08-06-2006
		US2006134671	22-06-2006
		EP1815024	08-08-2007
		AU2005312081	08-06-2006
		NO20072296	20-08-2007
		INDELNP200703794E	24-08-2007
		KR20070084488	24-08-2007
		JP2008520251T	19-06-2008
		MXPA07005764	01-08-2007
		BRPI0518036	28-10-2008
CN101111604 A	23-01-2008	US2006074565	06-04-2006
		WO2006052218	18-05-2006
		EP1809762	25-07-2007
		INDELNP200703233E	31-08-2007
		JP2008521383T	26-06-2008
WO2009126804A	15-10-2009	NONE	