



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

(21) International Application Number:
PCT/CN2011/072155

(22) International Filing Date:
25 March 2011 (25.03.2011)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant (for all designated States except US): **BIO-MERIEUX** [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 Alley, 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**; 10th Floor, Block A, Investment Plaza, 27 Jinrongdajie, Xicheng District, 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, RS, 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 DETERMINING IN VITRO PROBABILITY FOR INDIVIDUAL TO SUFFER FROM COLORECTAL CANCER

(57) Abstract: The present invention discloses a method for determining in vitro, in a peripheral blood sample, the probability for an individual to suffer from a colorectal cancer, using the comparison of the amount of expression products of nucleic acids of genes of the individual to be tested with the amount of expression products of nucleic acids of the same genes obtained from a CRC group of patients constituting the positive control and with the same genes obtained from a CNC group of individuals constituting the negative control; and a kit comprising specific binding partners for the expression products.



METHOD AND KIT FOR DETERMINING IN VITRO PROBABILITY FOR INDIVIDUAL TO SUFFER FROM
COLORECTAL CANCER

5 Field of the invention

The present invention relates to the detection of a colorectal cancer, especially to a method and kit for determining the probability to suffer from such a cancer.

10 Background

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 is the key to successful treatment and patient survival and represents a major public health challenge. 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 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 peripheral blood biomarkers aimed at early detection of CRC became a focus since several years, especially for its 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 particles, such as cells by suspending them in a stream of fluid and analyzing them by using an electronic detection apparatus.

The present inventors have found that differentially expressed genes represented 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 this step of purification of RNA.

Accordingly, the present invention relates to a method for determining *in vitro*, in a peripheral blood sample, the probability for an individual to suffer from a colorectal cancer, the method comprising the steps of:

a) determining, in the peripheral blood sample, the amount of at least one expression product from at least one nucleic acid sequence and no more than 7 nucleic acid sequences, said nucleic acid sequence being selected from the sequences identified in SEQ ID NOs: 1 to 11,

b) comparing the amount of said expression product determined in step a) with a reference amount of the expression product for a group of individuals previously diagnosed as colorectal cancer patients and with a reference amount of the expression product for a group of individuals previously verified as non colorectal cancer individuals,

c) performing analysis of results of step b), wherein

- if the result for the tested individual is close to or equal to the result obtained from the group of individuals previously diagnosed as colorectal cancer patients, then the tested individual is classified as a colorectal cancer patient, and
- if the result for the tested individual is close to or equal to the result obtained from the group of individuals previously verified as non colorectal cancer individuals, then the tested individual is classified as a non colorectal individual.

The amount of the expression product is directly linked to the expression level of a gene defined by its nucleic acid sequence.

The expression level of at least one of the above nucleic acids is a sufficient information for determining if the individual is a CRC patient or not. But, in a preferred embodiment of the invention, in the step a), it is determined the amount of the expression products from nucleic acid sequences selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 2 or SEQ ID NO: 3 or SEQ ID NO: 4, SEQ ID NO: 5 or SEQ ID NO: 6, SEQ ID NO: 7 or SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10 and SEQ ID NO: 11.

The amount of expression product(s) from the nucleic acid(s) is determined by bringing the expression product(s) into contact with at least one binding partner specific for each expression product.

Expression product(s) means RNA transcript(s) or polypeptides(s). Accordingly, in the method of the invention, it is determined the amount of at least one RNA transcript or at least one polypeptide.

The term RNA transcripts is intended to mean total RNA, i.e., coding or non coding RNA directly obtained from the peripheral blood sample or indirectly obtained from the blood sample after cell lysis. Especially, 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.

By way of indication, when the RNA is intracellular RNA, it can be extracted from the cells present in the blood sample by a step of lysis of, in order to release the nucleic acids contained in the cells of the individual to be tested. By way of example, use may be made of the methods of lysis as described in patent 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 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.

In the method of the invention the RNA transcript can be detected and quantified by hybridization, amplification or sequencing. Especially, to be detected and quantified, the RNA transcript is brought into contact with at least one probe or at least one primer under predetermined conditions which enable hybridization of said probe and/or said primer to the RNA transcript. But in another embodiment of the invention, DNA copies of the RNA transcript are prepared and said DNA copies are determined by bringing them into contact with at least one probe or at least one primer under predetermined conditions which enable hybridization of said probe and/or said primer to the DNA copies.

More precisely, in the methods described above RNA transcript or DNA copies are brought into contact with at least one hybridization probe and at least one primer and more particularly at least one hybridization probe and two primers.

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)), 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.

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, it is used at least two amplification primers, specific for a target gene, that allow the amplification 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 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, 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 plasmon resonance, or contact angle variation, or by physical methods such as atomic force spectroscopy, tunnel effect, etc.; radioactive molecules such as ³²P, ³⁵S or ¹²⁵I.

For the purpose of the present invention, the 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 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 excitation at the

appropriate wavelength. The detection probe in particular may be a "reporter probe" comprising a "color-coded barcode" according to NanoString™'s technology.

5 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
10 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
15 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.

20 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
25 (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
30 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,
35 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 cRNAs or cDNAs specific for a nucleic acid of 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.

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. 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).

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 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 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 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.^{sup.2}. The use of nylon membranes, referred to as "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.^{sup.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.

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.

By way of amplification, it is possible, to determine the expression level of the target gene as follows: 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.

By way of hybridization, the expression of a target gene can be determined as follows: 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-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 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 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 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 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 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 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 label is detected directly.

The expression of a target gene can also 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 the cDNAs of the mRNAs of the biological material. The polymerization of the complementary RNA of the cDNA is subsequently carried out 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 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.

When the expression product is a polypeptide it can be detected by bringing it in contact with at least one specific ligand, such as defined below. In a preferred embodiment the expressed polypeptide is brought into contact with at least two specific ligands, such as

defined below. Specific ligand means for example an antibody or an affinity protein named "NanofitinTM".

Nanofitins are affinity proteins with competitive features. They present a competitive affinity, similar to antibodies.

5 The term "antibody or antibodies" embraces polyclonal antibodies, monoclonal antibodies, humanized antibodies, recombinant antibodies. Their production methods are well known by the person skilled in the art.

10 The present invention also includes a kit for determining *in vitro* the probability for an individual to suffer from a colorectal cancer comprising at least one binding partner specific for at least one nucleic acid sequence and no more than 7 binding partners specific for 7 expression products of 7 nucleic acid
15 sequences, wherein the at least one binding partner is specific for at least one expression product of at least one nucleic acid sequence selected from the group consisting of sequences set forth in SEQ ID NOs: 1 to 11.

 Especially, the kit comprises a combination of 7 binding
20 partners which are specific for the expression products of 7 nucleic acid sequences having the sequences set forth in SEQ ID NOs: SEQ ID NO: 1, SEQ ID NO: 2 or SEQ ID NO: 3 or SEQ ID NO: 4, SEQ ID NO: 5 or SEQ ID NO: 6, SEQ ID NO: 7 or SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10 and SEQ ID NO: 11.

25 In the kit the specific binding partner comprises:

- at least one hybridization probe,
- or at least one hybridization probe and at least one primer, or
- at least one hybridization probe and two primers, or
- at least one specific ligand or at least two specific ligands,
30 such as antibody and/or affinity protein.

 Finally, the invention concerns the use of at least one specific binding partner for at least one expression product of at least one nucleic acid sequence and no more than 7 specific binding
35 partners for 7 expression products of 7 nucleic acid sequences, said at least one nucleic acid sequence having a sequence selected from the group consisting of nucleic acid sequences set forth in SEQ ID

NOS 1 to 11, in the manufacture of a composition for determining *in vitro* the probability for an individual to suffer from a colorectal cancer.

Especially, the use of a combination of 7 specific binding partners which are specific for 7 expression products of 7 nucleic acid sequences having the sequences set forth in SEQ ID NOS: SEQ ID NO: 1, SEQ ID NO: 2 or SEQ ID NO: 3 or SEQ ID NO: 4, SEQ ID NO: 5 or SEQ ID NO: 6, SEQ ID NO: 7 or SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10 and SEQ ID NO: 11.

Specific binding partner comprises:

- at least one hybridization probe,
- or at least one hybridization probe and at least one primer, or
- at least one hybridization probe and two primers, or

at least one specific ligand or at least two specific ligands, such as antibody and/or affinity protein.

Example:

I) Materials and Methods

1. Patients and Sample Collection

Peripheral blood samples from 161 colorectal patients (CRC) and 148 colonoscopy negative control patients (CNCs) were collected, between 2006 and 2010. The CRC patients were recruited at the Department of Colorectal Surgery, FDUSCC, 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. The CNCs, without any symptom of polyps or colorectal cancer, which had been confirmed by colonoscopy, were enrolled from the Community Hospital in Shanghai area and FDUSCC. For each patient, 2.5 ml of peripheral blood were collected into PAXgene™ Blood RNA tubes (PreAnalytiX GmbH, Hombrechtikon, CH) and processed according to manufacturer's guidelines.

The study involves two separate cohorts of participants. Cohort 1 consists of 100 CRC patients and 100 CNCs. For CRC patients, blood samples were collected in FDUSCC at least one week after colonoscopy, before surgery. For CNCs, blood samples were collected in a
5 Community Hospital in the Shanghai area one week before the colonoscopy. The gene expression profiles from these samples were analyzed as a train set to search for significant genes associated with CRC and identify molecular signature. Cohort 2 includes 61 CRC patients and 48 CNCs. Samples were collected in the same way as
10 cohort 1. Cohort 2 was used as an independent test set to verify the signature performance that observed in the cohort 1.

2. RNA Extraction and microarray experiments

Total RNA was extracted with the PAXgene™ Blood RNA System
15 (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
20 between 7 and 10 were 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 products were
25 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 Transferase (Roche Diagnostics Corp.,
30 Indianapolis, IN, U.S.A.) and GeneChip® DNA Labeling Reagent (Affymetrix Inc., Santa Clara, CA, U.S.A.). The labeled cDNA was hybridized onto the GeneChip HG U133 Plus 2.0 Array (Affymetrix) in a Hybridization Oven 640 (Agilent Technologies) at 60 rotations per minute, 50 °C for 18 hours. The HG U133 Plus 2.0 Array contains
35 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 scanned with the GeneChip® Scanner 3000 (Affymetrix).

3. Statistical Analysis

5 Microarray data quality control was performed according to the suggestions of standard Affymetrix quality control parameters. The Affymetrix expression arrays were preprocessed globally by Robust Multi-chip Average method (RMA) with background correction, quantile normalization and median polish summarization (Irizarry RA et al.,
10 Biostatistics 20 3;4:249-64).

For cohort 1 data, the probesets with extreme signal intensity (lower than $\log_2(50)$ or higher than $2E14$) were filtered out. Then, biological knowledge based filtering were performed using the
15 information of Entrez Gene Database (Maglot D et al., Nucleic Acids Research 2007;35:D26-31). Probesets without Entrez Gene ID annotation were removed. For multiple probesets mapping to the same Entrez Gene ID, only the probeset with the largest value of Inter Quantile Range were retained and the others were removed. After two-
20 steps filtering, 9,859 probesets were kept for the downstream analysis. To reduce the likelihood of batch effect, Combat method was applied to the filtered expression data (Johnson WE et al., Biostatistics 2007;8:118-27). Differentially Expressed Gene (DEG) analysis was performed by the Significance Analysis of Microarrays
25 (SAM) method (False Discovery Rate = 0.05; Type= "Two class unpaired"; test statistic = "t-statistic"; number of permutations = 1,000) (Tusher VG et al., PNAS USA 2001,98:5116-21). Significant gene selection and predictive model construction were performed using a 5-fold cross validation process with RFE-SVM method. Among
30 the 200 samples in train set, 160 were randomly selected to form a learning set; the predictive models were created with the different sizes ranging from 1 to 100 genes scored by RFE-SVM; and the model performance was assessed using the rest of 40 samples. This process was repeated 1,000 times. Our result suggested that a maximum 97%
35 accuracy was achievable with the 100-gene based SVM predictive models. The signature size optimization took into account the prediction performance, signature complexity and economy. Finally,

we identified seven core genes with overall 90% accuracy to meet our target performance. The seven genes were selected by t-test *P* value, fold change, biological function and not related to age or gender factors.

5

II) Results

1. Characteristics of the colorectal cancer and control patient populations

10 Among 309 participants in the two cohorts, there were 161 CRC patients and 148 CNCs. The demographic and clinical characteristics of the patients are summarized in the table 1.

Table 1: Clinical characteristics of the Patients

15

Variable	Train set		Test set	
	CRC (n=100)	Control (n=100)	CRC (n=61)	Control (n=48)
Age - yr				
Mean	57.6	56.5	55.4	55.2
Range	27-78	38-74	34-82	38-70
Sex - no. (%)				
Male	50 (50.0%)	50 (50.0%)	32 (52.5%)	3 (6.3%)
Female	50 (50.0%)	50 (50.0%)	29 (47.5%)	45 (93.7%)
Tumor site - no. (%)				
Colon	41 (41.0%)	-	33 (54.1%)	-
Rectum	59 (59.0%)	-	28 (45.9%)	-
Stage I	16 (16.%)	-	8 (13.1)%	-
Stage II	36 (36.0%)	-	19 (31.1%)	-
Stage III	24 (24.0%)	-	17 (27.9%)	-
Stage IV	24 (24.0%)	-	17 (27.9%)	-

2. 7-gene CRC biomarker panel: Identification and validation

Train set: the inventors performed significant gene selection and prediction model construction based on 5-fold cross validation process. The process was run for 1,000 iterations. Within each iteration, they recorded the unique top-7 gene set and its corresponding prediction model performance accessed by internal test fold. Eventually, the overall performance was estimated by taking the average performance of 1,000 prediction models in the internal test fold. The results show that an overall accuracy performance of 90.0% is achievable with prediction models. The inventors have selected the best 7-gene prediction model, for which 90.0% accuracy, 89.0% sensitivity and 91.0% specificity for the train set.

Test set: the inventors have then verified the performance of the signature of the above prediction model identified in the train set in an independent cohort (test set) including 109 samples, 61 CRCs and 48 CNCs. The overall performance of this signature are 83.0% (CI%:73.9, 88.9) accuracy, 84.0% (CI%:71.5, 91.4) sensitivity, and 81.0% (CI%:66.9, 86.6) specificity.

3. Analysis of discriminative capacities of individual genes from the signature observed from the train set.

The table 2 below summarizes the individual performance of said 7 genes. For each gene are given the individual characteristics like Probeset_id (Affymetrix probeset identification), T_test P value observed between 100 CNCs and 100 CRCs, and Fold Change observed between 100 CNCs and 100 CRCs.

Table 2

Probeset_id*	Gene Symbol**	SEQ ID NOS:	Mean signal***	t-test P value	Fold Change	Direction (in CRC)
227062_at	NEAT1	1	621	3.84×10^{-11}	1.46	up
223204_at	FAM198B	2, 3, 4	97	3.56×10^{-12}	1.52	up
205785_at	ITGAM	5, 6	95	1.35×10^{-17}	1.32	up
213906_at	MYBL1	7, 8	139	3.51×10^{-8}	1.38	down
209339_at	SIAH2	9	252	8.06×10^{-6}	1.25	up
1553589_a_at	PDZK1IP1	10	407	9.32×10^{-5}	1.37	down
1553991_s_at	VSIG10	11	65	1.47×10^{-14}	1.41	up

Up: means that the mean signal for the CRC group is higher than in the CNC group

Down means that the mean signal for the CRC group is lower than for CNC group

5 * means Probeset_id according to Affymetrix annotation version in 2010
(https://www.affymetrix.com/analysis/netaffx/xmlquery.affx?netaffx=netaffx4__annot&requested=403680)

** means the identified gene and its variants or related sequences to said gene or variants

10 *** means average signals observed for 100 CRCs and 100 CNCs array experiments

CLAIMS

1. A method for determining *in vitro*, in a peripheral blood sample, the probability for an individual to suffer from a colorectal cancer, the method comprising the steps of:

a) determining, in the peripheral blood sample, the amount of at least one expression product from at least one nucleic acid sequence and no more than 7 nucleic acid sequences, said nucleic acid sequence being selected from the sequences identified in SEQ ID NOs:

1 to 11,

b) comparing the amount of said expression product determined in step a) with a reference amount of the expression product for a group of individuals previously diagnosed as colorectal cancer patients and with a reference amount of the expression product for a group of individuals previously verified as non colorectal cancer individuals,

c) performing analysis of results of step b), wherein

- if the result for the tested individual is close to or equal to the result obtained from the group of individuals previously diagnosed as colorectal cancer patients, then the tested individual is classified as a colorectal cancer patient, and
- if the result for the tested individual is close to or equal to the result obtained from the group of individuals previously verified as non colorectal cancer individuals, then the tested individual is classified as a non colorectal individual.

2. The method as claimed in claim 1, wherein in step a), it is determined the amount of at least one expression product from the at least one nucleic acid by bringing said expression product into contact with at least one binding partner specific for said expression product.

3. The method as claimed in claim 1 or 2, wherein in step a), it is determined the amount of the expression products from nucleic acid sequences selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 2 or SEQ ID NO: 3 or SEQ ID NO: 4, SEQ ID NO: 5 or SEQ ID NO: 6,

SEQ ID NO: 7 or SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10 and SEQ ID NO: 11.

4. The method as claimed in claim 1, wherein the expression product
5 is at least one RNA transcript or at least one polypeptide.

5. The method as claimed in claim 4, wherein the expression product is at least one mRNA.

10 6. The method as claimed in claim 5, wherein the RNA transcript is detected and quantified by hybridization, by amplification or by sequencing.

7. The method as claimed in claim 4, wherein the RNA transcript is
15 brought into contact with at least one probe and/or at least one primer under predetermined conditions which enable hybridization of said probe and/or said primer to the RNA transcript.

8. The method as claimed in claim 6, wherein DNA copies of the RNA
20 transcript are prepared and the DNA copies are brought into contact with at least one probe and/or at least one primer under predetermined conditions which enable hybridization of said probe and/or said primer to the DNA copies.

25 9. The method as claimed in claim 4, wherein the expressed polypeptide is detected by bringing into contact with at least one specific ligand, in particular an antibody or an affinity protein.

10. The method as claimed in claim 9, wherein the expressed
30 polypeptide is brought into contact with at least two specific ligands, in particular two antibodies or two affinity protein or one antibody and one affinity protein.

11. A kit for determining *in vitro* the probability for an individual
35 to suffer from a colorectal cancer comprising at least one binding partner specific for at least one expression product of at least one nucleic acid sequence and no more than 7 binding partners specific

for 7 expression products of 7 nucleic acid sequences, wherein the at least one binding partner is specific for the expression product of at least one nucleic acid sequence selected from the group consisting of nucleic acid sequences set forth in SEQ ID NOs: 1 to 11.

12. The kit as claimed in claim 11, which comprises a combination of 7 binding partners which are specific for the expression products of 7 nucleic acid sequences having the sequences set forth in SEQ ID NOs: SEQ ID NO: 1, SEQ ID NO: 2 or SEQ ID NO: 3 or SEQ ID NO: 4, SEQ ID NO: 5 or SEQ ID NO: 6, SEQ ID NO: 7 or SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10 and SEQ ID NO: 11.

13. The kit as claimed in claim 11, wherein the at least one specific binding partner comprises at least one hybridization probe.

14. The kit as claimed in claim 13, wherein the specific binding partners comprise at least one hybridization probe and at least one primer.

15. The kit as claimed in claim 13, wherein the specific binding partners comprise at least one hybridization probe and two primers.

16. The kit as claimed in claim 11, wherein the at least one specific binding partner comprises at least one specific ligand, in particular one antibody or one affinity protein.

17. The kit as claimed in claim 16, wherein the specific binding partners comprises at least two specific ligands, in particular two antibodies or two affinity proteins or one antibody and one affinity protein.

18. Use of at least one specific binding partner for at least one expression product of at least one nucleic acid sequence and no more than 7 specific binding partners for 7 expression products of 7 nucleic acid sequences, said at least one nucleic acid sequence having a sequence selected from the group consisting of nucleic acid

sequences set forth in SEQ ID NOs 1 to 11, in the manufacture of a composition for determining *in vitro* the probability for an individual to suffer from a colorectal cancer.

5 19. The use as claimed 18, of a combination of 7 specific binding partners which are specific for the expression products of 7 nucleic acid sequences having the sequences set forth in SEQ ID NOs: SEQ ID NO: 1, SEQ ID NO: 2 or SEQ ID NO: 3 or SEQ ID NO: 4, SEQ ID NO: 5 or
10 10 and SEQ ID NO: 11.

20. The use, as claimed in claim 18, wherein the at least one specific binding partner comprises at least one hybridization probe.

15 21. The use, as claimed in claim 20, wherein the specific binding partners comprise at least one hybridization probe and at least one primer.

20 22. The use as claimed in claim 21, wherein the specific bonding partners comprise at least one hybridization probe and two primers.

23. The use as claimed in claim 18, wherein the at least one specific binding partner comprises at least one specific ligand, in particular an antibody or an affinity protein.

25 24. The use as claimed in claim 23, wherein the specific binding partners comprise at least two specific ligands, in particular two antibodies or two affinity proteins or one antibody and one affinity protein.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2011/072155

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:C12Q

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

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNABS, SIPOABS, DWPI, CNTXT, WOTXT, EPTXT, USTXT, CNKI, GOOGLE SCHOLAR, Elsevier Science, PubMed: colorectal cancer, CRC, colon cancer, large bowel cancer, FAM198B, AD021, AD036, C4orf18

GenBank; EMBL: sequences search on SEO ID NOs: 1-11

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	EP2177615 A1 (FRAUNHOFER GES FOERDERUNG ANGEWANDTEN EV) 21 Apr. 2010 (21.04.2010), Claims 1-7 and Table 34	1-24
Y	GENBANK, ACCESSION NO. NM_001031700, 2000	1-24
Y	GENBANK, ACCESSION NO. NM_016613, 2000	1-24
Y	GENBANK, ACCESSION NO. NM_001128424, 2000	1-24

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

* Special categories of cited documents:

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

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

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

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

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

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

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

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

“&” document member of the same patent family

Date of the actual completion of the international search

22 Dec. 2011(22.12.2011)

Date of mailing of the international search report

02 Feb. 2012 (02.02.2012)

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/CN2011/072155

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
EP2177615 A1	21.04.2010	NONE	