



(51) International Patent Classification:

C12Q 1/6886 (2018.01)

(21) International Application Number:

PCT/EP2018/078340

(22) International Filing Date:

17 October 2018 (17.10.2018)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

17197156.7

18 October 2017 (18.10.2017) EP

(71) Applicants:

**DEUTSCHES KREBS-
FORSCHUNGSZENTRUM** [DE/DE]; Im Neuenheimer
Feld 280, 69120 Heidelberg (DE). **UNIVERSITÄT
HEIDELBERG** [DE/DE]; Grabengasse 1, 69117 Heidel-
berg (DE).

(72) Inventors: **BURWINKEL, Barbara**; Montpellierstraße
13, 69115 Heidelberg (DE). **MAIERTHALER, Melanie**;
Weinbrennerstr. 5, 68753 Waghäusel (DE). **EDELMANN,
Dominic**; Bachgasse 55, 69469 Weinheim (DE). **BREN-
NER, Hermann**; Panoramastr. 27, 69257 Wiesenbach
(DE).

(74) Agent: **HERZOG FIESSER & PARTNER PATEN-
TANWÄLTE PARTG MBB**; Dick, Alexander, Duden-
straße 46, 68167 Mannheim (DE).

(81) Designated States (unless otherwise indicated, for every
kind of national protection available): AE, AG, AL, AM,
AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ,
CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO,
DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN,
HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP,
KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME,
MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ,
OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA,

(54) Title: MEANS AND METHODS FOR COLORECTAL CANCER CLASSIFICATION

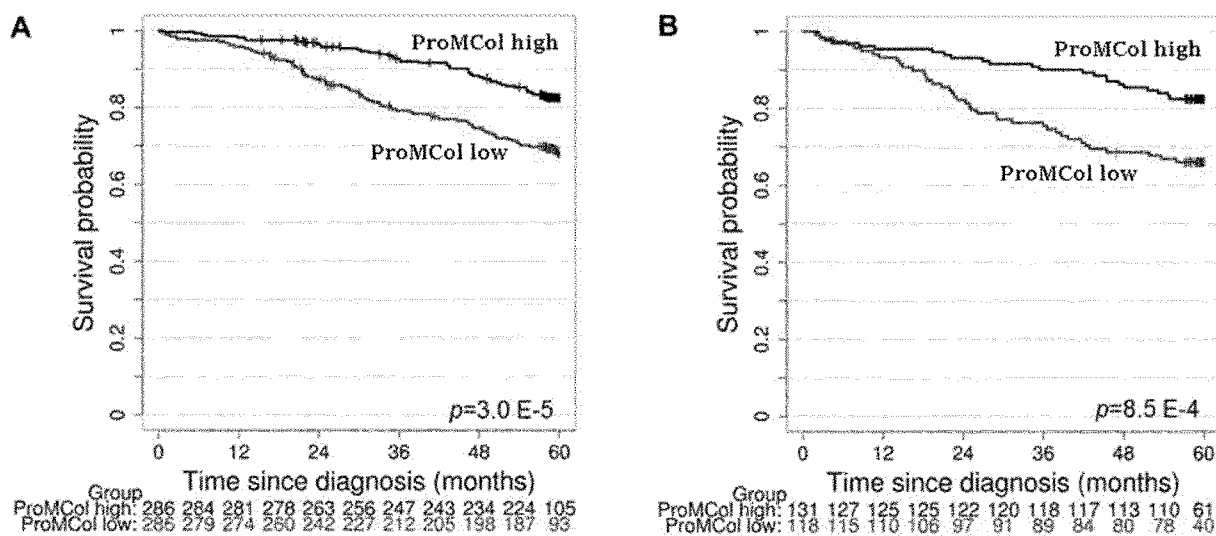


Fig. 4

(57) Abstract: The present invention relates to a method for determining a survival probability of a subject suffering from colorectal cancer comprising a) detecting the methylation status of at least two CpG sites related to at least two CpG sites selected from cg16336556, cg14270346, cg05646575, cg17431888, cg12510999, cg00832644, cg11056055, cg08804626, cg14983135, cg22522598, cg19184885, cg08729279, cg10758824, cg18195165, cg08617020, cg23750514, cg01131395, cg18736676, cg19340296, and cg16399624 in a sample of said subject and, b) based on the methylation status detected in step a), determining the survival probability of said subject; and to a method for patient monitoring comprising the aforementioned steps and providing close monitoring and/or lifestyle recommendations in case an unfavorable survival probability and/or an increased mortality risk is detected. The present invention further relates to a data collection comprising the positions of at least two of the CpG sites of the invention and kits, devices, and uses related thereto.

SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN,
TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

Published:

- *with international search report (Art. 21(3))*
- *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

Means and methods for colorectal cancer classification

5 The present invention relates to a method for determining a survival probability of a subject suffering from colorectal cancer comprising a) detecting the methylation status of at least two CpG sites related to at least two CpG sites selected from cg16336556, cg14270346, cg05646575, cg17431888, cg12510999, cg00832644, cg11056055, cg08804626, cg14983135, cg22522598, cg19184885, cg08729279, cg10758824, cg18195165, 10 cg08617020, cg23750514, cg01131395, cg18736676, cg19340296, and cg16399624 in a sample of said subject and, b) based on the methylation status detected in step a), determining the survival probability of said subject; and to a method for patient monitoring comprising the aforementioned steps and providing close monitoring and/or lifestyle recommendations in case an unfavorable survival probability and/or an increased mortality risk is detected. The 15 present invention further relates to a data collection comprising the positions of at least two of the CpG sites of the invention and kits, devices, and uses related thereto.

Colorectal cancer (CRC) is the third most common cancer worldwide accounting for 1.36 20 million new cases annually (Ferlay et al., International journal of cancer Journal international du cancer 2015;136(5):E359-86). The five-year survival rate of patients is highly dependent on the stage of the disease (Siegel et al.; CA Cancer J Clin 2012;62(4):220-41). At present, the most accurate means for the prediction of patient survival remains pathological staging according to the tumor-node-metastasis (TNM-) system but it has been recognized that even 25 patients within the same tumor stage display a strong heterogeneity for prognosis and treatment response (Nagtegaal et al.; Nat Rev Clin Oncol 2011;9(2):119-23; Boland & Goel; N Engl J Med 2016;374(3):277-8). Especially for stage II patients there is an ongoing debate if adjuvant chemotherapy should be recommended or not (Boland & Goel, loc. cit.; Benson et al.; J Clin Oncol 2004;22(16):3408-19; O'Connor et al., J Clin Oncol 2011;29(25):3381-8). 30 The current classification system provides only limited information for the clinical prognostication highlighting the need for additional prognostic markers to avoid a potential under- or over-treatment of patients.

In the research field of diagnostic, prognostic and predictive biomarkers for cancer, DNA methylation has gained increasing prominence (Esteller; *Lancet Oncol* 2003;4(6):351-8; Okugawa et al.; *Gastroenterology* 2015;149(5):1204-25 e12; Tang et al.; *Clinical epigenetics* 2016;8:115; Tang et al.; *Oncotarget* 2016;7(39):64191-202). DNA methylation, predominantly defined as an addition of a methyl-group at cytosine residues located adjacent to guanine bases (CpG dinucleotides), is one of the major epigenetic mechanisms, important in many physiological and pathophysiological processes (Egger et al.; *Nature* 2004;429(6990):457-63). Since many years, the dysregulation of DNA methylation has been known to have a key role in cancer development and progression (Feinberg & Tycko; *Nat Rev Cancer* 2004;4(2):143-53). In recent years, methods have become available which allow testing thousands of CpG dinucleotides in the human genome for their methylation status; e.g., Illumina® Kits make screening of more than 450 thousand CpG sites possible; this compares to an estimated number of more than 28 million CpG sites in the human genome (Luo et al. (2014), *BioMed Research International*, Vol 2014, Art. ID 784706). It was also established that CpG sites in the vicinity of an aberrantly methylated CpG site tend to also be affected by aberrant methylation (Sofer et al. (2013), *Bioinformatics* 29(22): 2884; Yang et al. (2015), *Int J Cancer* 136: 1845; Tang et al. (2016), *Oncotarget* 7(39): 64191).

Meanwhile, the accessibility of genome-wide methylation profiling revealed innovative prognostic methylation markers in different cancers (Sandoval et al.; *J Clin Oncol* 2013;31(32):4140-7; Bjaanaes et al.; *Mol Oncol* 2016;10(2):330-43; Stefansson et al.; *Mol Oncol* 2015;9(3):555-68; Cicek et al.; *Hum Mol Genet* 2013;22(15):3038-47; Wei et al.; *Nat Commun* 2015;6:8699) but most of the studies on CRC prognosis lack an adequate sample size. So far, among all methylation-based prognostic biomarker candidates for CRC, the CpG island methylator phenotype (CIMP) status has been the most promising indicator for prognostication (Okugawa et al., loc. cit), however, to date there is no consensus definition for CIMP leading to inconsistent study results (Issa; *Nat Rev Cancer* 2004;4(12):988-93; Jia et al.; *Clinical epigenetics* 2016;8:25). A better understanding of this disease and detection of novel markers for the prediction of patients' prognosis may enable individual treatment options in order to improve patients' outcomes and quality of life.

There is, thus, a need in the art for improved methods for prognosticating in colorectal cancer patients, and in particular for improved genetic and/or epigenetic markers for such determination. This problem is solved by the methods and means of the present invention

with the features of the independent claims. Preferred embodiments, which may be realized in an isolated fashion or in arbitrary combination, are listed in the dependent claims.

Accordingly, the present invention relates to a method for determining a survival probability of a subject suffering from colorectal cancer comprising

- a) detecting the methylation status of at least two CpG sites related to at least two CpG sites of Table 1 in a sample of said subject and,
- b) based on the methylation status detected in step a), determining the survival probability of said subject.

As used in the following, the terms “have”, “comprise” or “include” or any arbitrary grammatical variations thereof are used in a non-exclusive way. Thus, these terms may both refer to a situation in which, besides the feature introduced by these terms, no further features are present in the entity described in this context and to a situation in which one or more further features are present. As an example, the expressions “A has B”, “A comprises B” and “A includes B” may both refer to a situation in which, besides B, no other element is present in A (i.e. a situation in which A solely and exclusively consists of B) and to a situation in which, besides B, one or more further elements are present in entity A, such as element C, elements C and D or even further elements. It will be understood that any component defined herein as being included may preferably be explicitly excluded from the claimed invention by way of proviso or negative limitation.

Further, as used in the following, the terms "preferably", "more preferably", "most preferably", "particularly", "more particularly", "specifically", "more specifically" or similar terms are used in conjunction with optional features, without restricting further possibilities. Thus, features introduced by these terms are optional features and are not intended to restrict the scope of the claims in any way. The invention may, as the skilled person will recognize, be performed by using alternative features. Similarly, features introduced by "in an embodiment of the invention" or similar expressions are intended to be optional features, without any restriction regarding further embodiments of the invention, without any restrictions regarding the scope of the invention and without any restriction regarding the possibility of combining the features introduced in such way with other optional or non-optional features of the invention.

Moreover, if not otherwise indicated, the term "about" relates to the indicated value with the commonly accepted technical precision in the relevant field, preferably relates to the indicated value $\pm 20\%$, more preferably $\pm 10\%$, most preferably $\pm 5\%$. Further, the term "essentially" indicates that deviations having influence on the indicated result or use are absent, i.e. potential deviations do not cause the indicated result to deviate by more than $\pm 20\%$, more preferably $\pm 10\%$, most preferably $\pm 5\%$. Thus, "consisting essentially of" means including the components specified but excluding other components except for materials present as impurities, unavoidable materials present as a result of processes used to provide the components, and components added for a purpose other than achieving the technical effect of the invention. For example, a composition defined using the phrase "consisting essentially of" encompasses any known acceptable additive, excipient, diluent, carrier, and the like. Preferably, a composition consisting essentially of a set of components will comprise less than 5% by weight, more preferably less than 3% by weight, even more preferably less than 1%, most preferably less than 0.1% by weight of non-specified component(s). In the context of nucleic acid sequences, the term "essentially identical" indicates a %identity value of at least 80%, preferably at least 90%, more preferably at least 98%, most preferably at least 99%. As will be understood, the term essentially identical includes 100% identity. The aforesaid applies to the term "essentially complementary" mutatis mutandis. Unless otherwise noted, nucleic acid sequences and amino acid sequences are noted according to conventional notation. Thus, nucleic acid sequences noted in the direction 5' to 3', and amino acid sequences in the direction N-terminal to C-terminal.

The method for determining a survival probability of the present invention, preferably, is an in vitro method. Moreover, it may comprise steps in addition to those explicitly mentioned above. For example, further steps may relate, e.g., to obtaining a sample for step a), deriving recommendations for further proceeding and/or providing treatment after obtaining the result of step b), and/or further steps as specified herein below. The method, preferably, comprises further including at least one, more preferably at least three, more preferably at least five, most preferably all of the standard clinical factors gender, age, tumor stage, tumor location, smoking behavior and MSI status in the determination, preferably as specified herein in the Examples. Moreover, one or more of said steps may be performed by automated equipment.

As used herein, the term "survival probability" relates to the probability that a subject will be alive during certain period of time. As will be understood by the skilled person, a survival

probability is also a measure for a mortality risk, i.e. for the probability that said subject dies within the indicated period of time. Preferably, said period of time is at most 6 years, more preferably at most 5 years, more preferably at most 50 months, even more preferably at most 40 months, most preferably at most 30 months. Thus, preferably, the survival probability may be a favorable survival probability, i.e. a survival probability indicating a low probability for dying within one of the aforesaid time frames. Preferably, the survival probability for the aforesaid time frames in case a favorable survival probability is determined is at least 0.75, more preferably at least 0.8, still more preferably at least 0.9, most preferably at least 0.95. Also preferably, the survival probability may be an unfavorable survival probability, i.e. a survival probability indicating a decreased probability for surviving one of the aforesaid time frames. Preferably, the survival probability for the aforesaid time frames in case an unfavorable survival probability is determined is at most 0.74, more preferably at most 0.7, even more preferably at most 0.7, most preferably at most 0.65. In accordance with the above, "determining a survival probability" of a subject, as used herein, relates to determining the probability according to which the subject will survive the aforesaid time frame, which also is a measure for the probability to die within one of the aforesaid time frames. Preferably, survival probability (ps) and mortality risk (pm) are correlated according to equation $ps = 1 - pm$; preferably, said mortality risk is a probability to die from a specific disease, most preferably from colorectal cancer. Preferably, the aforesaid time frames are calculated from the time, preferably the day, the sample is obtained from the subject.

As specified elsewhere herein, the subject, preferably, is a subject known to suffer from cancer, preferably colorectal cancer. Thus, the method of the present invention, preferably, does not provide diagnosis that a subject is, at the time of assessment, afflicted with disease, in particular colorectal cancer. Thus, in an embodiment, determining a survival probability is not diagnosing a specific disease, more preferably is not diagnosing disease. Thus, preferably, the method for determining a survival probability is not required to be performed by a medical practitioner, more preferably is not performed by a medical practitioner. Preferably, the result of the method of the present invention is not a diagnosis of disease. As will be understood by the skilled person, in case an unfavorable survival probability is determined according to the method of the present invention, the subject and/or the counseling medical practitioner may decide to or recommend to perform life-style changes in order to improve its survival probability; also, treatment methods, in particular aggressive treatment methods, may be recommended, e.g. surgery, high-dose chemotherapy and/or high-dose radiotherapy. Thus,

preferably, in case an unfavorable survival probability is determined, the subject is recommended to be treated with at least one of surgery, chemotherapy, and radiotherapy; also preferably, in case a favorable survival probability is determined, the subject is recommended to be treated by avoiding chemotherapy and/or radiotherapy. Thus, preferably, in case a favorable survival probability is determined, said subject is treated by surgery only, by surgery and radiotherapy, or by surgery and chemotherapy. In an embodiment, detecting an unfavorable survival probability, preferably, provides an indication that a subject has an increased probability to, preferably within the time frames as specified above, experience severe aggravation of disease, preferably at least of one of the diseases as specified herein. In a further embodiment, detecting an unfavorable survival probability, preferably, provides an indication that a subject has an increased probability to, preferably within the time frames as specified above, die from disease, preferably at least one of the diseases as specified herein.

The term "subject", as used herein, relates to an animal, preferably a mammal, and, more preferably, a human. Preferably, the subject according to the present invention is a subject of at least 40 years of age, more preferably at least 50 years of age, even more preferably at least 60 years of age, most preferably at least 65 years of age. Preferably, the subject has been diagnosed with cancer, more preferably colorectal cancer, even more preferably with non-metastatic colorectal cancer. In contrast, the term "apparently healthy subject" relates to a subject not known to suffer from colorectal cancer, preferably not suspected to suffer from colorectal cancer based on physical examination, more preferably not showing any symptom of disease, even more preferably not known to suffer from cancer and not showing any symptoms of cancer based on physical examination. As will be understood by the skilled person, the subject and the apparently healthy subject preferably are corresponding subjects from the same species, preferably from the same race.

The term "sample", as used herein, refers to a sample of a body fluid, to a sample of separated cells or to a sample from tissue of the subject; preferably, the term refers to a tumor cell-comprising sample of a body fluid, to a sample of separated tumor cells (e.g. circulating/dissimilated tumor cells in peripheral blood) or to a sample from tumor tissue of the subject. Samples of body fluids can be obtained by well known techniques and include, preferably, samples of blood, plasma, serum, or urine. Tissue or organ samples may be obtained from any tissue or organ by, e.g., biopsy. Separated cells may be obtained from the body fluids or the tissues or organs by separating techniques such as centrifugation or cell

sorting. Preferably, cell-, tissue- or organ samples are obtained from tumor tissues. Preferably, the sample is a sample comprising tumor cells, more preferably a tumor sample, preferably a formalin-fixed tumor sample, more preferably a formalin-fixed, paraffin embedded tumor sample. Preferably, the sample is a sample comprising colorectal cancer cells, more preferably a tumor sample of a colorectal cancer, preferably a formalin-fixed tumor sample of a colorectal cancer, more preferably a formalin-fixed, paraffin embedded tumor sample of a colorectal cancer.

The term "colorectal cancer" is, in principle, known to the skilled person as relating to a cancer originating in the colon (colon cancer) or in the rectum (rectal cancer). As will be understood by the skilled person, metastases of other cancers having the primary tumor in other parts of the body are not colorectal cancer, even if said metastases are situated in the colon or rectum. Preferably, the colorectal cancer is an adenocarcinoma, a carcinoid tumor, a gastrointestinal stromal tumor, a lymphoma, or a sarcoma. More preferably, the colorectal cancer is an adenocarcinoma. As used herein, colorectal cancer may be of any of cancer stages I to IV.

The terms "CpG" and "CpG site" are known to the skilled person. Preferably, the terms relate to a site in DNA, preferably chromosomal DNA of a subject, having the nucleotide sequence 5'-CG-3'. As is also known to the skilled person, CpG sites can be methylated by DNA methyltransferases at the cytosine residue to yield a 5-methylcytosine residue, and methylation at a specific CpG site may be inherited or may be a de novo methylation acquired during life time of the subject. The CpG sites as referred to herein are those of Table 1. The CpG site locations indicated in Table 1 refer to the positions in the human reference genome GRCh37 as provided by the Genome Reference Consortium (www.ncbi.nlm.nih.gov/grc) on 2009/02/27. This assembly is also referred to as hg19.

The term "CpG site related to" a CpG site X, with X being a specific CpG site of Table 1, as used herein, relates to a CpG site in the vicinity of the specific CpG site. Preferably, a CpG site related to a specific CpG site is a CpG site at most 5 kb downstream or upstream, more preferably at most 2 kb downstream or upstream, even more preferably at most 1 kb downstream or upstream, still more preferably 0.5 kb downstream or upstream of the specific CpG site. As will be understood, most preferably, the CpG site related to a specific CpG site is the specific CpG site of Table 1 itself. Thus, preferably, a CpG site related to a CpG site of

Table 1 is a CpG site of Table 1 as specified herein above, i.e. is a CpG site specifically indicated in Table 1. In accordance, the present invention preferably relates to a method for determining a survival probability of a subject suffering from colorectal cancer comprising

- a) detecting the methylation status of at least two CpG sites selected from the CpG sites of Table 1 in a sample of said subject and,
- b) based on the methylation status detected in step a), determining the survival probability of said subject.

The term "at least two CpG sites related to at least two CpG sites of Table 1", as used herein, relates to at least to CpG sites related to two non-identical CpG sites of Table 1. As will be understood, in case more than two CpG sites are analyzed, also more than one CpG site related to the same CpG site of Table 1 may preferably be analyzed; e.g. in case three CpG sites are analyzed, one CpG site related to cg16336556 and two CpG sites related to cg14270346 might be analyzed. More preferably, however, the number of CpG sites analyzed is identical to the number of non-identical CpG sites of Table 1 these CpG sites analyzed are related to.

Table 1: CpG sites of the invention; positions on human chromosome and nucleotide number of the CpG sites refer to the human genome sequence assembly GRCh37/hg19.

	CpG ID¹	Chr.	Position²	PCA weight
1	cg16336556	2	33295138	0.31
2	cg14270346	9	38026076	0.31
3	cg05646575	12	132221835	0.30
4	cg17431888	1	183203863	0.29
5	cg12510999	2	28022223	0.26
6	cg00832644	1	162343785	0.24
7	cg11056055	7	116356738	0.24
8	cg08804626	6	1953086	0.23
9	cg14983135	7	48129822	0.22
10	cg22522598	19	15360953	0.22
11	cg19184885	17	45699322	0.22
12	cg08729279	2	216708378	0.21
13	cg10758824	1	164606935	0.21

14	cg18195165	7	5549570	0.19
15	cg08617020	16	87894491	0.19
16	cg23750514	11	2790418	0.18
17	cg01131395	13	47207812	0.15
18	cg18736676	6	109053447	0.15
19	cg19340296	6	52288804	0.12
20	cg16399624	10	134553188	0.07

¹Unique CpG locus identifier from the 450K methylation array

²All coordinates referring to RefSeq genes (hg19) and NCBI build 37

Chr: chromosome, UTR: untranslated region, PCA: Principal component analysis

5 Preferably, the CpG sites analyzed according to the method of the present invention comprise CpG sites related to CpG sites selected from the list consisting of cg16336556, cg14270346, cg05646575, cg17431888, cg12510999, cg00832644, cg11056055, cg08804626, cg14983135, cg22522598, cg19184885, cg08729279, and cg10758824; preferably consisting of cg16336556, cg14270346, cg05646575, cg17431888, cg12510999, cg00832644, and
10 cg11056055.

Preferably, the CpG sites analyzed according to the method of the present invention comprise CpG sites selected from the list consisting of cg16336556, cg14270346, cg05646575, cg17431888, cg12510999, cg00832644, cg11056055, cg08804626, cg14983135,
15 cg22522598, cg19184885, cg08729279, and cg10758824; preferably consisting of cg16336556, cg14270346, cg05646575, cg17431888, cg12510999, cg00832644, and cg11056055.

As used herein, the term "methylation status" relates to a state of a specific CpG site in a cell
20 being methylated or not, more preferably relates to the extent to which a specific CpG site is methylated in a population of cells, or not. As is understood by the skilled person, in a diploid cell, there are four occurrences of a specific CpG site, i.e. two alleles, with each allele comprising the two strands of DNA making up double-stranded DNA; thus, the methylation status of a single CpG site may be all four CpGs non-methylated; one CpG methylated; two,
25 three, or four CpGs methylated. Preferably, in the method of the present invention, the methylation of only one strand of a given DNA is analyzed, e.g. by hybridizing a primer upstream of said CpG site as specified herein below. Also, in a population of cells, in

particular in a mixed population of cells, the methylation status of a CpG site is not necessarily identical for all cells of said population. Thus, preferably, the methylation status is detected as the number of cells comprising a specific CpG site at least one, preferably at least twice, in methylated form in a given number of cells; or the methylation status is detected as the number of methylated forms of a specific CpG site detected in a given number of cells. Preferably, the methylation status of at least 10, more preferably at least 25, most preferably at least 100 cells is detected in such case. More preferably, the methylation status is detected as a relative methylation status, e.g. in comparison to a population of a corresponding cell population obtained from one or more apparently healthy subjects. Most preferably, the methylation status is detected as a ratio of the number of individual CpG sites at a given position found to be methylated to the total number of individual CpG sites at said given position analyzed, which is known as the methylation beta value (i.e., preferably, $\beta = \text{methylated signal} / (\text{unmethylated signal} + \text{methylated signal})$), or as a figure derivable therefrom by standard mathematical operations. In accordance with the above, beta-values range from 0 to 1, with 0 representing completely unmethylated and 1 represents completely methylated. A known parameter derivable from the beta-value is the methylation M value (methylated/unmethylated) which can be calculated from $\beta = 2M/(2M+1)$; thus $M = \log_2[\beta / (1 - \beta)]$. Thus, the methylation status of a CpG site in a population of cells, preferably, is the average degree of methylation of said CpG site in a population of at least 10, preferably at least 25, more preferably at least 100 cells. In an embodiment, the methylation status may also be expressed as a ratio of the number of individual CpG sites at a given position found to be unmethylated to the total number of individual CpG sites at said given position analyzed, i.e. as a non-methylation status. More preferably, the methylation status is expressed as a ratio of the number of individual CpG sites at a given position found to be methylated to the total number of individual CpG sites at said given position analyzed.

Methods for determining the methylation status of a CpG site are known in the art. Preferably, the method comprises isolating genomic DNA from said sample, preferably from cells comprised in said sample. Preferably, the method comprises contacting said DNA with a methylation-sensitive restriction enzyme having a nucleic acid sequence comprising the sequence 5'-CG-3' as a recognition sequence; preferably, the method further comprises contacting a further aliquot of said DNA with a corresponding non-methylation-sensitive restriction enzyme having the same nucleic acid sequence comprising the sequence 5'-CG-3' as a recognition sequence. More preferably, the method comprises treating said DNA, before

or after isolation, with a bisulfite, preferably sodium bisulfite. Preferably, the method further comprises annealing an oligonucleotide specifically annealing to a sequence immediately upstream of said CpG site and comprising a 3'-terminal sequence 5'-CG-3' and/or an oligonucleotide specifically annealing to a sequence immediately upstream of said CpG site and comprising a 3'-terminal sequence 5'-CA-3' to said genomic DNA, preferably to said bisulfite-treated genomic DNA, per CpG site. Preferably, the method further comprises performing a one-nucleotide extension reaction after said annealing in such case. Also preferably, the method comprises annealing per CpG site an oligonucleotide specifically annealing to a sequence immediately upstream of said CpG site and having a C as the terminal nucleotide, and performing pyrosequencing using said oligonucleotide as a sequencing primer. Preferably, in the method comprising annealing of an oligonucleotide, methylation of only the strand complementary to said oligonucleotide is analyzed. Also preferably, for a given CpG site, methylation of the CpG site of only one strand of DNA is analyzed, namely the CpG site as indicated above in Table 1; thus, preferably, for each CpG site indicated in Table 1, only one oligonucleotide is used in analysis.

According to the method for determining a survival probability, the methylation status of at least two CpG sites selected from Table 1 is determined. As is understood by the skilled person, accuracy of prediction may be increased by determining the methylation status of an increased number of CpG sites; thus, preferably, the methylation status of at least three, preferably at least five, more preferably at least eight, most preferably at least 13 of said CpG sites of Table 1 is determined. More preferably, the methylation status of at least three, preferably at least five, more preferably at least eight, most preferably at least 13 CpG sites related to, more preferably CpG sites selected from cg16336556, cg14270346, cg05646575, cg17431888, cg12510999, cg00832644, cg11056055, cg08804626, cg14983135, cg22522598, cg19184885, cg08729279, and cg10758824. More preferably, the methylation status of at least three, preferably at least five, more preferably at least six CpG sites related to, more preferably CpG sites selected from cg16336556, cg14270346, cg05646575, cg17431888, cg12510999, cg00832644, and cg11056055 is detected. Most preferably, the methylation status of all methylation sites of Table 1 is determined.

Preferably, a survival probability is determined by comparing the methylation status determined in a sample to a corresponding reference. Thus, preferably, the method for determining a survival probability comprises comparing the methylation status determined for

a CpG site in a sample to a reference. Thus, preferably, the method comprises further step a1) comparing the methylation status of said at least two CpG sites of step a) to references; and in step b) the determining is based on the comparison of step a1).

5 As used herein, the term "reference" relates to a reference value or a reference range, preferably derived from a population of subjects, preferably a population of apparently healthy subjects as specified herein above, more preferably a population of subjects suffering from colorectal carcinoma as specified elsewhere herein. Preferably, a reference value or reference range may be obtained from a second sample, which is a sample of healthy tissue,
10 more preferably of healthy tissue of the same tissue type, more preferably of healthy tissue from the same subject, most preferably of the same tissue type and from the same subject as the tumor sample. As used herein, the term "corresponding reference" relates to a reference value or reference range obtained by applying the same method, but to a different, i.e. reference, sample. Thus, in case the methylation status of cg16336556 is determined in a
15 sample, determining the methylation status of cg16336556 in a sample of an apparently healthy subject and/or determining the methylation status of cg16336556 in a sample of healthy tissue of the same patient provides a preferred corresponding reference. Also, if a score is calculated e.g. from a multitude of CpG sites, calculating a corresponding score from the same CpG sites of a sample of an apparently healthy subject and/or a sample of healthy
20 tissue of the same patient provides a preferred corresponding reference. As indicated above, at least two CpG sites are evaluated according to the present invention. As is understood by the skilled person, the value detected for a specific CpG site is compared to a reference for a corresponding CpG site, i.e. to a reference value or reference range pertaining to the CpG site having the same position in the genome. Thus, in case e.g. the average degree of methylation
25 is determined for the eight first CpG sites of Table 1, each of these values is compared to a corresponding reference value, respectively. As is also understood by the skilled person, values are compared to corresponding values, i.e. average degree of methylation values are compared to average degree of methylation values, numbers of cells comprising the CpG site in methylated form are compared to numbers of cells comprising the CpG site in methylated
30 form, and the like. Preferably, for CpG sites related to the specific CpG sites of Table 1, the above applies mutatis mutandis.

Preferably, the reference is derived from healthy tissue of the same patient, from a population of apparently healthy subjects. More preferably, the reference includes a value of a

methylation status or a score derived therefrom representing an average value of a population of subjects suffering colorectal cancer; more preferably the reference is a median of a methylation status or a score derived therefrom of a population of subjects suffering colorectal cancer. In the aforesaid cases, preferably, an unfavorable health state is determined if at least one of said CpG sites deviates from, preferably significantly deviates from, more preferably is lower than, most preferably is significantly lower than, the reference value. More preferably, an unfavorable survival probability is determined if a methylation status deviating from, preferably significantly deviating from, more preferably being lower than, most preferably significantly lower than, the reference is detected for at least two, more preferably at least four, even more preferably at least six, still more preferably at least eight, most preferably more than ten CpG sites. Most preferably, an unfavorable survival probability is determined if a methylation status deviating from, preferably significantly deviating from, more preferably being lower than, most preferably being significantly lower than, the reference is detected for at least two, more preferably at least four, even more preferably at least six, still more preferably at least eight, most preferably at least ten CpG sites selected from the CpG sites of Table 1, preferably selected from the first 13 CpG sites of Table 1. Preferably, a decrease of the methylation status of a CpG site compared to the reference is indicative of an unfavorable survival probability; and a non-decreased methylation status, i.e. preferably being similar or increased compared to the reference is indicative of a favorable survival probability. More preferably, a multitude of CpG sites is analyzed and the respective methylation statuses are combined into a score. Said score may be obtained by, e.g. summing up the β values of the respective CpG sites; preferably, said sum is compared to the corresponding sum obtained from a reference. More preferably, the β values are weighted for obtaining a score, more preferably with the weighting factors of Table 1 ("PCA weight");

Thus, in case all CpG sites of Table 1 are analyzed, the score preferably is calculated as score

$$= 0.31 * \beta_{(cg16336556)} + 0.31 * \beta_{(cg14270346)} + 0.30 * \beta_{(cg05646575)} + 0.29 * \beta_{(cg17431888)} + 0.26 * \beta_{(cg12510999)} + 0.24 * \beta_{(cg00832644)} + 0.24 * \beta_{(cg11056055)} + 0.23 * \beta_{(cg08804626)} + 0.22 * \beta_{(cg14983135)} + 0.22 * \beta_{(cg22522598)} + 0.22 * \beta_{(cg19184885)} + 0.21 * \beta_{(cg08729279)} + 0.21 * \beta_{(cg10758824)} + 0.19 * \beta_{(cg18195165)} + 0.19 * \beta_{(cg08617020)} + 0.18 * \beta_{(cg23750514)} + 0.15 * \beta_{(cg01131395)} + 0.15 * \beta_{(cg18736676)} + 0.12 * \beta_{(cg19340296)} + 0.07 * \beta_{(cg16399624)} \text{ (eq. 1).}$$

As will be appreciated, other weighting factors may be used; preferably, however, these weighting factors for all CpG sites used in the calculation are derived from the aforesaid weighting factors by a common mathematical operation, more preferably multiplication or division by the same factor. As will also be appreciated, the score calculated according to eq. 1 may assume values of from 0 (all CpG

sites having a β value of 0, i.e. being unmethylated) to 4.30 (all CpG sites having a β value of 1, i.e. being fully methylated). Preferably, a methylation status increased compared to the median methylation status of a population of subjects suffering from colorectal cancer is indicative of a favorable survival probability; also preferably, a methylation status decreased compared to the median methylation status of a population of subjects suffering from colorectal cancer is indicative of an unfavorable survival probability. Thus, preferably, a score, preferably calculated according to eq. 1, increased compared to the median score of a population of subjects suffering from colorectal cancer is indicative of a favorable survival probability; also preferably, a score, preferably calculated according to eq. 1, decreased compared to the score of a population of subjects suffering from colorectal cancer is indicative of an unfavorable survival probability. More preferably, a score calculated according to eq. 1 being higher than about 2.85, more preferably higher than 2.86, is indicative of a favorable survival probability; also preferably, a score being lower than about 2.85, more preferably lower than or equal to 2.86, is indicative of an unfavorable survival probability.

Methods for determining a, preferably significant, more preferably statistically significant, deviation of a methylation status from a reference are known to the skilled person; preferably, a value and a reference value are determined to be essentially identical if the difference between two values is, preferably, not significant and shall be characterized in that the value is within at least the interval between 1st and 99th percentile, 5th and 95th percentile, 10th and 90th percentile, 20th and 80th percentile, 30th and 70th percentile, 40th and 60th percentile of the reference value, preferably, the 50th, 60th, 70th, 80th, 90th or 95th percentile of the reference value. Statistical tests for determining whether two amounts are essentially identical are well known in the art and are also described elsewhere herein. Conversely, an observed difference for two values, on the other hand, shall preferably be statistically significant. A difference in value is, preferably, significant outside of the interval between 45th and 55th percentile, 40th and 60th percentile, 30th and 70th percentile, 20th and 80th percentile, 10th and 90th percentile, 5th and 95th percentile, 1st and 99th percentile of the reference value. Whether a difference is statistically significant can be determined without further ado by the person skilled in the art using various well known statistic evaluation tools, e.g., determination of confidence intervals, p-value determination, Student's t-test, Mann-Whitney test etc.. Details are found in Dowdy and Wearden, Statistics for Research, John Wiley & Sons, New York 1983. Preferred confidence intervals are at least 90%, at least 95%,

at least 97%, at least 98% or at least 99 %. The p-values are preferably ≤ 0.1 , more preferably ≤ 0.05 , still more preferably ≤ 0.01 , even more preferably ≤ 0.005 , or, most preferably ≤ 0.0001 . Preferably, the probability envisaged by the present invention allows that the determination will be correct for at least 60%, more preferably at least 70%, still more preferably at least 80%, or, most preferably, at least 90% of the subjects of a given cohort or population. Further methods of evaluating statistical significance of differences in methylation are described herein below in the Examples.

As will be understood by the skilled person, other methods of establishing a reference and/or a score according to the invention can be envisaged.. E.g., a reference value may be established by determining the mean methylation status of a population of subjects suffering from colorectal cancer known to have an unfavorable survival probability, and may be used as a cutoff value; as will be understood, in such case a methylation status similar to the reference would be indicative of an unfavorable survival probability. Further, preferably, the lower limit of normal (LLN) may be used as a cutoff. Also, values derived from any of the aforesaid populations may be divided in halves, tertiles, quartiles, pentiles, or the like. Also, the lowest or the lower two tertiles of values of a population of subjects suffering from colorectal cancer may be regarded as being indicative of unfavorable survival probability. The specific choice of reference and/or score will mainly be governed on the specific sensitivity and specificity required, but also by other parameters such as the particular population of interest. The skilled person has means and methods at hand enabling appropriate election.

Advantageously, it was found in the work underlying the present invention that the methylation status of the indicated CpG sites is an independent indicator of the mortality risk of a subject suffering from colorectal cancer. Thus, a prognostic classifier, preferably based on 20 CpGs, could be constructed, which improves prognosis prediction for overall survival and disease-free survival colorectal cancer patients and which provides for a significant reduction of prediction error.

The definitions made above apply mutatis mutandis to the following. Additional definitions and explanations made further below also apply for all embodiments described in this specification mutatis mutandis.

The present invention further relates to a method for treating colorectal cancer in a subject suffering from colorectal cancer comprising

a) determining a survival probability of said subject according to the method according to the present invention;

5 b) treating said subject with at least one of surgery, chemotherapy, and radiotherapy if an unfavorable survival probability is determined; and treating said subject by avoiding chemotherapy and/or radiotherapy if a favorable survival probability is determined; and, thereby,

c) treating colorectal cancer in said subject.

10

The method of treating a subject of the present invention, preferably, is an in vivo method. Moreover, it may comprise steps in addition to those explicitly mentioned above. For example, further steps may relate, e.g., to diagnosing colorectal cancer before the steps of the method for determining a survival probability, providing further therapeutic options, or
15 administering one or more therapeutic measures to said subject, depending on the result of said method. Moreover, one or more of said steps may be performed by automated equipment.

According to the invention, the possibility to establish a survival probability for a subject
20 enables the medical practitioner to better select an appropriate therapy. Preferably, in the method of treating a subject, the primary tumor is removed by surgery, including classical surgery, i.e. resection of tumor, ablation (e.g. radiofrequency ablation), and cryotherapy (cryosurgery). Thus, preferably, the method of treating a subject of the present invention, preferably, comprises the step of removal of the primary tumor, more preferably before or
25 after further treatment is administered. Preferably, the method of treating a subject further comprises providing at least one of close monitoring and/or lifestyle recommendations and/or treatment methods to said subject.

The term "close monitoring", as used herein, relates to medically examining a subject for
30 signs of relapse and/or metastasis at least once within 3 months, preferably within two months, more preferably within one month for a period of at least 12 months, more preferably at least 18 months, still more preferably at least 24 months, most preferably at least 35 months.

The term "lifestyle recommendations", as used herein, relates to recommendations decreasing the probability of relapse and/or metastasis. Preferably, such recommendations are recommendations to reduce or quit alcohol consumption, to reduce or quit smoking, to reduce body weight, to increase exercise and/or to use healthy nutrition.

5

Treatment methods for a patient suffering from colorectal cancer, in particular after removal of the primary tumor, preferably include chemotherapy, radiotherapy, targeted therapy, and immunotherapy.

10 As used herein, the term "chemotherapy" relates to treatment of a subject with an antineoplastic drug. Preferably, chemotherapy is a treatment including alkylating agents (e.g. cyclophosphamide), platinum (e.g. carboplatin), antimetabolites (e.g. 5-Fluorouracil), anthracyclines (e.g. doxorubicin, epirubicin, idarubicin, or daunorubicin), topoisomerase II inhibitors (e.g. etoposide, irinotecan, topotecan, camptothecin, or VP16), anaplastic
15 lymphoma kinase (ALK)-inhibitors (e.g. Crizotinib or AP26130), aurora kinase inhibitors (e.g. N-[4-[4-(4-Methylpiperazin-1-yl)-6-[(5-methyl-1H-pyrazol-3-yl)amino]pyrimidin-2-yl]sulfanylphenyl]cyclopropanecarboxamide (VX-680)), or Iodine131-1-(3-iodobenzyl)guanidine (therapeutic metaiodobenzylguanidine), or histone deacetylase (HDAC) inhibitors, alone or any suitable combination thereof. It is to be understood that
20 chemotherapy, preferably, relates to a complete cycle of treatment, i.e. a series of several (e.g. four, six, or eight) doses of antineoplastic drug or drugs applied to a subject, which may be separated by several days or weeks without such application.

The terms "radiation therapy" and "radiotherapy" are known to the skilled artisan. The term
25 relates to the use of ionizing radiation to treat or control cancer.

The term "targeted therapy", as used herein, relates to application to a patient a chemical substance known to block growth of cancer cells by interfering with specific molecules known to be necessary for tumorigenesis or cancer or cancer cell growth. Examples known to
30 the skilled artisan are small molecules like, e.g. PARP-inhibitors (e.g. Iniparib), antiangiogenic agents (e.g. Bevacizumab, Ramucirumab, Ziv-aflibercept), signalling inhibitors (e.g. cetuximab or panitumumab), or kinase inhibitors (e.g. Regorafenib).

The term "immunotherapy" as used herein relates to the treatment of cancer by modulation of the immune response of a subject. Said modulation may be inducing, enhancing, or suppressing said immune response, e.g. by administration of at least one cytokine, and/or of at least one antibody specifically recognizing cancer cells. The term "cell based immunotherapy" relates to a cancer therapy comprising application of immune cells, e.g. T-cells, preferably tumor-specific NK cells, to a subject.

The present invention also relates to a method for patient monitoring comprising the steps of the method for determining a survival probability and/or of the method for treating colorectal cancer of the present invention and the further steps or steps of providing close monitoring and/or lifestyle recommendations in case an unfavorable survival probability and/or an increased mortality risk is detected.

The present invention further relates to a use of means for determining the methylation status of, at least two CpG sites related to at least two CpG sites of Table 1, preferably at least two CpG sites of Table 1, for manufacturing a diagnostic means or device for determining a survival probability of a subject suffering from colorectal cancer.

The present invention also relates to a data collection, preferably comprised on a data carrier, comprising the positions of at least two, preferably at least three, more preferably at least five, even more preferably at least eight, still more preferably at least 13, most preferably at least 16 CpG sites selected from Table 1 and/or CpG sites related thereto; preferably of from three to all, more preferably of from ten to all, even more preferably of from 15 to all CpG sites selected from Table 1 and/or CpG sites related thereto. Preferably, the data collection further comprises reference values or reference ranges for the methylation status of said CpG sites, preferably as specified elsewhere herein, more preferably as specified herein in the Examples.

The term "data collection" refers to a collection of data which may be physically and/or logically grouped together. Accordingly, the data collection may be implemented in a single data storage medium or in physically separated data storage media being operatively linked to each other. Preferably, the data collection is implemented by means of a database. Thus, a database as used herein comprises the data collection on a suitable storage medium. Moreover, the database, preferably, further comprises a database management system. The database management system is, preferably, a network-based, hierarchical or object-oriented

database management system. Furthermore, the database may be a federal or integrated database. More preferably, the database will be implemented as a distributed (federal) system, e.g. as a Client-Server-System. More preferably, the database is structured as to allow a search algorithm to compare a test data set with the data sets comprised by the data collection.

5 Specifically, by using such an algorithm, the database can be searched for similar or identical data sets being indicative for a survival probability as set forth above (e.g. a query search). Consequently, the information obtained from the data collection can be used, e.g., as a reference for the methods of the present invention described above.

10 The term “data storage medium” as used herein encompasses data storage media which are based on single physical entities such as a CD, a CD-ROM, a hard disk, optical storage media, a diskette, or a sheet of paper. Moreover, the term further includes data storage media consisting of physically separated entities which are operatively linked to each other in a manner as to provide the aforementioned data collection, preferably, in a suitable way for a
15 query search.

Further, the present invention relates to a kit comprising means for determining the methylation status of at least two CpG sites related to at least two CpG sites of Table 1, preferably at least two CpG sites selected from the CpG sites of Table 1, more preferably
20 selected from the list consisting of cg16336556, cg14270346, cg05646575, cg17431888, cg12510999, cg00832644, cg11056055, cg08804626, cg14983135, cg22522598, cg19184885, cg08729279, and cg10758824 in a sample of a subject suffering from colorectal cancer, and a data collection according to the present invention.

25 Moreover, the present invention relates to a device comprising an analysis unit comprising means for determining the methylation status of at least two CpG sites related to at least two CpG sites of Table 1, preferably of at least two CpG sites selected from the CpG sites of Table 1, more preferably selected from the list consisting of in a sample of a subject suffering from colorectal cancer, and an evaluation unit comprising a data collection according to the
30 present invention.

In view of the above, the following embodiments are particularly envisaged:

1. A method for determining a survival probability of a subject suffering from colorectal
35 cancer comprising

- a) detecting the methylation status of at least two CpG sites related to at least two CpG sites of Table 1 in a sample of said subject and,
 - b) based on the methylation status detected in step a), determining the survival probability of said subject.
- 5 2. The method of embodiment 1, wherein said at least two CpG sites are selected from the list consisting of cg16336556, cg14270346, cg05646575, cg17431888, cg12510999, cg00832644, cg11056055, cg08804626, cg14983135, cg22522598, cg19184885, cg08729279, and cg10758824; preferably consisting of cg16336556, cg14270346, cg05646575, cg17431888, cg12510999, cg00832644, and cg11056055.
- 10 3. The method of embodiment 1 or 2, wherein the methylation status of at least three, preferably at least eight, more preferably at least 13, most preferably at least 16 of said CpG sites is detected.
4. The method of any one of embodiments 1 to 3, wherein the methylation status of 20 methylation sites is determined.
- 15 5. The method of any one of embodiments 1 to 4, wherein said detecting the methylation status of a CpG site is detecting the average degree of methylation of said site from at least 10, preferably at least 25, more preferably at least 100 cells.
6. The method of any one of embodiments 1 to 5, wherein a decreased average degree of methylation compared to a reference is indicative of an unfavorable survival probability.
- 20 7. The method of any one of embodiments 1 to 6, wherein said method comprises further step a1) comparing the methylation status of said at least two CpG sites of step a) to a reference or to references; and wherein in step b) the determining is based on the comparison of step a1).
8. The method of embodiment 7, wherein said reference is a reference value or a
25 reference range.
9. The method of embodiment 7 or 8, wherein said reference is obtained from a population of subjects suffering from colorectal cancer, preferably non-metastasizing colorectal cancer.
10. The method of any one of embodiment 1 to 9, wherein determining said survival
30 probability comprises determining a mortality risk.
11. The method of any one of embodiments 1 to 10, wherein said mortality risk is a disease-specific mortality risk.
12. The method of embodiment 10 or 11, wherein said mortality risk not an overall mortality risk.
- 35 13. The method of any one of embodiments 1 to 12, wherein said sample is a tissue sample, preferably is a tumor sample, more preferably is a sample of tumor cells.
14. The method of any one of embodiments 1 to 13, wherein said methylation status is

detected in cells of said subject, preferably is detected in tumor cells.

15. The method of any one of embodiments 1 to 14, wherein said method comprises isolating genomic DNA from said sample.

5 16. The method of any one of embodiments 1 to 15, wherein said subject is a mammal, preferably a human.

17. The method of any one of embodiments 1 or 16, wherein said method comprises treating said genomic DNA with a bisulfite.

10 18. The method of any one of embodiments 1 to 17, wherein said method comprises annealing an oligonucleotide specifically annealing to a sequence immediately upstream of said CpG site and comprising a 3'-terminal sequence 5'-CG-3' and/or an oligonucleotide specifically annealing to a sequence immediately upstream of said CpG site and comprising a 3'-terminal sequence 5'-CA-3' to said genomic DNA, preferably to said bisulfite-treated genomic DNA, per CpG site.

15 19. The method of embodiment 18, wherein said method comprises performing a one-nucleotide extension reaction after said annealing.

20 20. The method of any one of embodiments 1 to embodiment 17, wherein said method comprises annealing per CpG site an oligonucleotide specifically annealing to a sequence immediately upstream of said CpG site and having a C as the terminal nucleotide and performing pyrosequencing using said oligonucleotide as a sequencing primer.

21. The method of any one of embodiments 1 to 20, wherein said determining a survival probability is not diagnosing disease.

22. The method of any one of embodiments 1 to 21, wherein the CpG sites correlate with positions in the human genome as shown in Table 1.

25 23. A method for treating colorectal cancer in a subject suffering from colorectal cancer comprising

a) determining a survival probability of said subject according to the method according to any one of embodiments 1 to 22;

30 b) treating said subject with at least one of surgery, chemotherapy, and radiotherapy if an unfavorable survival probability is determined; and treating said subject by avoiding chemotherapy and/or radiotherapy if a favorable survival probability is determined; and, thereby,

c) treating colorectal cancer in said subject.

24. The method of embodiment 23, wherein said chemotherapy is high-dose chemotherapy and/or wherein said radiotherapy is high-dose radiotherapy.

35 25. Use of the methylation status of genomic DNA or means for the determination thereof in a sample of a subject suffering from colorectal cancer for determining a survival probability of said subject, preferably for predicting the mortality risk of said subject.

26. The use of embodiment 25, wherein said use comprises detecting the methylation status of at least two CpG sites selected from the CpG sites of Table 1, preferably selected from the list consisting of cg16336556, cg14270346, cg05646575, cg17431888, cg12510999, cg00832644, cg11056055, cg08804626, cg14983135, cg22522598, cg19184885, cg08729279, and cg10758824.
27. Use of means for determining the methylation status of at least two CpG sites related to at least two CpG sites of Table 1, preferably at least two CpG sites of Table 1, for manufacturing a diagnostic means or device for determining a survival probability of a subject suffering from colorectal cancer.
28. A method for patient monitoring comprising the steps of the method according to any one of embodiments 1 to 24 and providing close monitoring and/or lifestyle recommendations in case an unfavorable survival probability and/or an increased mortality risk is detected.
29. A data collection, preferably comprised on a data carrier, comprising the positions of at least two, preferably at least three, more preferably at least five, even more preferably at least eight, most preferably at least 15 CpG sites selected from Table 1; preferably of from three to all, more preferably of from ten to all, even more preferably of from 15 to all CpG sites selected from Table 1 and/or CpG sites related thereto.
30. The data collection of embodiment 29, further comprising reference values or reference ranges for the methylation status of said CpG sites.
31. A kit comprising means for determining the methylation status of at least two CpG sites related to at least two CpG sites of Table 1, preferably at least two CpG sites selected from the CpG sites of Table 1, more preferably selected from the list consisting of cg16336556, cg14270346, cg05646575, cg17431888, cg12510999, cg00832644, cg11056055, cg08804626, cg14983135, cg22522598, cg19184885, cg08729279, and cg10758824 in a sample of a subject suffering from colorectal cancer, and a data collection according to embodiment 29 or 30.
32. A device comprising an analysis unit comprising means for determining the methylation status of at least two CpG sites related to at least two CpG sites of Table 1, preferably of at least two CpG sites selected from the CpG sites of Table 1, more preferably selected from the list consisting of in a sample of a subject suffering from colorectal cancer, and an evaluation unit comprising a data collection according to embodiment 29 or 30.
33. Use of means for determining the methylation status of at least two CpG sites related to at least two CpG sites of Table 1, preferably of at least two CpG sites selected from the CpG sites of Table 1, and a data collection according to embodiment 29 or 30 for manufacturing a diagnostic means or device for determining a survival probability of a subject suffering from colorectal cancer.

All references cited in this specification are herewith incorporated by reference with respect to their entire disclosure content and the disclosure content specifically mentioned in this specification.

5 Figure Legends

Fig. 1: Study design and analysis flowchart

Fig. 2: Illustration of weighting in the IHW model. The left side of the plot, with β -value difference (x-axis) ranging between -0.5 and 0, represents the CpG sites with lower methylation in tumor than in normal mucosa tissue, the right side with β -value difference between 0 and 0.5, the CpG sites with higher methylation in tumor than in normal mucosa tissue. In all 20 folds, illustrated by the different colored lines, the highest weighting (y-axis) was given to CpG sites with clearly lower methylation levels in tumor than in normal mucosa tissue and no weight was given to CpG sites with higher methylation levels in tumor tissue.

Fig. 3: Prediction error curve and AUC curve for OS (A, B) and DSS (C, D) of CRC patients with non-metastatic disease. In comparison to the reference, in the covariate model (Clinical covariates) the prediction error was calculated based on the clinical variables age, gender, smoking behavior, MSI-status, tumor stage and tumor location (colon proximal, colon distal, rectum). By the combination of the clinical variables with our ProMCol classifier (Clinical covariates + ProMCol) the prediction error was further reduced.

Fig. 4: Kaplan-Meier plots for OS of all CRC patients with non-metastatic disease (stage I-III) dependent on the methylation status of the combined methylation marker analyzed in the screening cohort (A) and in the validation cohort (B). OS of only stage II and III CRC patients dependent on the methylation status of the combined methylation marker is shown for the screening cohort (C) and the validation cohort (D). Group low: low expression of the combined methylation marker (value $\leq$ median), group high: high expression of the combined methylation (value $>$ median). Given p-values are based on log-rank tests between the respective groups. The shape of the Kaplan-Meier plots support the assumption of proportional hazards.

The following Examples shall merely illustrate the invention. They shall not be construed, whatsoever, to limit the scope of the invention.

Example 1: Study cohort

5 All included patients were enrolled in the ongoing population-based case-control study DACHS (Darmkrebs: Chancen der Verhütung durch Screening), described in detail elsewhere (Brenner et al.; Ann Intern Med 2011;154(1):22-30; Lilla et al.; Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 2006;15(1):99-107). In total,
10 22 hospitals in the study area of the Rhine-Neckar-Odenwald region of southwestern Germany were involved in recruitment. Extensive patient information was gained by standardized questionnaires, carried out by trained interviewers. Follow-up (FU) information concerning vital status, date and cause of death was obtained by the local population registries and health authorities three and five years after CRC diagnosis (Hoffmeister et al.; Journal of
15 the National Cancer Institute 2015;107(6):djv045). The study was approved by the ethical committees of the Medical Faculty of the University of Heidelberg and of the Medical Chambers of Baden-Württemberg and Rhineland-Palatinate. Written informed consent was obtained from each participant. In the present study, only patients with non-metastatic disease recruited between 2003 and 2007 were included (for whom 5-year FU was complete at the
20 time of this analysis). Genome-wide DNA methylation data are so far available for patients recruited until 2007 only. Eligible for the screening cohort (n=689) were all DACHS patients whose tumors were examined in the pathology institutes in Heilbronn, Ludwigshafen, Mannheim and Speyer whereas for the validation cohort (n=408) only tumors examined in the pathology institute of Heidelberg were selected. In addition to the patients' tumor tissues, the
25 validation cohort contained corresponding normal mucosa tissue samples, enabling the analysis of tumor and normal mucosa tissue pairs for 34 patients. Patients who had received neoadjuvant therapy were excluded for the survival analysis. The median FU-time for all patients in the screening and the validation cohort was approximately five (4.99) years. Table 2 gives an overview of the clinicopathological characteristics of the included patients.

30

Table 2: Clinicopathological characteristics of the screening (A) and the validation (B) cohort

A. Screening cohort		Non-metastatic CRC	Metastatic CRC
Total number		587	101
Gender	male	319 (54.3%)	62 (61.4%)
	female	268 (45.7%)	39 (38.6%)

Age distribution at diagnosis (years)	range mean median	33-94 69.3 69	46-93 67.8 68
Stage at diagnosis	stage I stage II stage III stage IV	124 (21.1%) 224 (38.2%) 239 (40.7%)	101 (100%)
Neoadjuvant treatment	yes no no information	14 572 1	5 94 2
CRC location	colon rectum	395 (67.3%) 192 (32.7%)	69 (68.3%) 32 (31.7%)
CC location	proximal distal	231 (39.3%) 164 (27.9%)	29 (28.7%) 40 (39.6%)
CRC phenotype*	MSS MSI no information	476 (81.1%) 67 (11.4%) 44 (7.5%)	93 (92.1%) 3 (3.0%) 5 (5.0%)
Smoking behavior	never former current no information	256 (43.6%) 239 (40.7%) 91 (15.5%) 1 (0.2%)	43 (42.6%) 39 (38.6%) 19 (18.8%) -

B. Validation cohort		Non-metastatic CRC	Metastatic CRC
Total number		308	58
Gender	male female	177 (57.5%) 131 (42.5%)	33 (56.9%) 25 (43.1%)
Age distribution (years)	range mean median	37-91 68.3 69	37-89 68.2 69.5
Stage at diagnosis	stage I stage II stage III stage IV	63 (20.5%) 131 (42.5%) 114 (37.0%)	58 (100%)
Neoadjuvant treatment	yes no	34 274	0 58
CRC location	colon rectum no information	190 (61.7%) 118 (38.3%)	41 (70.7%) 16 (27.6%) 1 (1.7%)
CC location	proximal distal	102 (33.1%) 88 (28.6%)	22 (37.9%) 19 (32.8%)
CRC phenotype¹	MSS MSI no information	275 (89.9%) 29 (9.5%) 4 (1.3%)	56 (96.8%) 1 (1.6%) 1 (1.6%)
Smoking behavior	never former current no information	133 (43.5%) 120 (39.2%) 54 (17.6%) 1 (0.3%)	21 (36.2%) 26 (44.8%) 10 (17.2%) 1 (1.7%)

* MSI was determined using a mononucleotide marker panel consisting out of BAT25, BAT26, and CAT25 23

Example 2: DNA isolation

Tissue samples were collected from all de-central pathology institutes in the tissue bank of the National Center for Tumor Diseases (NCT, Heidelberg, Germany). All samples were provided in accordance with the regulations of the tissue bank and the approval of the ethics committee of Heidelberg University. For each sample, tumor DNA was isolated from four formalin-fixed and paraffin-embedded (FFPE) tissue slices à 5 µm. A hematoxylin-stained slice of every tumor block was evaluated by an experienced pathologist in order to mark the regions with high tumor cell content. After deparaffinization, the DNA of the manually microdissected tumor tissue of the screening cohort was isolated following a semi-automated protocol using the Maxwell® 16 MDx instrument (Promega, USA) in combination with the DNeasy Blood & Tissue Kit (Qiagen, Germany) whereas the DNA of the validation cohort was extracted manually with the DNeasy Blood & Tissue Kit according to manufacturer's recommendations. For both cohorts the isolated DNA was eluted with 50µl elution buffer.

Example 3: Methylation profiling using the Infinium HumanMethylation450 BeadChip array (Illumina)

DNA quality and quantity was checked using the Quant-iT™ PicoGreen® dsDNA Assay Kit (Thermo Fisher Scientific, USA) and the integrated Infinium HD FFPE QC Assay (Illumina, USA). Samples were randomized in 96 sample batches according to OSAT 24 and replicates were included to check for batch effects. A minimum of 100 ng DNA, but if available, 250 ng DNA diluted in 45 µl RNase-free water were used per sample for the following experimental steps of the methylation profiling, all performed according to manufacturer's recommendations. After bisulfite conversion using the Zymo EZ96 DNA Methylation Kit (Zymo Research, USA), the DNA restoration step (Illumina, USA) for FFPE samples was conducted, followed by the genome-wide methylation analysis using the Infinium HumanMethylation450 BeadChip (Illumina, USA) interrogating over 485,000 CpG loci. Samples of the screening cohort were analyzed more than one year apart from samples of the validation cohort. The methylation score of each CpG site was defined as β -value, ranging from 0 to 1, with 0 representing a completely unmethylated status and 1 a fully methylated status.

Example 4: Preprocessing and normalization of methylation data

Preprocessing and statistical analyses were all performed using the computational environment R, version 3.3.1 (www.r-project.org/). Raw data files generated by the iScan

array scanner were read and preprocessed using the 'minfi' package, included in the Bioconductor collection of R packages. The standard Illumina normalization procedure ('preprocessIllumina') was used to correct for technical differences between the Infinium I and II assay designs. For all analyses filtering criteria were applied according to Sturm et al. (Cancer cell 2012;22(4):425-37). In the screening cohort, probes that failed in more than 10% samples, based on detection p-value using a significance level of 0.01, were excluded. To allow for an independent validation, no filtering was applied to the validation cohort. Instead, here exactly the same CpG sites were used as in the screening cohort.

Missing information in all clinical variables, except for microsatellite instability (MSI), were imputed with the R-package 'mice' using information from the other clinical variables as listed in Table 1. For the imputation of MSI status, a random forest based on 100 CpG sites was applied. These 100 CpG sites were pre-elected using distance correlation sure independence screening (DC-SIS) (Li et al.; J Am Stat Assoc 2012;107(499):1129-39).

Example 5: Statistical analyses

A paired Wilcoxon signed-rank test was used to find the differentially methylated CpG sites between the tumor and normal tissue pairs. The difference of tumor and normal tissues for every CpG site was estimated via the (pseudo-)median of the sample differences. For statistical inference, a filtering step for standard deviation > 0.05 was applied to the selected CpG sites. A variable screening for prognostic CpG sites for overall survival (OS) was performed for single CpG sites using marginal testing based on a Cox model adjusted for age as continuous covariate and gender, smoking behavior, MSI-status, tumor stage (2 vs. 1, 3 vs. 1) and tumor location (colon proximal, colon distal or rectum) as categorical covariates. Subsequently, the p-values were adjusted via independent hypothesis weighting (IHW) (Ignatiadis et al.; Nat Methods 2016;13(7):577-80) with 20 folds and a significance level of 0.1. As auxiliary covariate for IHW the estimate of the mean difference in methylation between tumor and normal mucosa tissue was used as ordinal covariate, manually partitioned into 16 intervals $([-1,-0.35), [-0.35,-0.3), [-0.3,-0.25), \dots, [0.25,0.3), [0.3,0.35), [0.35,1])$. For the 1000 CpG sites with the smallest adjusted IHW p-value, we evaluated the apparent Brier score regarding 3-year survival in the screening cohort, using the R package 'pec'. Finally, the 20 CpG sites with the smallest apparent Brier Score in the screening cohort were selected for the construction of the methylation-based prognostic classifier. The first principal component of the β -values of these 20 CpG sites was calculated. The ProMCol classifier was

obtained by multiplying the β -values of the 20 CpG sites with the respective weights of the first principal component (this is essentially the score of the first principal component omitting the centering of the β -values). For determining the number of CpG sites that we used for the classifier, we applied a 10 fold internal cross validation approach with 3 repetitions.

5

Two Cox-models were fitted using the R package ‘survival’. The first model included only the clinical covariates (as specified above), the second one included both the clinical covariates and the prognostic classifier. The given hazard ratios (HR) are based on a change from the lower to the upper quartile of the prognostic classifier. Both, the Cox model based only on clinical covariates and the Cox model based on clinical covariates and the prognostic classifier (which were fitted using the screening cohort data) were independently validated in the validation cohort. The prediction error curve and the Brier score for the 3-years survival in the validation cohort were evaluated using the loss function approach described in Gerds and Schumacher (Biometrical Journal 2006;48(6):1029–40) using the R package ‘pec’. The area under the curve (AUC) curves were calculated following the incident/dynamic approach described in the work of Song and Zhou (Statistica Sinica 2008(18):947-65) using the R package ‘survAUC’.

Complementary, we performed a replication analysis, fitting new Cox models to the validation cohort. Kaplan-Meier curves were generated for the two subgroups ‘ProMCol high’ and ‘ProMCol low’ using the median value of the prognostic classifier as cut-off. Log-rank tests were used to assess the significance of the survival differences between the groups. Moreover, the Kaplan-Meier plots support the assumption of proportional hazards. In all analyses, a two-tailed significance level of 0.05 was used.

25

Example 6: Analysis of CpG based methylation data from the genome-wide analysis

After preprocessing and filtering for standard deviation >0.05 , 269,306 probes were included in the analysis. The replicates ($n=11$) analyzed in both cohorts showed correlations with Spearman correlation coefficient $\rho \geq 0.98$. As the tumor and normal mucosa tissue pairs ($n=34$), available from the validation set, were used for the IHW procedure these tumor samples were excluded from the validation of our classifier later on. Of 688 patients in the screening cohort and 366 residual patients in the validation cohort, 52 patients had to be excluded because of neoadjuvant treatment. Highly distinct pathological and clinical characteristics of patients with non-metastatic and metastatic disease require a separate

30

analysis of these two patients groups. As the statistical power was too limited for the analysis of the metastatic cases (94 patients in the screening and 52 in the validation cohort), impeding the identification of reliable prognostic biomarkers, we focused our analysis on the patients with non-metastatic CRC. Finally, 572 patients with non-metastatic disease were included for further analysis in the screening cohort and 249 patients in the validation cohort. An overview over the study design and the analysis steps is given in Figure 1.

Example 7: Identification of candidate CpG sites for the OS of patients with non-metastatic CRC

As a starting point, we determined differentially methylated CpG sites across the genome in the analyzed 34 tumor and normal mucosa tissue pairs. The difference in methylation between tumor and normal tissue served as covariate for adjustment of the p-values via IHW in the Cox model, used for the identification of prognostic CpG sites for OS. IHW improves the power of large-scale multiple testing by adjusting the p-values using weighted multiple-testing using data-driven weights. Figure 2 illustrates the assigned weights according to the β -value difference of tumor and normal mucosa tissue. As a result of the IHW procedure, we obtained high weights for CpG sites with low methylation levels in tumor compared to normal mucosa tissue. In contrast, the IHW procedure assigned no weight to CpG sites with higher methylation levels in tumor tissue than in normal mucosa tissue. Consequently, the IHW procedure removed 114,405 CpG sites without a positive weight (weights = 0) from the further model building process.

Example 8: Construction of the ProMCol classifier in the screening cohort

After the adjustment of the p-values via IHW, only the 1000 CpG sites showing the smallest p-values in the survival analysis were used for the construction of the prognostic classifier using the screening cohort. The Brier Score, a measure for the accuracy of probabilistic predictions, was calculated for the 3-years survival of the patients with non-metastatic disease. The 20 CpG sites with the best Brier Score (depicted with their characteristics in Table 3) were selected for the prognostic classifier.

Table 3: CpG sites with the best Brier Score

	CpG ID¹	Chr.	Position²	UCSC RefGene	Location in gene	Enhancer	DHS	PCA weight
1	cg16336556	2	33295138	LTBP1	Body	TRUE	TRUE	0,31
2	cg14270346	9	38026076	SHB	Body	TRUE	TRUE	0,31
3	cg05646575	12	132221835	SFRS8	Body	TRUE	TRUE	0,30
4	cg17431888	1	183203863	LAMC2	Body	TRUE	TRUE	0,29

5	cg12510999	2	28022223	RBKS	Body	TRUE	TRUE	0,26
6	cg00832644	1	162343785	C1orf111	3'UTR	NA	NA	0,24
7	cg11056055	7	116356738	MET	Body	TRUE	TRUE	0,24
8	cg08804626	6	1953086	GMD5	Body	TRUE	TRUE	0,23
9	cg14983135	7	48129822	UPP1	5'UTR	NA	NA	0,22
10	cg22522598	19	15360953	BRD4	Body	TRUE	TRUE	0,22
11	cg19184885	17	45699322	NPEPPS	3'UTR	TRUE	TRUE	0,22
12	cg08729279	2	216708378			TRUE	TRUE	0,21
13	cg10758824	1	164606935	PBX1	Body	TRUE	NA	0,21
14	cg18195165	7	5549570	FBXL18	Body	NA	NA	0,19
15	cg08617020	16	87894491	SLC7A5	Body	NA	TRUE	0,19
16	cg23750514	11	2790418	KCNQ1	Body	TRUE	TRUE	0,18
17	cg01131395	13	47207812	LRCH1	Body	TRUE	NA	0,15
18	cg18736676	6	109053447			TRUE	NA	0,15
19	cg19340296	6	52288804	EFHC1	Body	NA	TRUE	0,12
20	cg16399624	10	134553188	INPP5A	Body	NA	NA	0,07

¹ Unique CpG locus identifier from the 450K methylation array

² All coordinates referring to RefSeq genes (hg19) and NCBI built 37

Chr: chromosome, UTR: untranslated region, DHS: Dnase I Hypersensitivity Site (experimentally determined by the ENCODE project), PCA: Principal component analysis

5

Example 9: Associations of the ProMCol classifier with survival of patients with non-metastatic CRC

Using the ProMCol classifier, we fitted a Cox regression analysis in the screening cohort. In one model only the standard clinical variables (gender, age, tumor stage, tumor location, smoking behavior and MSI-status) were included, in the other model the clinical variables in combination with the ProMCol classifier. Here, the ProMCol classifier was significantly associated with OS of patients with non-metastatic disease with HR=0.51, 95% CI=0.41-0.63 and p=6.2E-10 (results are shown in Table 4A). Patients with a high methylation status showed a better prognosis for OS than patients with a low methylation status, meaning the higher the ProMCol classifier value, the better the patients' prognosis. For an independent validation of the ProMCol classifier, the prediction error for the OS of patients with non-metastatic CRC was calculated in the validation cohort. This analysis revealed a clear advantage in prediction probability for adding the ProMCol classifier to the model (Figure 3A). Whereas the prediction error, calculated by the model using only clinical variables was 0.127 for the three-year survival of the patients, a model combining the clinical variables with the ProMCol classifier improved the prediction with a smaller error value of 0.120. For the four-year survival the prediction error could be reduced from 0.153 with clinical variables to 0.140 using the combination of clinical variables and the ProMCol classifier. As an addition, we computed a time-dependent AUC curve (Figure 3B), showing for the three-year (four-

year) OS of the patients an increase of AUC from 0.705 (0.704) calculated only with clinical variables to 0.750 (0.743) calculated with clinical variables and the ProMCol classifier.

Unadjusted Kaplan-Meier plots for OS are shown for all patients with non-metastatic disease (stage I-III) and specified for stage II and III patients dependent on the methylation status of the ProMCol classifier in both cohorts in Figure 4. Patients with a high methylation status expressed by the ProMCol classifier value > 2.86 (corresponding to the median of the classifier in the screening cohort) clearly showed a better prognosis for OS than patients with a low methylation status corresponding to a classifier value ≤ 2.86 ($p=3.0E-5$ in the screening and $p=8.5E-4$ in the validation cohort). After three (five) years of FU, 92.8% (82.5%) of the CRC patients in the screening cohort with a ProMCol classifier value > 2.86 in contrast to 79.2% (68.2 %) with a classifier value ≤ 2.86 , were still alive. In the validation cohort, the absolute three-year and five-year survival rates of CRC patients with a ProMCol classifier value > 2.86 were 90.1 % and 82.4 %, respectively, compared to 75.4 % and 66.1 % for patient with a low ProMCol classifier value (≤ 2.86).

To ensure the specificity for CRC, we analyzed the association of the ProMCol classifier with disease-specific survival (DSS) of patients with non-metastatic disease according to our procedure for OS. The multivariate Cox regression analysis revealed a significant association in the screening cohort with $HR=0.50$, 95% $CI=0.38-0.67$, $p=1.40E-6$ (Table 4B). In the independent validation, the prediction error concerning DSS was reduced from 0.097 (calculated only with clinical variables) to 0.092 for the three-year survival and from 0.124 to 0.112 for the four-year survival (Figure 3B). Corresponding to this, the AUC for DSS (Figure 3D) was increased from 0.728 to 0.772 for the three-year survival and from 0.728 to 0.768 for the four-year survival.

As an addition, we computed a dynamic AUC curve (Figure 3B), showing for the three-year (four-year) OS of the patients an increase of AUC from 0.705 (0.704) calculated only with clinical variables to 0.750 (0.743) calculated with clinical variables and the ProMCol classifier.

In addition, we performed a replication analysis, fitting a new Cox model for the clinical variables and our ProMCol classifier for OS (Table 5A) and DSS (Table 5B) on the validation cohort. In this cohort, the ProMCol classifier showed a significant association with OS

(HR=0.61, 95% CI=0.45-0.82 and $p=0.001$) and DSS (HR=0.55, 95% CI=0.38-0.79, $p=0.001$) of patients with non-metastatic disease. For non-DSS, no significant association ($p=0.302$) was found in the validation cohort.

5 Table 4

A: Multivariate Cox regression analysis of the ProMCol classifier with OS in the screening cohort*

output	Model only with clinical covariates			Model with clinical covariates and ProMCol classifier		
	HR	95% CI	<i>p</i> -value	HR	95% CI	<i>p</i> -value
Age (continuous, per year)	1.07	1.05-1.09	4.9E-11	1.08	1.05-1.10	1.2E-11
Smoke (former vs. never)	0.94	0.62-1.42	0.769	0.96	0.64-1.45	0.858
Smoke (current vs. never)	1.13	0.63-2.01	0.685	1.07	0.60-1.92	0.809
Cancer stage (2 vs. 1)	1.45	0.84-2.51	0.181	1.24	0.71-2.15	0.451
Cancer stage (3 vs. 1)	2.60	1.56-4.35	2.6E-04	2.03	1.20-3.42	0.008
Gender (male vs. female)	1.45	0.98-2.15	0.063	1.67	1.12-2.49	0.012
MSI (yes vs. no)	0.86	0.56-1.32	0.482	0.85	0.56-1.30	0.466
Location (colon, dist vs. colon, prox)	0.95	0.62-1.46	0.819	0.97	0.63-1.49	0.878
Location (rectum vs. colon, prox)	0.97	0.63-1.49	0.884	1.06	0.69-1.63	0.776
ProMCol classifier	-	-	-	0.51	0.41-0.63	6.2E-10

B: Multivariate Cox regression analysis of the ProMCol classifier with DSS in the screening cohort

output	Model only with clinical covariates			Model with clinical covariates and ProMCol classifier		
	HR	95% CI	<i>p</i> -value	HR	95% CI	<i>p</i> -value
Age (continuous, per year)	1.05	1.03-1.08	1.1E-04	1.05	1.03-1.08	7.1E-05
Smoke (former vs. never)	0.74	0.43-1.30	0.299	0.76	0.43-1.32	0.323
Smoke (current vs. never)	1.00	0.47-2.11	0.998	0.97	0.46-2.05	0.930
Cancer stage (2 vs. 1)	3.71	1.28-10.77	0.016	3.12	1.07-9.11	0.038
Cancer stage (3 vs. 1)	7.87	2.82-21.92	7.9E-05	6.02	2.14-16.93	6.6E-04
Gender (male vs. female)	1.21	0.73-2.01	0.465	1.40	0.83-2.36	0.201
MSI (yes vs. no)	0.82	0.47-1.45	0.502	0.83	0.48-1.45	0.520
Location (colon, dist vs. colon, prox)	0.74	0.41-1.33	0.312	0.78	0.43-1.42	0.424
Location (rectum vs. colon, prox)	1.00	0.58-1.73	0.999	1.11	0.64-1.91	0.713
ProMCol classifier	-	-	-	0.50	0.38-0.67	1.4E-06

HR=Hazard ratio, CI= Confidence interval; HRs for the ProMCol are based on a change from the lower to the upper quartile of the prognostic classifier.

* This model was used for the evaluation of the prediction error (validation) of the ProMCol classifier (see Figure 2)

Table 5

A: Exploratory replication analysis of the Cox regression analysis for OS in the validation cohort

Model only with clinical covariates				Model with clinical covariates and ProMCol classifier		
output	HR	95% CI	p-value	HR	95% CI	p-value
Age (continuous, per year)	1.07	1.04-1.10	5.1E-06	1.06	1.03-1.09	1.1E-05
Smoke (former vs. never)	0.74	0.40-1.38	0.346	0.70	0.38-1.28	0.247
Smoke (current vs. never)	1.37	0.63-2.99	0.422	1.48	0.68-3.22	0.318
Cancer stage (2 vs. 1)	3.49	1.04-11.74	0.043	2.96	0.87-10.03	0.081
Cancer stage (3 vs. 1)	8.13	2.48-26.65	5.4E-04	6.31	1.90-20.93	0.003
Gender (male vs. female)	0.88	0.50-1.57	0.674	0.89	0.51-1.57	0.698
MSI (yes vs. no)	1.83	0.84-3.98	0.126	1.71	0.78-3.73	0.177
Location (colon, dist vs. colon, prox)	1.24	0.67-2.30	0.496	1.17	0.63-2.18	0.620
Location (rectum vs. colon, prox)	1.16	0.59-2.28	0.662	1.15	0.59-2.24	0.686
ProMCol classifier	-	-	-	0.61	0.45-0.82	0.001

B: Exploratory replication analysis of the Cox regression analysis for DSS in the validation cohort

Model only with clinical covariates				Model with clinical covariates and ProMCol classifier		
output	HR	95% CI	p-value	HR	95% CI	p-value
Age (continuous, per year)	1.05	1.02-1.09	0.001	1.05	1.02-1.09	0.002
Smoke (former vs. never)	0.90	0.42-1.90	0.779	0.80	0.38-1.68	0.550
Smoke (current vs. never)	1.80	0.75-4.30	0.189	1.97	0.82-4.72	0.128
Cancer stage (2 vs. 1)	7.11	0.93-54.37	0.059	5.81	0.75-44.71	0.091
Cancer stage (3 vs. 1)	20.36	2.75-150.75	0.003	14.92	1.99-111.66	0.008
Gender (male vs. female)	0.74	0.37-1.47	0.388	0.79	0.41-1.54	0.497
MSI (yes vs. no)	1.93	0.71-5.24	0.198	1.86	0.68-5.12	0.230
Location (colon, dist vs. colon, prox)	1.78	0.80-3.96	0.160	1.67	0.75-3.75	0.210
Location (rectum vs. colon, prox)	1.83	0.78-4.29	0.164	1.80	0.77-4.20	0.175
ProMCol classifier	-	-	-	0.55	0.38-0.79	0.001

HR=Hazard ratio, CI= Confidence interval; HRs for the ProMCol are based on a change from the lower to the upper quartile of the prognostic classifier.

5 List of non-standard references cited in this specification:

Benson et al.; J Clin Oncol 2004;22(16):3408-19

Bjaanaes et al.; Mol Oncol 2016;10(2):330-43

Boland & Goel; N Engl J Med 2016;374(3):277-8

10 Brenner et al.; Ann Intern Med 2011;154(1):22-30

Cicek et al.; Hum Mol Genet 2013;22(15):3038-47

Egger et al.; Nature 2004;429(6990):457-63

Esteller; Lancet Oncol 2003;4(6):351-8

Feinberg & Tycko; Nat Rev Cancer 2004;4(2):143-53

15 Ferlay et al., International journal of cancer Journal international du cancer

2015;136(5):E359-86

Gerds & Schumacher (Biometrical Journal 2006;48(6):1029–40

Hoffmeister et al.; Journal of the National Cancer Institute 2015;107(6):djv045

Ignatiadis et al.; Nat Methods 2016;13(7):577-80

5 Issa; Nat Rev Cancer 2004;4(12):988-93

Jia et al.; Clinical epigenetics 2016;8:25

Li et al.; J Am Stat Assoc 2012;107(499):1129-39

Lilla et al.; Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive
10 Oncology 2006;15(1):99-107

Luo et al. (2014), BioMed Research International, Vol 2014, Art. ID 784706

Nagtegaal et al.; Nat Rev Clin Oncol 2011;9(2):119-23

O'Connor et al., J Clin Oncol 2011;29(25):3381-8

Okugawa et al.; Gastroenterology 2015;149(5):1204-25 e12

15 Sandoval et al.; J Clin Oncol 2013;31(32):4140-7

Siegel et al.; CA Cancer J Clin 2012;62(4):220-41

Sofer et al. (2013), Bioinformatics 29(22): 2884

Song & Zhou; Statistica Sinica 2008(18):947-65

Stefansson et al.; Mol Oncol 2015;9(3):555-68

20 Sturm et al.; Cancer cell 2012;22(4):425-37

Tang et al.; Clinical epigenetics 2016;8:115

Tang et al.; Oncotarget 2016;7(39):64191-202

Yang et al. (2015), Int J Cancer 136: 1845

Wei et al.; Nat Commun 2015;6:8699

25

Claims

1. A method for determining a survival probability of a subject suffering from colorectal cancer comprising
 - 5 a) detecting the methylation status of at least two CpG sites related to at least two CpG sites of Table 1 in a sample of said subject and,
 - b) based on the methylation status detected in step a), determining the survival probability of said subject.
- 10 2. The method of claim 1, wherein said at least two CpG sites are at least two CpG sites of Table 1, preferably selected from the list consisting of cg16336556, cg14270346, cg05646575, cg17431888, cg12510999, cg00832644, cg11056055, cg08804626, cg14983135, cg22522598, cg19184885, cg08729279, and cg10758824.
3. The method of claim 1 or 2, wherein the methylation status of at least eight, preferably at least 13 of said CpG sites is detected.
- 15 4. The method of any one of claims 1 to 3, wherein said detecting the methylation status of a CpG site is detecting the average degree of methylation of said site from at least 10, preferably at least 100 cells.
5. The method of any one of claims 1 to 4, wherein a decreased average degree of methylation compared to a reference is indicative of an unfavorable survival probability.
- 20 6. The method of any one of claims 1 to 5, wherein said method comprises further step a1) comparing the methylation status of said at least two CpG sites of step a) to a reference or to references; and wherein in step b) the determining is based on the comparison of step a1).
- 25 7. The method of claim 5 or 6 wherein said reference is obtained from a population of subjects suffering from colorectal cancer, preferably non-metastasizing colorectal cancer.
8. The method of any one of claims 1 to 7, wherein said sample is a tissue sample, preferably is a sample of tumor cells.
- 30 9. The method of any one of claims 1 to 8, wherein said subject is a human
10. A method for patient monitoring comprising the steps of the method according to any one of claims 1 to 9 and providing close monitoring and/or lifestyle recommendations in case an unfavorable survival probability and/or an increased mortality risk is detected.
- 35 11. A data collection, preferably comprised on a data carrier, comprising the positions of at least two, preferably at least eight CpG sites selected from Table 1 and/or CpG sites related thereto.

12. The data collection of claim 11, further comprising reference values or reference ranges for the methylation status of said CpG sites.
- 5 13. A kit comprising means for determining the methylation status of at least two CpG sites related to at least two CpG sites of Table 1, preferably selected from the list consisting of cg16336556, cg14270346, cg05646575, cg17431888, cg12510999, cg00832644, cg11056055, cg08804626, cg14983135, cg22522598, cg19184885, cg08729279, and cg10758824 in a sample of a subject suffering from colorectal cancer, and a data collection according to claim 11 or 12.
- 10 14. A device comprising an analysis unit comprising means for determining the methylation status of at least two CpG sites related to at least two CpG sites of Table 1 in a sample of a subject suffering from colorectal cancer, and an evaluation unit comprising a data collection according to claim 11 or 12.
- 15 15. Use of means for determining the methylation status of at least two CpG sites related to at least two CpG sites of Table 1 and a data collection according to claim 11 or 12 for manufacturing a diagnostic means or device for determining a survival probability of a subject suffering from colorectal cancer.

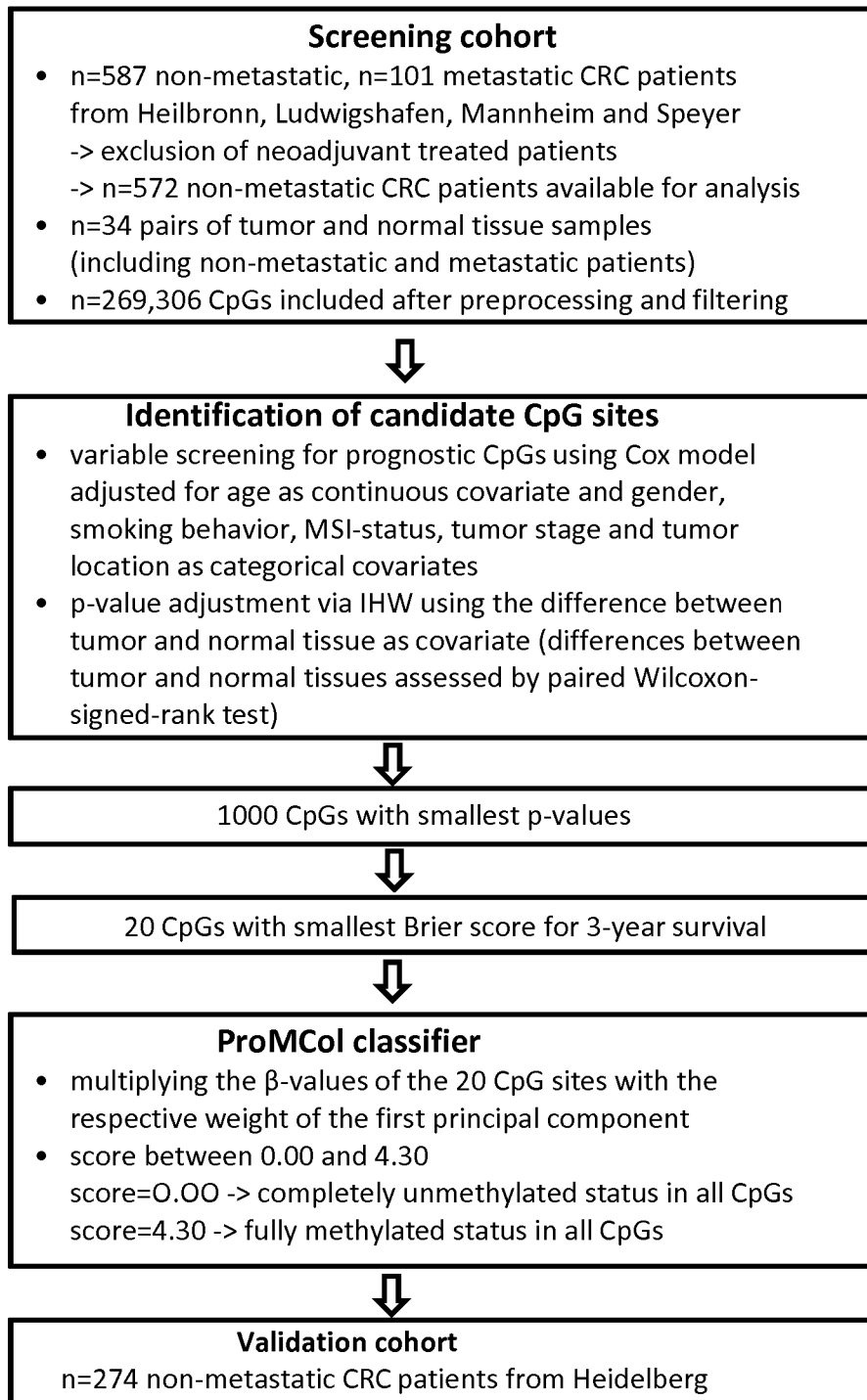


Fig. 1

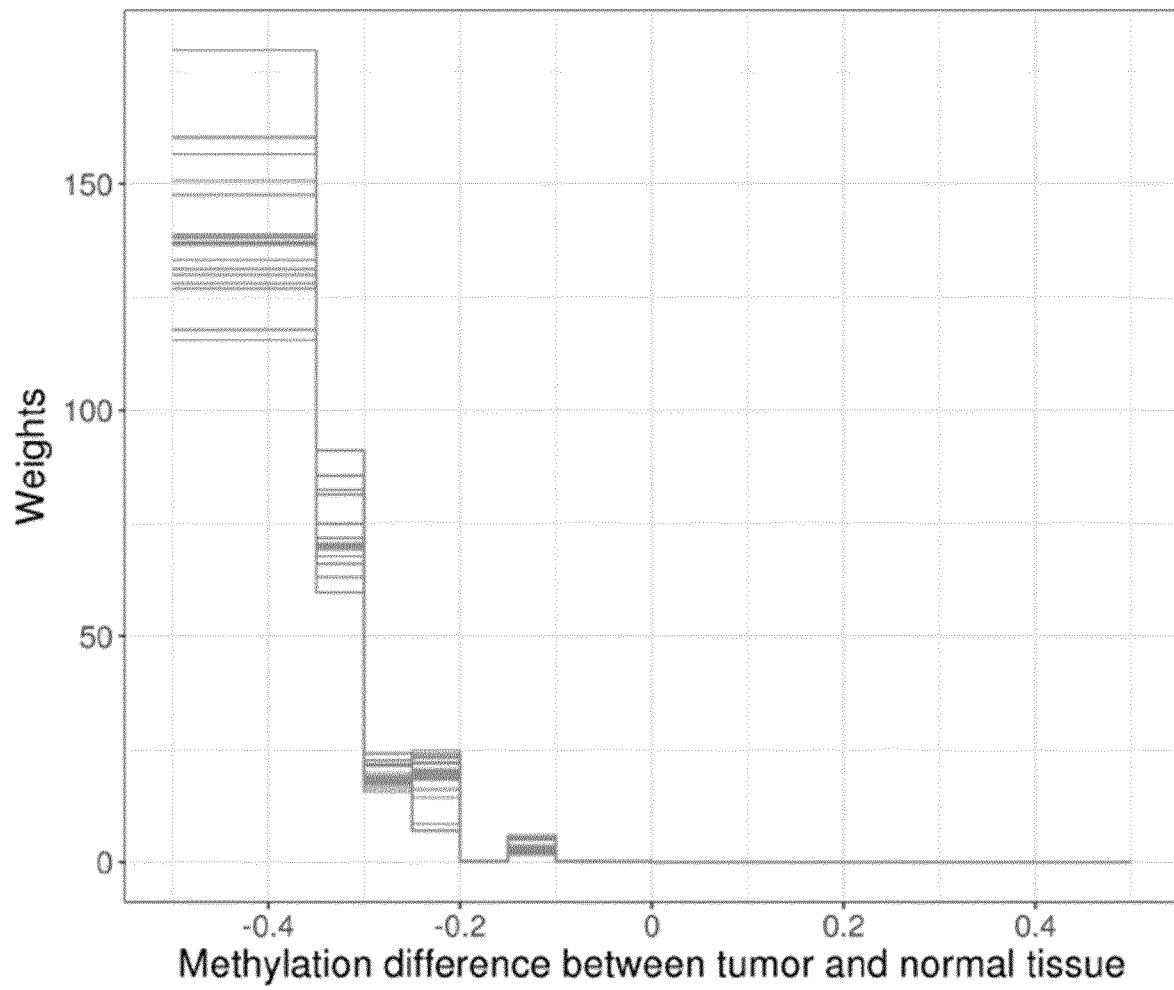


Fig. 2

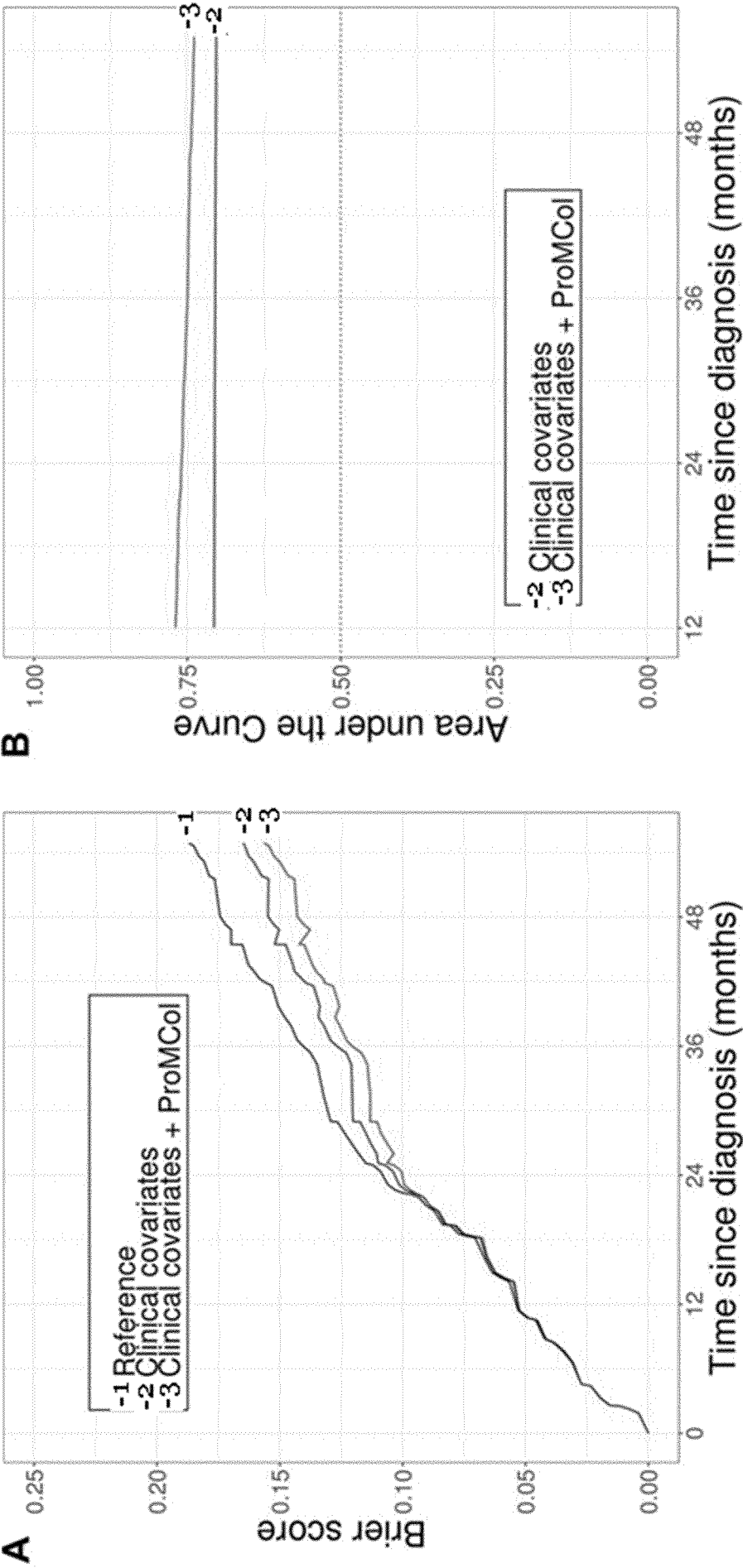


Fig. 3

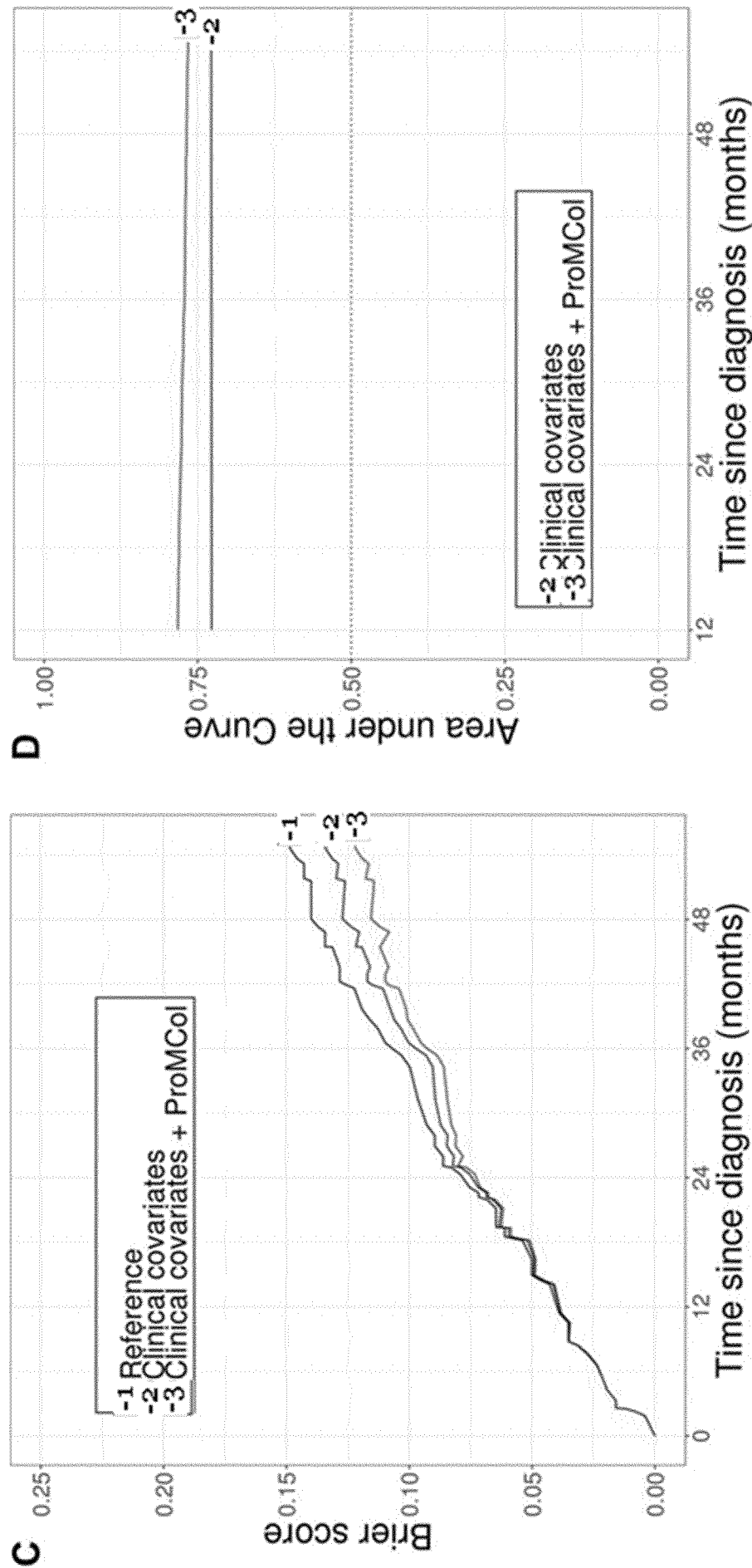


Fig. 3 (continued)

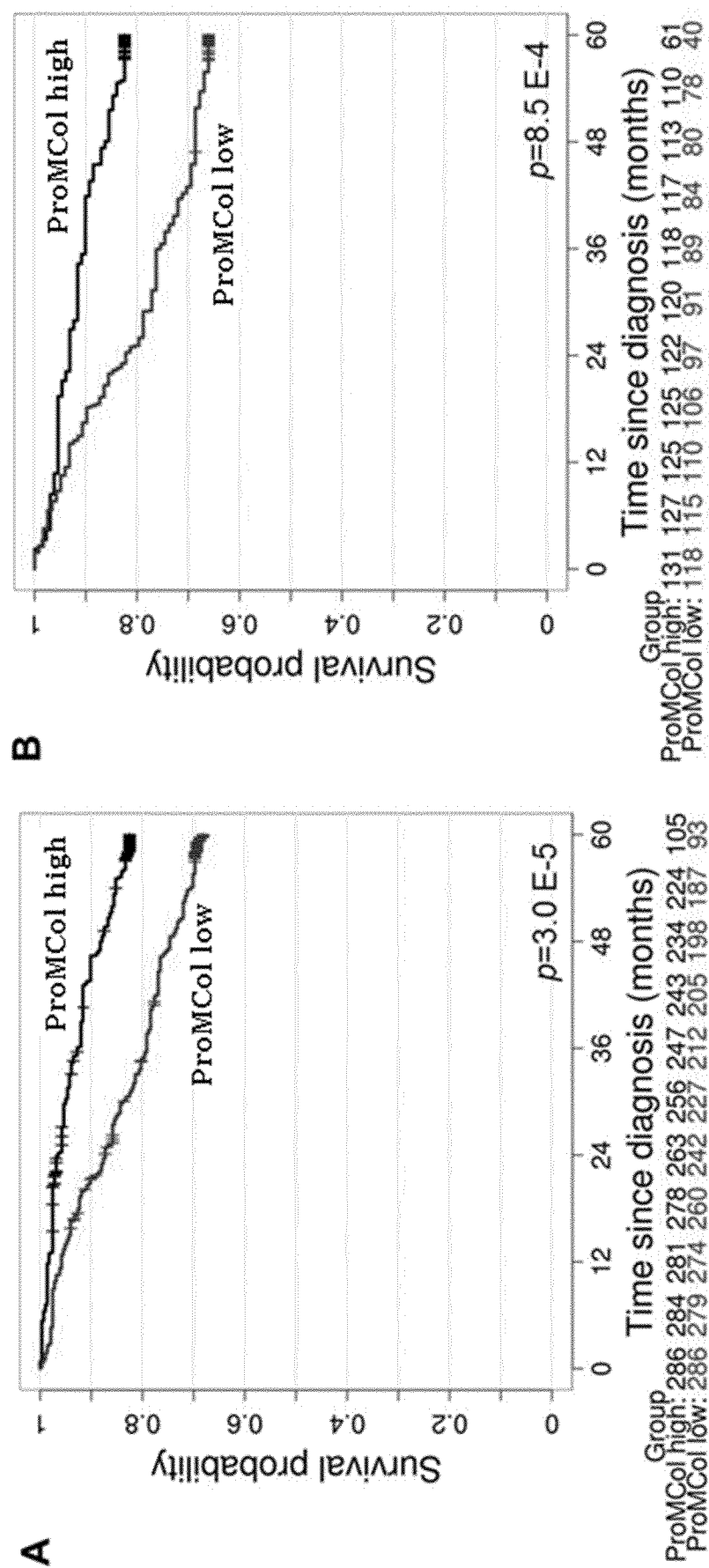


Fig. 4

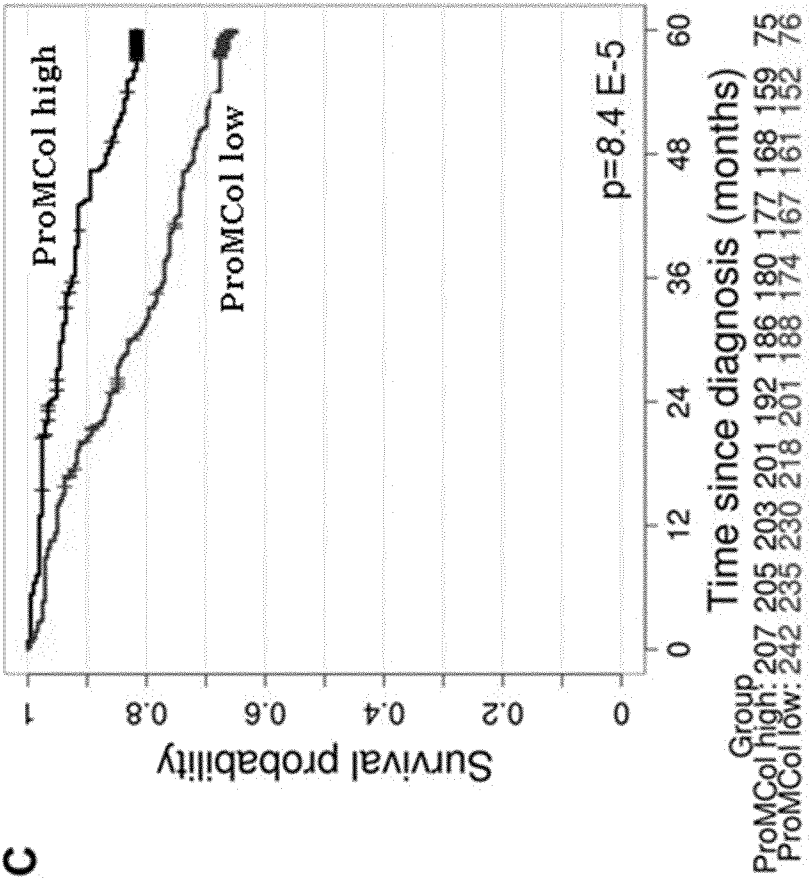
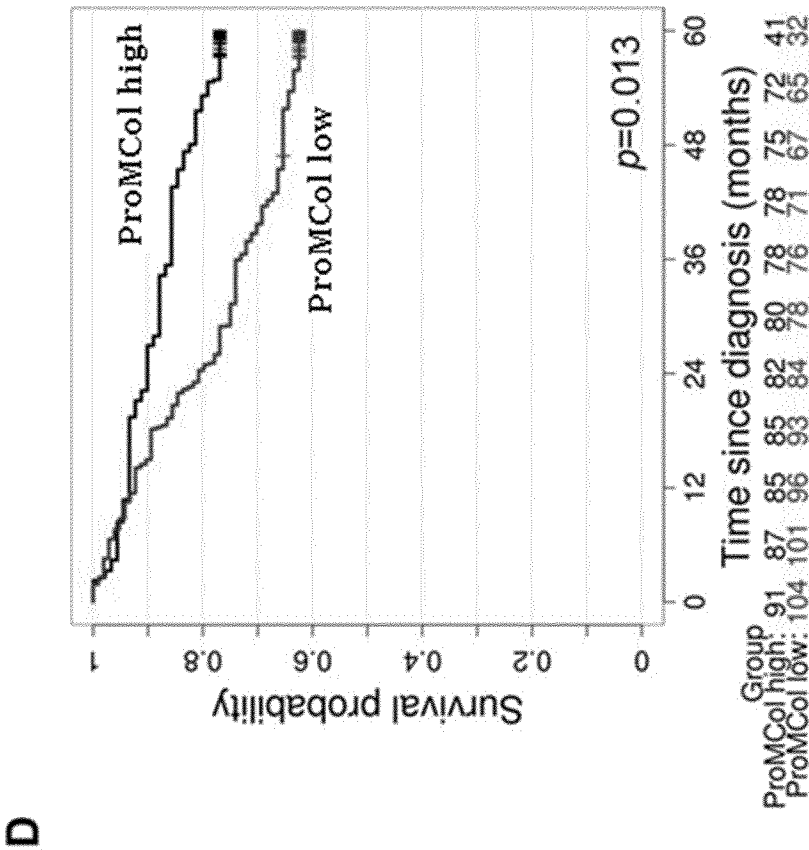


Fig. 4 (continued)

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2018/078340

A. CLASSIFICATION OF SUBJECT MATTER

INV. C12Q1/6886

ADD.

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

C12Q

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

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

EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	XIAOKE HAO ET AL: "DNA methylation markers for diagnosis and prognosis of common cancers", PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA, vol. 114, no. 28, 26 June 2017 (2017-06-26), pages 7414-7419, XP055450701, US ISSN: 0027-8424, DOI: 10.1073/pnas.1703577114 -----	1-15
A	US 2013/345144 A1 (HOON DAVE S B [US] ET AL) 26 December 2013 (2013-12-26) ----- -/-	1-15



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

12 December 2018

Date of mailing of the international search report

25/02/2019

Name and mailing address of the ISA/

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

Authorized officer

Cornelis, Karen

INTERNATIONAL SEARCH REPORT

International application No.
PCT/EP2018/078340

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-15(partially)

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2018/078340

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	M. WALLNER ET AL: "Methylation of Serum DNA Is an Independent Prognostic Marker in Colorectal Cancer", CLINICAL CANCER RESEARCH, vol. 12, no. 24, 15 December 2006 (2006-12-15), pages 7347-7352, XP055532806, US ISSN: 1078-0432, DOI: 10.1158/1078-0432.CCR-06-1264 -----	1-15
A	WO 2006/050573 A1 (GARVAN INST MED RES [AU]; MAS JORDI FRIGOLA [ES]; PEINADO MIGUEL A [ES] 18 May 2006 (2006-05-18) -----	1-15
X	MARINA BIBIKOVA ET AL: "High density DNA methylation array with single CpG site resolution", GENOMICS, ACADEMIC PRESS, SAN DIEGO, US, vol. 98, no. 4, 26 July 2011 (2011-07-26), pages 288-295, XP028304083, ISSN: 0888-7543, DOI: 10.1016/J.YGENO.2011.07.007 [retrieved on 2011-08-02] the whole document -----	11-14
X	"Data Sheet: Epigenetics; Infinium HumanMethylation450 BeadChip", 11 October 2012 (2012-10-11), XP055090635, Retrieved from the Internet: URL: http://res.illumina.com/documents/products/datasheets/datasheet_humanmethylation450.pdf [retrieved on 2013-11-28] the whole document -----	11-14
X	WO 2013/007702 A1 (EPIGENOMICS AG [DE]; LEWIN JOERN [DE]; KRISPIN MANUEL [DE]) 17 January 2013 (2013-01-17) the whole document -----	1-10,15

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2018/078340

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2013345144 A1	26-12-2013	EP 1869212 A2	26-12-2007
		EP 2272984 A1	12-01-2011
		US 2007020646 A1	25-01-2007
		US 2013345144 A1	26-12-2013
		US 2016122826 A1	05-05-2016
		WO 2006084051 A2	10-08-2006

WO 2006050573 A1	18-05-2006	CA 2587817 A1	18-05-2006
		EP 1824993 A1	29-08-2007
		US 2009042184 A1	12-02-2009
		WO 2006050573 A1	18-05-2006

WO 2013007702 A1	17-01-2013	AU 2012282528 A1	23-01-2014
		BR 112014000443 A2	14-02-2017
		CA 2840149 A1	17-01-2013
		CN 103732759 A	16-04-2014
		CN 108048573 A	18-05-2018
		DK 2729579 T3	08-01-2018
		EA 201400117 A1	30-06-2014
		EP 2729579 A1	14-05-2014
		ES 2654561 T3	14-02-2018
		HR P20180001 T1	09-03-2018
		HU E036036 T2	28-06-2018
		IL 230303 A	31-05-2018
		JP 6397762 B2	26-09-2018
		JP 2014520520 A	25-08-2014
		KR 20140044385 A	14-04-2014
		LT 2729579 T	26-02-2018
		MX 358117 B	06-08-2018
		SI 2729579 T1	31-07-2018
		US 2015031021 A1	29-01-2015
		WO 2013007702 A1	17-01-2013

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-15(partially)

A method to determine survival probability of a subject suffering from colorectal cancer by determining the methylation status of at least 2 CpG sites related to Table 1, whereby one of the CpG sites is cg16336556, a data collection with the positions of these CpG sites, a kit and device for determining the methylation status of these sites.

2-20. claims: 1-15(partially)

A method to determine survival probability of a subject suffering from colorectal cancer by determining the methylation status of at least 2 CpG sites related to Table 1, whereby one of the CpG sites is one of the CpGs of Table 1, a data collection with the positions of these CpG sites, a kit and device for determining the methylation status of these sites.
