



(86) Date de dépôt PCT/PCT Filing Date: 2019/01/25  
(87) Date publication PCT/PCT Publication Date: 2019/08/08  
(45) Date de délivrance/Issue Date: 2022/09/06  
(85) Entrée phase nationale/National Entry: 2020/07/10  
(86) N° demande PCT/PCT Application No.: EP 2019/051868  
(87) N° publication PCT/PCT Publication No.: 2019/149633  
(30) Priorité/Priority: 2018/01/30 (EP18154190.5)

(51) Cl.Int./Int.Cl. *G01N 33/574* (2006.01)  
(72) Inventeurs/Inventors:  
BUND, TIMO, DE;  
ZUR HAUSEN, HARALD, DE;  
DE VILLIERS, ETHEL-MICHELE, DE;  
TESSMER, CLAUDIA, DE;  
HEIKENWALDER, MATHIAS, DE;  
WEBER, ACHIM, CH;  
BURK-KORNER, AMELIE, DE  
(73) Propriétaire/Owner:  
DEUTSCHES KREBSFORSCHUNGSZENTRUM  
STIFTUNG DES OFFENTLICHEN RECHTS, DE  
(74) Agent: BENOIT & COTE INC.

(54) Titre : UTILISATION DE LA PROTEINE REP BMMF1 EN TANT QUE BIOMARQUEUR POUR LE CANCER  
COLORECTAL  
(54) Title: USE OF BMMF1 REP PROTEIN AS BIOMARKER FOR COLORECTAL CANCER

(57) Abrégé/Abstract:

The present invention relates to the use of BMMF Rep-protein as biomarker for colon cancer.

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

(19) World Intellectual Property

Organization

International Bureau

(43) International Publication Date

08 August 2019 (08.08.2019)



(10) International Publication Number

WO 2019/149633 A1

(51) International Patent Classification:

G01N 33/574 (2006.01)

Published:

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

— with sequence listing part of description (Rule 5.2(a))

(21) International Application Number:

PCT/EP2019/051868

(22) International Filing Date:

25 January 2019 (25.01.2019)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

18154190.5 30 January 2018 (30.01.2018) EP

(71) Applicant:

**DEUTSCHES KREBS-  
FORSCHUNGSZENTRUM STIFTUNG DES ÖF-  
FENTLICHEN RECHTS** [DE/DE]; Im Neuenheimer  
Feld 280, 69120 Heidelberg (DE).

(72) **Inventors:** **BUND, Timo**; Buchenweg 3, 69221 Dossen-  
heim (DE). **ZUR HAUSEN, Harald**; Eichenstraße 1,  
69483 Wald-Michelbach (DE). **DE VILLIERS, Ethel-  
Michele**; Eichenstraße 1, 69483 Wald-Michelbach (DE).  
**TESSMER, Claudia**; Höhenstraße 23, 74869 Schwarzach  
(DE). **HEIKENWÄLDER, Mathias**; Mannheimer Straße  
15, 69115 Heidelberg (DE). **WEBER, Achim**; Univer-  
sitätsspital Zürich, Hadlaubstraße 45, 8006 Zürich (CH).  
**BURK-KÖRNER, Amelie**; Römerstraße 208, 69126 Hei-  
delberg (DE).

(74) **Agent:** **ANDREA SCHÜSSLER**; Truderinger Strasse 246,  
81825 Munich (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,  
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).

(54) **Title:** USE OF BMMF1 REP PROTEIN AS BIOMARKER FOR COLORECTAL CANCER(57) **Abstract:** The present invention relates to the use of BMMF Rep-protein as biomarker for colon cancer.

WO 2019/149633 A1

## USE OF BMMF1 REP PROTEIN AS BIOMARKER FOR COLORECTAL CANCER

The invention relates to the use of a DNA-replication-associated (Rep) protein as a biomarker for colorectal cancer (CRC).

### 5 Background of the Invention

Thousands of people around the world have been diagnosed with colon cancer, many of them ultimately dying of the disease. Patients are typically treated with colon resection surgery, followed by radiation therapy or systemic chemotherapy, the therapy being based on macroscopic traits of the tumor and the tumor stage. The 5-  
10 year relapse-free survival rate is improved in some patients receiving chemotherapy after colon surgical resection surgery, while this statistic is not improved in others.

Colorectal cancer (CRC) is the second leading cause of cancer death in the US. Most CRC patients are diagnosed late due to lack of predictive biomarkers and poor screening rate attributable to the inconvenience of current screening methods. To  
15 enhance earlier detection, there is a need for biomarkers that will facilitate early detection and further insights into the pathogenesis of CRC.

CRC is a cancer that evolved as a consequence of uncontrolled cell growth in the colon or rectum. These malignancies may develop as a consequence of pre-existing benign adenomas where genetic alterations promote the transition from normal to  
20 cancerous growth. Epigenetic events have been recognized as an important mechanism for this transition.

A blood-based protein biomarker panel for the detection of colorectal cancer as described by Fung et al., 2015 has not yet brought the desired results as regards predictability. Thus, unfortunately, there does not exist a biomarker for CRC which  
25 may be used both for prognostic and diagnostic purposes.

## Description of the Present Invention

In the present application the inventors have created a model for colon cancer development that is shown in Fig. 1.

The inventors found that the uptake of BMMF (Bovine Meat and Milk Factor) agents within the first months of life either by substitution of breast-feeding during weaning by cow milk products or by the uptake of dairy or beef products, in general, leads to the early infection of newborns with BMMF antigens in the colon. Based on the decline of maternal antibodies and the frequently observed weakness of the immune system often coupled with induction of immune tolerances of the newborn during this very early period of life, these agents might either directly escape colonic immune response or a situation of immune tolerance against these agents might be induced. Within the next years or decades – depending on the immune system of the host – more and more BMMF antigens accumulate within the colonic *lamina propria*. This accumulation may be triggered also by the uptake of specific molecules that may represent receptors for BMMFs. These molecules are also taken up by consumption of cow products and are metabolized into receptors on the surface of the host cells. When a certain level of antigen is reached by continuous uptake of BMMFs in combination with focal spreading of infection, the host immune response induces a state of chronic and local inflammation producing a stable increase of reactive oxygen species (ROS) and cyclooxygenase-2 (Cox-2) which dramatically increases the probability of deregulated cell proliferation with concomitant fixation of random mutations in surrounding cells induced by ROS.

Available results based on Ki67 (Ki67 = proliferation marker) IHC staining show that especially the basal epithelial cells derived from epithelial stem cells located close to the endpoint of the crypts and migrating towards the distal luminal part of the crypts feature an intrinsically high proliferation enabling stochastic manifestation of mutations as a basic requirement for tumorigenesis and development of colon cancer (Fig. 2). Thus, BMMFs represent a specific and local trigger for induction of chronic inflammation within the lamina propria leading to an increase of ROS which induces proliferation and mutation in surrounding epithelial cells eventually leading to the formation of polyps as precursors for colon cancer.

In detail, a selection of 11 pairs of donor tissue samples (each tumor tissue and peripheral tissue) and 4 samples from different donors of colon polyp tissue were subjected to IHC staining with mouse monoclonal anti-Rep antibodies. With regard to the 11 tumor and peripheral samples, 9 out of 11 tissues showed strong specific antibody staining with at least two anti-Rep antibodies. Exemplarily, the staining with an anti-Rep antibody (e.g. antibody 10-3) shows specific detection of protein targets in tissue samples 4798, 4799, 4802, 4806, 4809 and 4813 in the tumor as well as in the peripheral tissue. In general, the staining intensity is higher for peripheral tissue, where intensely stained cytoplasmic aggregates of smaller size concentrate within distinct islands of cells within the *lamina propria* between the characteristic Crypts of Lieberkühn of the colon tissue (all pictures showing traversal cuts throughout the Crypts of Lieberkühn with the (inner) lumen of the crypts having contact to the colonic fluids in the center of the ring-like arrangement of epithelial cells of the crypts; Fig. 2). Staining of tumor tissues also shows intense staining of smaller sized aggregates mainly within the cytoplasmic regions of the stroma cells within the tumor tissue. The same aggregate-like structures are also visible in the *lamina propria* of colon polyps by the use of four different anti-Rep antibodies for staining of sequential cuts. Both, for colon peripheral tissue and for colon polyps, the regions with highest Rep-specific antibody detection correlate with regions with highest detection levels for CD68 positive cells pointing towards a localization of the Rep-specific antigens in inflammatory islands, i.e. regions with especially high levels of inflammatory monocytes, circulating macrophages, or tissue macrophages.

As additionally shown in the present application, the comparison of Rep IRS scores (IRS, immunoreactive score) of peritumoral and tumoral tissues confirm a significantly increased Rep detection in the peritumoral tissue of the analyzed CRC patients.

The inventors concluded that for cancer induction the amount of BMMF antigen within the *lamina propria* between the colon crypts is important for the induction of chronic local inflammation, induction of diffusing ROS/NOS which finally induces random mutations in dividing stem and daughter cells of neighboring crypts which might turn into early polyps/tumor precursor cells by manifestation of key mutations. Thus, a correlation of peritumoral BMMF levels and intensity of CD68-positive

macrophages should be observable. Indeed, by testing IRS values for Rep-detection against IRS-values for CD68-detection within the corresponding CRC patients, a significant correlation of Rep and CD68 detection is observed which shows that higher detection levels of Rep correlate with higher levels of CD68 detection.

- 5 To test a prognostic score of Rep detection in peritumoral regions of CRC patients on patient survival time, Kaplan Meier curves were calculated based on patient data which were grouped into "Rep low" and "Rep high". High levels of Rep in CRC patients correlate with a significant reduction of patient overall survival time, c.f. Fig. 13 and Fig. 14. The experiments show that increased Rep levels correlate with a
- 10 hazard ratio of 4.7 indicating a negative correlation of Rep detection and patient survival and underlining the prognostic score of Rep detection in CRC tissue. Patients scored "Rep low" show a 5- and 10-year survival probability of 92%, whereas patients with "Rep high" show significantly decreased survival probabilities of 74% after 5 years and 65% after 10 years, representing a decrease of 20 to 30%.
- 15 The inventors also tested the antibody level of colon cancer patients by contacting the Rep protein with a serum specimen suspected of containing anti-Rep protein antibodies under conditions that permit the Rep protein to bind to any such antibody present in the specimen. The inventors recognized that the serum reactivity is reduced in colon cancer patients in comparison to healthy controls.
- 20 So far a spectrum of 18 different, but partially related, DNA molecules were isolated from different test material (bovine sera, milk, brain tissue of one multiple sclerosis patient autopsies) (Funk, Gunst et al. 2014, Gunst, zur Hausen et al. 2014, Lamberto, Gunst et al. 2014, Whitley, Gunst et al. 2014; WO 2015/062726 A2; WO 2016/005054 A2). The 18 isolates were divided into four different groups BMMF1 through
- 25 BMMF4, according to their molecular characteristics (zur Hausen et al., 2017). Three of these groups revealed a remarkable degree of similarity to *Acinetobacter baumannii* and *Psychrobacter* plasmids. The fourth group comprised 3 isolates being representatives of *Gemycircularviridae*. Putative Rep genes were identified as part of the BMMF's DNA sequences obtained by *in silico* comparisons to available
- 30 sequences. Amplification using abutting primers in the rep gene led to the isolation of full and partial circular DNA genomes from bovine sera (Funk et al., 2014). This

was extended to samples from commercially available milk products for the presence of specific circular single-stranded DNA genomes. Full-length circular single-stranded DNA molecules of 14 different isolates of (~1100 to 3000 nucleotides) were cloned and sequenced (Whitley et al., 2014; Gunst et al., 2014; Funk et al., 2014; Lamberto et al., 2014). Four additional isolates were obtained from human brain and serum (all from patients with multiple sclerosis) (Whitley et al., 2014; Gunst et al., 2014; Lamberto et al., 2014).

Among these isolates two DNA molecules closely related to transmissible spongiform encephalopathy (TSE)-associated isolate Sphinx 1.76 (1,758 bp; accession no. HQ444404, (Manuelidis L. 2011)) were isolated from brain tissue from an MS patient. These isolates were MSBI1.176 (MSBI, multiple sclerosis brain isolate) (1,766 bp) and MSBI2.176 (1,766 bp) which are designated as "MSBI1 genome" and "MSBI2 genome", respectively. MSBI1.176 shares 98% nucleotide similarity to the sequence of Sphinx 1.76. The large open reading frames (ORFs) of the isolates encode a putative DNA replication protein sharing high similarity between them. Another common feature is the presence of iteron-like tandem repeats. The alignment of this repeat region indicates a variation in the core of single nucleotides. This iteron-like repeats may constitute the binding sites for Rep proteins. The sequences of the isolates have been deposited in the EMBL Databank under accession numbers LK931491 (MSBI1.176) and LK931492 (MSBI2.176) (Whitley C. et al. 2014) and have been aligned and described in WO 2016/005054 A2.

Further isolates were obtained from cow milk. These Cow milk isolates (CMI) were CMI1.252, CMI2.214 and CMI3.168 which are designated as "CMI1 genome", "CMI2 genome" and "CMI3 genome", respectively. The sequences of the isolates have been deposited in the EMBL Databank under accession numbers LK931487 (CMI1.252), LK931488 (CMI2.214) and LK931489 (CMI3.168) and have been aligned and described in WO 2016/005054 A2.

The present inventors have found that both CMI genomes and MSBI genomes show a significant production of transcribed RNA and the encoded Rep protein is expressed in peripheral tissue around the colon cancer islands and polyps. The present inventors have found that the encoded Rep proteins (MSBI1 Rep, MSBI2

Rep, CMI1 Rep, CMI2 Rep, CMI3 Rep) represent a biomarker for colon cancer. As DNA-replication-associated protein (RepB) the Rep protein has DNA binding activity and can be essential for initiation of replication of episomal or viral DNA molecules. Rep proteins show a marked potential of self-oligomerization and aggregation, which  
 5 was described within prokaryotic systems *in vivo* and *in vitro* (Giraldo, Moreno-Diaz de la Espina et al. 2011, Torreira, Moreno-Del Alamo et al. 2015).

The inventors have raised monoclonal antibodies against Rep protein. In particular embodiments the anti-Rep antibodies bind to epitopes of Rep protein that are exemplified in Fig. 3. Particular preferred antibodies bind to epitopes within an amino  
 10 acid sequence selected from the group consisting of amino acids from 1 to 136, from 137 to 229 and from 230 to 324 of SEQ ID NO:1. For example, the antibody binds to an epitope comprised by SEQ ID NO:2 or SEQ ID NO:3.

Brief description of the drawings:

- 15 Figure 1 shows the proposed model for colon cancer development
- Figure 2 shows results based on Ki67 IHC staining (available online data). It may be seen that especially the basal epithelial cells derived from epithelial stem cells located close to the endpoint of the crypts and migrating towards the distal luminal part of the crypts feature an intrinsically high  
 20 proliferation enabling stochastic manifestation of mutations as a basic requirement for tumorigenesis and development of colon cancer
- Figure 3 shows characteristics of the raised antibodies and the localization of epitopes within Rep
- Figure 4 shows IHC data obtained by employment of antibody 10-3 (Group C  
 25 anti-Rep antibody) to colon cancer tissue and surrounding non-cancerous peripheral tissue; designations of patients: #4799, #4809, #4798, #4802, #4813 and #4806
- Figure 5 shows IHC data obtained by employment of anti-Rep antibodies 3-6 and 10-3 (Group C anti-Rep antibodies) and commercially available anti-

CD68 antibodies to peripheral tissue surrounding the colon tumour of patient #4809

- Figure 6 shows IHC data obtained by employment of anti-Rep antibodies Group A and C antibodies (antibodies 1-5, 3-6, 10-3 and 11-5) and commercially available anti-CD68 antibodies to peripheral tissue surrounding colon benign polyps
- Figure 7 Distribution of IRS values for immunohistochemical staining of tumoral tissue region of CRC patients with an anti-Rep antibody
- Figure 8 Distribution of IRS values for immunohistochemical staining of peritumoral tissue region of CRC patients with an anti-Rep antibody
- Figure 9 Distribution of IRS values for immunohistochemical staining of tumoral tissue region of CRC patients with an anti-CD68 antibody
- Figure 10 Distribution of IRS values for immunohistochemical staining of peritumoral tissue region of CRC patients with an anti-CD68 antibody
- Figure 11 Representation of IRS values for peritumoral and tumoral tissues of CRC patients
- Figure 12 IRS values of Rep detection are plotted against CD68 positivity, which is grouped into 7 levels (0-6)
- Figure 13 Kaplan-Meier representation of survival of CRC patients grouped into "Rep low" (IRS values 0-1.9) and "Rep high" (IRS values 2-6)
- Figure 14 Representation of the prognostic Hazard Ratio of Rep detection in peritumoral CRC patient tissues based on Cox-correlation of Kaplan-Meier survival curves of CRC patients grouped into "Rep low" (IRS values 0-1.9) and "Rep high" (IRS values 2-6)
- Figure 15 Kaplan-Meier representation of survival of CRC patients grouped into "CD68 low" (IRS values 0-2.9) and "CD68 high" (IRS values 3-6)
- Figure 16 Immunodetection of BMMF-Rep antigens in tissue lysate of an exemplary CRC patient (peritumoral and tumoral) as well as of four

polyps of individual donors after immunostaining with anti-BMMF-Rep antibody 5-2. 100 ng affinity-purified MSBI1.176 Rep protein was loaded as positive control

- 5 The invention provides the teaching that Rep proteins may represent biomarkers for an enhanced risk to develop colon cancer and are useful as a marker for determining the overall survival prognosis of CRC patients.

The terms "colon cancer" and "colorectal cancer" (CRC) are used interchangeably. CRC is a cancer that evolved as a consequence of uncontrolled cell growth in the  
 10 colon or rectum. These malignancies may develop as a consequence of pre-existing benign adenomas where genetic alterations promote the transition from normal to cancerous growth. The terms "colon cancer" or CRC mean pre-stages, early stages or late stages of the disease and metastases derived therefrom.

In an alternative embodiment the present invention may also encompass the  
 15 systematic testing of healthy colonic tissue (tissue from individuals without cancer diagnosis or a specific hint for the disease) to assess the disease risk in the future. This means that the present invention is also suitable to determine the predisposition for developing CRC.

"Rep protein" as used herein refers to a DNA-replication-associated protein (RepB).  
 20 The Rep protein comprises DNA binding activity and could be essential for initiation of replication of episomal/viral DNA molecules. In general Rep protein refers to a Rep protein from the group of the Small Sphinx Genome (Whitley et al., 2014). In particular, the Rep protein is a MSBI1 genome-encoded Rep protein (MSBI1 Rep), a MSBI2 genome-encoded Rep protein (MSBI2 Rep), a CMI1 genome-encoded Rep  
 25 protein (CMI1 Rep), a CMI2 genome-encoded Rep protein (CMI2 Rep) or CMI3 genome-encoded Rep protein (CMI3 Rep). Preferably, the MSBI1 Rep protein is encoded by MSBI1.176 deposited in the EMBL databank under the acc. no. LK931491 and has the amino acid sequence as depicted in SEQ ID NO:1 or the Rep protein is MSBI2 encoded by MSBI2.176 deposited in the EMBL databank under the  
 30 acc. no. LK931492 and has the amino acid sequence as depicted in SEQ ID NO:8

(Whitley, Gunst et al. 2014). In another preferred embodiment the CMI1 Rep protein is encoded by CMI1.252 deposited in the EMBL databank under the acc. no. LK931487 and has the amino acid sequence as depicted in SEQ ID NO:10. In another preferred embodiment the CMI2 Rep protein is encoded by CMI2.214  
 5 deposited in the EMBL databank under the acc. no. LK931488 and has the amino acid sequence as depicted in SEQ ID NO:11. In another preferred embodiment the CMI3 Rep protein is encoded by CMI3.168 deposited in the EMBL databank under the acc. no. LK931489 and has the amino acid sequence as depicted in SEQ ID NO:12. In a particular preferred embodiment the Rep protein comprises a N-terminal  
 10 region conserved among BMMF1 genomes consisting essentially of amino acids from 1 to 229 of SEQ ID NO:1 and a C-terminal variable region specific for MSBI1.176 consisting essentially from amino acids 230 to 324 of SEQ ID NO:1. The N-terminal conserved region comprises a putative, first DNA binding domain consisting essentially of amino acids from 1 to 136 of SEQ ID NO: 1 and a second  
 15 putative DNA binding domain consisting essentially of amino acids from 137 to 229 of SEQ ID NO:1. The C-terminal domain shows little sequence homology with any known protein and consists of amino acids 230 to 324.

“Rep protein” also encompasses fragments and variants of the protein with SEQ ID NO:1 or SEQ ID NO:8 which are capable of binding an anti-Rep antibody specific for  
 20 Rep protein having the amino acid sequence of SEQ ID NO:1 or SEQ ID NO:8. Preferably, such a fragment is an immunogenic fragment of the protein having the amino acid sequence of SEQ ID NO:1 or SEQ ID NO:8 which encompasses at least one epitope for an anti-Rep protein antibody against the Rep protein of SEQ ID NO:1 or SEQ ID NO:8 and, preferably, comprises at least 7, 8, 9, 10, 15, 20, 25 or 50  
 25 contiguous amino acids. In particular embodiments the fragment comprises or consists essentially of a domain of the Rep protein, for example, the N-terminal conserved region, the C-terminal variable region, the first or second DNA binding domain. A variant of the protein with SEQ ID NO:1 or SEQ ID NO:8 comprises one or more amino acid deletions, substitutions or additions compared to SEQ ID NO:1 and  
 30 has a homology of at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% to the amino acid sequence of SEQ ID NO:1 or SEQ ID NO:8, wherein the variant is capable of binding an anti-Rep antibody specific for a Rep protein having

the amino acid sequence of SEQ ID NO:1 or SEQ ID NO:8. Included within the definition of variant are, for example, polypeptides containing one or more analogues of an amino acid (including, for example, unnatural amino acids, peptide nucleic acid (PNA), etc.), polypeptides with substituted linkages, as well as other modifications known in the art, both naturally occurring and non-naturally occurring. The term Rep protein includes fusion proteins with a heterologous amino acid sequence, with a leader sequence or with a Tag-sequence and the like. In certain embodiments of the invention protein tags are genetically grafted onto the Rep protein described above, for example the Rep protein selected from the group consisting of MSBI1, MSBI2, CMI1, CMI2 or CMI3. In particular at least one protein tag is attached to a polypeptide consisting of an amino acid sequence as depicted in any one of SEQ ID NOs:1-3,8-12,14. Such protein tags may be removable by chemical agents or by enzymatic means. Examples of protein tags are affinity or chromatography tags for purification. For example the Rep protein may be fused to a Tag-sequence, for example, selected from the group consisting of His<sub>6</sub>-Tag (SEQ ID NO:4), T7-Tag (SEQ ID NO:5), FLAG-Tag (SEQ ID NO:6) and Strep-II-Tag (SEQ ID NO:7). a His-Tag (SEQ ID No:4), a T7-Tag (SEQ ID NO:5), FLAG-Tag (SEQ ID NO:6) or StreptII-Tag (SEQ ID NO:7). Further, fluorescence tags such as green fluorescence protein (GFP) or its variants may be attached to a Rep-protein according to the invention.

20

In a particular preferred embodiment the MSBI1 genome-encoded Rep protein (MSBI1 Rep) is codon-optimized for the production in human cell lines (e.g. HEK-293, HEK293TT, HEK293T, HEK293FT, HaCaT, HeLa, SiHa, CaSki, HDMEC, L1236, L428, BJAB, MCF7, Colo678, any primary cell lines) as well as bovine (e.g. MAC-T) or murine cell lines (e.g. GT1-7). This is described in detail in PCT/EP2017/075774.

25

The Rep protein of the invention, including the Rep fragments and Rep variants as defined above, can be prepared by classical chemical synthesis. The synthesis can be carried out in homogeneous solution or in solid phase. The polypeptides according to this invention can also be prepared by means of recombinant DNA techniques.

30

"Subject" as used herein refers to a mammalian individual or patient, including murines, cattle, for example bovines, simians and humans. Preferably, the subject is a human patient.

5 "Anti-Rep antibody" as used herein refers to an antibody binding at a detectable level to Rep protein which affinity is more strongly to the Rep protein of the invention than to a non-Rep protein. Preferably, the antigen affinity for Rep protein is at least 2 fold larger than background binding. In particular the anti-Rep antibody is specific for the MSBI1 Rep having the amino acid sequence of SEQ ID NO:1 or MSBI2 Rep. In particular embodiments the antibody is cross-specific for MSBI1 Rep, MSBI2 Rep, 10 CMI1 Rep, CMI2 Rep and/or CMI3 Rep. In certain embodiments the anti-Rep antibody is cross-specific for at least two, preferably all, of MSBI1 Rep, MSBI2 Rep, CMI1 Rep, CMI2 Rep and/or CMI3 Rep.

The inventors also tested the antibody level of colon cancer patients by contacting the Rep protein with a specimen suspected of containing anti-Rep protein antibodies 15 under conditions that permit the Rep protein to bind to any such antibody present in the specimen. Such conditions will typically be physiologic temperature, pH and ionic strength using an excess of Rep protein. The incubation of the Rep protein with the specimen is followed by detection of immune complexes comprised of the antigen. In certain embodiments either the Rep protein is coupled to a signal generating 20 compound, e.g. detectable label, or an additional binding agent, e.g. secondary anti-human antibody, coupled to a signal generating compound is used for detecting the immune complex.

Anti-Rep antibodies can be detected and quantified in assays based on Rep protein 25 as protein antigen, which serves as target for the mammalian, e.g. human, antibodies suspected in the specimen. Preferably, the Rep protein is purified and the specimen can be, for example, serum or plasma. The methods include immobilization of Rep protein on a matrix followed by incubation of the immobilized Rep protein with the specimen. Finally, the Rep-bound antibodies of the formed immunological complex 30 between Rep protein and antibodies of the specimen are quantified by a detection binding agent coupled to a signal generating compound, e.g. secondary HRP-(horseradish-peroxidase)-coupled detection antibody allowing for HRP-substrate

based quantification. This signal generating compound or label is in itself detectable or may be reacted with an additional compound to generate a detectable product.

Design of the immunoassay is subject to a great deal of variation, and many formats are known in the art. Protocols may, for example, use solid supports, or  
5 immunoprecipitation. Most assays involve the use of binding agents coupled to signal generating compounds, for example labelled antibody or labelled Rep protein; the labels may be, for example, enzymatic, fluorescent, chemiluminescent, radioactive, or dye molecules. Assays which amplify the signals from the immune complex are also known; examples of which are assays which utilize biotin and avidin or  
10 streptavidin, and enzyme-labeled and mediated immunoassays, such as ELISA assays.

The immunoassay may be in a heterogeneous or in a homogeneous format, and of a standard or competitive type. Both standard and competitive formats are known in  
15 the art.

In an immunoprecipitation or agglutination assay format the reaction between the Rep protein and the anti-Rep antibody forms a network that precipitates from the solution or suspension and forms a visible layer or film of precipitate. If no anti-Rep  
20 antibody is present in the specimen, no visible precipitate is formed.

In further embodiments the inventors used methods wherein an increased amount of Rep protein in a sample correlates with a diagnosis or predisposition of colon cancer. In such embodiments the Rep protein in the sample is detected by anti-Rep  
25 antibodies.

"Sample" as used herein refers to a biological sample encompassing cancerous colon tissue, peripheral tissue surrounding the cancerous colon tissue and (benign) colon polyps. The samples encompass tissue samples such as tissue cultures or  
30 biopsy specimen.

Such methods comprise the steps of detecting Rep protein in a sample from a subject by anti-Rep antibodies. In such methods Rep protein is detected in tissue samples by immunohistochemical methods or immunofluorescence microscopy.

5 In certain embodiments anti-Rep antibodies are used for the detection or capturing of the Rep protein in the sample.

The term „antibody“, preferably, relates to antibodies which consist essentially of pooled polyclonal antibodies with different epitopic specificities, as well as distinct monoclonal antibody preparations. As used herein, the term „antibody“(Ab) or  
10 „monoclonal antibody“ (Mab) is meant to include intact immunoglobulin molecules as well as antibody fragments (such as, for example, Fab and F(ab')<sub>2</sub> fragments) which are capable of specifically binding to Rep protein. Fab and F(ab')<sub>2</sub> fragments lack the Fc fragment of intact antibody, clear more rapidly from the circulation, and may have less non-specific tissue binding than an intact antibody. Thus, these fragments are  
15 preferred, as well as the products of a FAB or other immunoglobulin expression library. Moreover, antibodies useful for the purposes of the present invention include chimeric, single chain, multifunctional (e.g. bispecific) and humanized antibodies or human antibodies.

20 In certain embodiments the antibody or antigen binding fragment thereof is coupled to a signal generating compound, e.g., carries a detectable label. The antibody or antigen binding fragment thereof can be directly or indirectly detectably labeled, for example, with a radioisotope, a fluorescent compound, a bioluminescent compound, a chemiluminescent compound, a metal chelator or an enzyme. Those of ordinary  
25 skill in the art will know of other suitable labels for binding to the antibody, or will be able to ascertain such, using routine experimentation.

Anti-Rep antibodies are, preferably, raised (generated) against a Rep protein having the amino acid sequence of SEQ ID NO:1 or SEQ ID NO:8 or a fragment thereof by  
30 methods well known to those skilled in the art.

In certain embodiments anti-Rep antibodies are used in the methods of the invention which are capable of binding to several or all kinds of Rep proteins from the group of the Small Sphinx Genome (anti-Small-Sphinx-like Rep antibody or anti-SSLRep antibody). Such anti-SSLRep antibody binds to an epitope within the conserved N-terminal region of the Rep protein from amino acids 1 to 229 of SEQ ID NO:1. In particular embodiments anti-Rep antibodies of the anti-SSLRep type are used which bind to an epitope within SEQ ID NO:2 (amino acids 32-49 of SEQ ID NO:1) or SEQ ID NO:3 (amino acids 197-216 of SEQ ID NO:1). The peptide fragments of SEQ ID NO:2 and SEQ ID NO:3 are highly conserved among the Rep proteins from the Small Sphinx Genome group and appear to be exposed due to their hydrophilic character. Anti-Rep antibodies of the anti-SSLRep type may be produced by immunization, for example of mice or guinea pig, by peptides consisting essentially of the amino acid sequences as depicted in SEQ ID NOs:2 or 3; or by other immunogenic fragments, preferably comprising at least 8-15 amino acids, derived from the conserved N-terminal Rep protein region from amino acids 1 to 229 of SEQ ID NO:1.

In further embodiments anti-Rep antibodies specific for MSBI1 Rep protein are used. Such antibodies may be produced, for example, by immunization of a mammal such as mice or guinea pig with a full-length Rep protein having the amino acid sequence of SEQ ID NO:1.

Preferably, the methods of the invention use anti-Rep antibodies which are capable of detecting Rep protein up to ranges from picogram to femtogram.

Examples of such groups of anti-Rep antibodies are shown in Table 1:

| Antibody Group | Rep-Protein Localisation                      | Specificity                                       | Antibody                 | DSMZ deposit   |
|----------------|-----------------------------------------------|---------------------------------------------------|--------------------------|----------------|
| Group A        | cytoplasm +<br>nuclear membrane<br>(+nucleus) | MSBI1 +<br>small-sphinx-like<br>All BMMF1<br>Reps | AB01 523-1-1<br>(Ab 1-5) | DSM<br>ACC3327 |
| Group B        | speckles in<br>cytoplasm                      | MSBI1 +<br>small-sphinx-like                      | AB02 304-4-1<br>(Ab 5-2) | DSM<br>ACC3328 |

|                |                                                |                   |                                                                                                  |                                  |
|----------------|------------------------------------------------|-------------------|--------------------------------------------------------------------------------------------------|----------------------------------|
| <b>Group C</b> | cytoplasm +<br>nuclear membrane<br>(+ nucleus) | MSBI1<br>specific | MSBI1 381-6-2<br>(Ab 3-6)<br><br>MSBI1 572-13-<br>19 (Ab 10-3)<br><br>MSBI1 617-1-3<br>(Ab 11-5) | DSM<br>ACC3329                   |
| <b>Group D</b> | speckles in<br>cytoplasm                       | MSBI1<br>specific | D1: MSBI1 961-<br>2-2<br>D2: MSBI1 761-<br>5-1 (Ab 13)                                           | DSM<br>ACC3331<br>DSM<br>ACC3330 |

Anti-Rep antibodies of group A have an epitope within the amino acid sequence depicted in SEQ ID NO:3 (aa 198-217 of SEQ ID NO:1) and are capable of detecting MSBI1 Rep and Rep proteins comprising this conserved epitope of the Small Sphinx Genome group (e.g. MSBI2, CMI1, CMI4). In immunofluorescence assays such anti-Rep antibodies detect a specific Rep localisation pattern, wherein the main localisation is homogeneously distributed over the cytoplasm and nuclear membrane; and additional weak and homogeneously distributed localisation is seen in the nucleus. An example of such a group A antibody is antibody AB01 523-1-1 (also called antibody 1-5; DSM ACC3327) which was employed in the examples as group A antibody.

Anti-Rep antibodies of group B have an epitope within the amino acid sequence depicted in SEQ ID NO:2 (aa 33-50 of SEQ ID NO:1) and are capable of detecting MSBI1 Rep and Rep proteins comprising this conserved epitope of the Small Sphinx Genome group (e.g. MSBI2, CMI1, CMI4). In immunofluorescence assays such anti-Rep antibodies detect specifically speckles (cytoplasmatic aggregations) of the Rep protein (often in the periphery of the nuclear membrane). An example of such a group B antibody is the antibody designated as AB02 304-4-1 (also called antibody 5-2; DSM ACC3328) which was employed in the examples as group B antibody.

Anti-Rep antibodies of group C detect specifically a structural epitope of MSBI1 (SEQ ID NO:1). In immunofluorescence assays such anti-Rep antibodies detect a specific Rep localisation pattern, wherein the main localisation is homogeneously distributed over the cytoplasm and nuclear membrane; and additional weak and homogeneously

distributed localisation is seen in the nucleus. An example of such a group C antibody is antibody MSBI1 381-6-2 (also called antibody 3-6; DSM ACC3329) which was employed in the Example as group C antibody with an epitope in the sequence of aa 137-324. Another example of an antibody of a group C antibody is antibody MBSI1 572-13-19 (also called antibody 10-3) detecting an epitope in the C-terminal domain of MSBI 1 Rep (aa 230-324). Another example of an antibody of a group C antibody is antibody MBSI1 617-1-3 (also called antibody 11-5) detecting an epitope in the N-terminal domain of MSBI 1 Rep (aa 1-136).

Anti-Rep antibodies of group D detect specifically a structural epitope of MSBI1 (SEQ ID NO:1), where antibody MSBI1 961-2-2 designated as "D1" (DSM ACC3331) detects an epitope depicted in SEQ ID NO:9 (aa 281-287) in the C-terminal domain of MSBI1. Antibody MSBI1 761-5-1 (also called antibody 13; DSM ACC3328) designated as "D2" detects a 3D structural epitope of MSBI1 which is exclusively accessible under in vivo conditions and is not accessible in Western Blots. In immunofluorescence assays such anti-Rep antibodies detect specifically speckles (cytoplasmatic aggregations) of the Rep protein (often in the periphery of the nuclear membrane).

The invention is further illustrated by, but not limited to, the following examples:

20

#### **Example 1: Detection of BMMF protein targets in colon tissue**

Anti-Rep antibodies 10-3 (1:500 dilution), 3-6 (1:500), 11-5 (1:100) and 1-5 (1:100) were used for detection of BMMF-Rep related protein targets in colon tissue based on immunohistochemistry (IHC). CD68-specific staining of monocytes, circulating macrophages, and tissue macrophages was performed based on a rabbit anti-CD68 antibody (Cell Signaling Technology Europe BV, Catalogue #76437, 1:800). Therefore paraffin embedded tissue sections of colon tissue samples were incubated with the respective antibody on a Leica BOND automated IHC staining system (EDTA epitope retrieval buffer). Colorimetric staining was performed with the Leica DAB staining kit (Leica biosystems, rabbit anti-mouse secondary antibody ab125904

(1:500) from Abcam). The stained tissue sections were digitalized on a Zeiss optical light microscope.

## Results

5 A selection of 11 pairs of donor tissue samples (each tumor tissue and peripheral tissue) and 4 samples from different donors of colon polyp tissue were subjected to IHC staining with mouse monoclonal anti-Rep antibodies. With regard to the 11 tumor and peripheral samples, 9 out of 11 tissues showed strong specific antibody staining with at least two anti-Rep antibodies. Exemplarily, the staining with antibody 10-3  
10 shows specific detection of protein targets in tissue samples 4798, 4799, 4802, 4806, 4809 and 4813 in the tumor as well as in the peripheral tissue. In general, the staining intensity is higher for peripheral tissue, where intensely stained cytoplasmic aggregates of smaller size concentrate within distinct islands of cells within the *lamina propria* between the characteristic Crypts of Lieberkühn of the colon tissue (all  
15 pictures showing traversal cuts throughout the Crypts of Lieberkühn with the (inner) lumen of the crypts having contact to the colonic fluids in the center of the ring-like arrangement of epithelial cells of the crypts). Staining of tumor tissues also shows intense staining of smaller sized aggregates mainly within the cytoplasmic regions of the stroma cells within the tumor tissue. The same aggregate-like structures are also  
20 visible in the *lamina propria* of colon polyps by the use of four different anti-Rep antibodies for staining of sequential cuts. Both, for colon peripheral tissue and for colon polyps, the regions with highest Rep-specific antibody detection correlate with regions with highest detection levels for CD68 positive cells pointing towards a localization of the Rep-specific antigens in regions with especially high levels of  
25 inflammatory monocytes, circulating macrophages, or tissue macrophages.

For verification of IHC results, part of the patient tissue material was lysed in 8M urea denaturing lysis buffer with a tissue homogenizer and subjected to SDS-PAGE. The areas of interest were cut, tryptically digested and analyzed by mass spectroscopy for the presence of BMMF1 specific peptide targets. A set of 4 pairs (tumor and  
30 peripheral tissue) with strong IHC signal levels was chosen together with one pair with no IHC signal. In line with the IHC results, only the 4 strongly IHC-positive

peripheral tissues allowed for Mass-Spec detection of one BMMF1-group-specific peptide sequence.

To confirm specificity of the antibody detection in IHC, DNA isolations from the five samples with strongest IHC signal detection were subjected to rolling circle  
5 amplification and PCR with BMMF-group1 specific primers and DNA sequencing to analyze the presence of BMMF DNA in the same tissue material. Indeed, based on a set of the 5 strongest IHC sample pairs (tumor and peripheral tissue), within each donor MSB11-specific DNA was isolated from at least one tissue type (tumor or peripheral tissue).

10

### **Example 2: Immunohistochemical staining of CRC tissue multi arrays (TMAs)**

The detection of BMMF1 Rep protein antigens in tissue material of patients with colorectal cancer (CRC) was tested based on immunohistochemical staining of CRC tissue multi arrays (TMAs) with a specific mouse monoclonal anti-BMMF1 Rep  
15 antibody. Therefore, TMAs representing a total of 259 CRC patients were used, which offered a total of each two tissue spots of tumoral as well as two spots of peritumoral tissue regions per patient. All patients were treated in the University Hospital in Heidelberg in 2003 and 2004. The TMA samples were provided by the tissue bank of the National Center for Tumor Diseases (NCT, Heidelberg, Germany)  
20 in accordance with the regulations of the tissue bank and the approval of the ethics committee of Heidelberg University by permission of Hermann Brenner, Michael Hoffmeister and Jenny Chang-Claude (Tissue Bank of the National Center for Tumor Diseases (NCT) Heidelberg, Germany and Institute of Pathology, Heidelberg University Hospital, Germany).

### **25 Tissue staining**

The TMAs were stained fully automatically on a BOND MAX machine (Leica Biosystems) with EDTA epitope retrieval and the given antibody incubations (Table 2). Detection was performed by using Bond Polymer Refine Detection Kit (DS9800 Leica DAB Kit) including DAB chromogen and hematoxylin counterstain. Slides were

scanned with a digital slide scanner (Hamamatsu) and analyzed based on Hamamatsu NDP viewer software.

Table 2: Experimental parameters used for IHC staining

| <u>Antibody</u>   | <u>Source</u>         | <u>Host</u> | <u>Dilution</u> | <u>Final concentration in µg/ml</u> | <u>Incubation time</u>     |
|-------------------|-----------------------|-------------|-----------------|-------------------------------------|----------------------------|
| <u>Primary</u>    |                       |             |                 |                                     |                            |
| Rep mAb #3-6      | T. Bund. DKFZ         | mouse       | 1:500           | 3.9                                 | 30 min at room temperature |
| CD68              | Cell signaling #76437 | rabbit      | 1:1000          |                                     |                            |
| <u>Secondary</u>  |                       |             |                 |                                     |                            |
| rabbit anti-mouse | Abcam #125904         | rabbit      | 1:500           |                                     | 20 min at room temperature |

5

#### BMMF Signal quantification

Two researchers independently scored the tissues spots according to the following scoring criteria. For each antibody, the percentage of stained cells (positivity) and intensity (I) were determined. For Rep staining only stromal staining was evaluated representing the dominant localization of detected BMMF antigen with negligible stainings in tumor region in tumoral tissues spots or crypts in peritumoral spots. The positivity (POS) of Rep staining was assessed using a three-level scale in which 0 indicated no positive tissue parts at all, 1 indicated 1-10% positive, 2 indicated 11-30%, 3 indicated more than 30% of the individual tissue spots showing stained aggregates. For CD68/macrophage staining the positivity score was set as 0 for negative tissue spots, 1 indicated <20% positive cells, 2 indicated 20-60%, 3 indicated more than 60% of positive cells. Intensity (I) was generally graded as follows: 0 = no reaction, 1 = mild, 2 = intense staining. For statistical analysis the immunoreactive score IRS was calculated as follows:

$$IRS = [(I(Investigator_1) + I(Investigator_2))/2] * [(POS(Investigator_1) + POS(Investigator_2))/2]$$

; minimum value = 0, maximum value = 6.

An overview is shown in Table 3. Out of 259 patients, 12 patients were excluded because of bad tissue quality and one because of missing clinical data. In total, data  
5 from 246 patients was processed for further analysis.

Table 3: Parameters used for quantification of TMA staining (IRS: Immunoreactive Score).

|                                                                                                                     | Intensity<br>(I) | I           | Positivity<br>(proportion<br>positive<br>POS) | of<br>cells,<br>POS |
|---------------------------------------------------------------------------------------------------------------------|------------------|-------------|-----------------------------------------------|---------------------|
| <b><u>Rep staining</u></b>                                                                                          | 0                | no reaction | 0                                             | 0                   |
|                                                                                                                     | 1                | mild        | 1                                             | 1-10%               |
|                                                                                                                     | 2                | strong      | 2                                             | 11-30%              |
|                                                                                                                     |                  |             | 3                                             | >31%                |
| <b><u>CD68 staining</u></b>                                                                                         | 0                | no reaction | 0                                             | 0                   |
|                                                                                                                     | 1                | mild        | 1                                             | <20%                |
|                                                                                                                     | 2                | strong      | 2                                             | 20-60%              |
|                                                                                                                     |                  |             | 3                                             | >60%                |
| <b><u>IRS</u></b> = $[(I(Investigator_1) + I(Investigator_2))/2] * [(POS(Investigator_1) + POS(Investigator_2))/2]$ |                  |             |                                               |                     |

#### Statistical analysis

10 Statistical analysis was performed using R-3.5.2 (<http://www.r-project.org>) and R-studio 2016. In order to compare tumoral and peritumoral tissue a 2-sided Wilcoxon signed rank test was performed. Interdependence between immunostaining and clinical data was calculated using Kruskal-Wallis test. Survival curves were plotted using the Kaplan-Meier method and analyzed using the log-rank test. Hazard ratio  
15 was determined by Cox-regression analysis. P-Values less than 0.05 were considered to be statistically significant.

## Results

In total, scoring information was collected for n=246 CRC patients. In Rep scorings for tumoral tissue regions 49,3% of the patient tissues showed positivity, while  
 5 scorings for peritumoral tissues showed 99,1% of the patient tissues to be positive. The corresponding distributions of IRS values shows an asymmetric distribution for the tumoral tissue with an enrichment of cases with IRS values of 0 and 1 (Figure 7). This trend is inverted for the distribution of the IRS values for the peritumoral scoring with an enrichment of IRS values of >3 (Figure 8).

10 With respect to the quantification of signal obtained by staining of tumoral (Figure 9) and peritumoral tissue regions (Figure 10) with CD68 antibodies >99% of the patients showed positivity with a balanced distribution of IRS values in both cases.

Comparison of Rep IRS scores of peritumoral and tumoral tissues suggested a significantly increased Rep detection ( $p = > 2.2e-16$ ) in the peritumoral tissue of the  
 15 analyzed CRC patients (Figure 11).

The inventors concluded that for cancer induction the amount of BMMF antigen within the *lamina propria* between the colon crypts is important for the induction of chronic local inflammation, induction of diffusing ROS/NOS which finally induces random mutations in dividing stem and daughter cells of neighboring crypts which  
 20 might turn into early polyps/tumor precursor cells by manifestation of key mutations. Thus, a correlation of peritumoral BMMF levels and intensity of CD68-positive macrophages should be observable. Indeed, by testing IRS values for Rep-detection against IRS-values for CD68-detection within the corresponding CRC patients, a significant correlation of Rep and CD68 detection is observed ( $p=0.00045$  by Kruskal  
 25 Wallis test), which shows that higher detection levels of Rep correlate with higher levels of CD68 detection (Figure 12).

To test a prognostic score of Rep detection in peritumoral regions of CRC patients on patient survival time, Kaplan Meier curves were calculated based on a set of n=243 patients (77 events based on CRC-specific death, 166 censored), which were  
 30 grouped into "Rep low" for IRS values of 0-1.9 and "Rep high" for IRS values of 2-6

(Figure 13). High levels of Rep in CRC patients correlate with a significant reduction of patient overall survival time ( $p=0.0017$ ).

Increased Rep levels correlate with a hazard ratio of 4.7 (95% CI: 1-19) indicating a negative correlation of Rep detection and patient survival ( $p=0.00475$ ) and underlining the prognostic score of Rep detection in CRC tissue (Figure 14).

Patients scored "Rep low" show a 5 and 10-year survival chance of 92%, whereas patients with "Rep high" show chances of 74% after 5 years and 65% after 10 years, representing a decrease of 20 to 30%. The prognostic effect of Rep detection is not explainable by CD68 detection, as survival analysis based on "CD68 high" and "CD68 low" Kaplan Meier analysis shows no significant differences between the two survival curves ( $p=0.271$ , n.s.)(Figure 15).

### **Example 3: Detection of BMMF Rep antigens by immunoblotting of tissue lysates of CRC patients**

Peritumoral or tumoral tissue material of the same patient (e.g. patient 4800 P (peritumoral) or 4800 T (tumoral)) or tissue material from early polyps of four individual donors was lysed in 8 M urea, 100 mM  $\text{NaH}_2\text{PO}_4$ , 10 mM Tris, 5 mM DTT, pH 8.0 with the use of protease inhibitors (P8340 Sigma-Aldrich) in a tissue homogenizer at 4 °C with silica beads before SDS-PAGE and immunodetection with anti-BMMF Rep antibodies, in particular anti-Rep Ab 5-2 (1:250 dilution). Target bands at around 42 kDa were clearly visible with strongest intensities for the peritumoral CRC patient sample 4800 and even stronger bands for tissue lysates of early polyps of the four individual patients (Figure 16).

### **Example 4: PCR amplification of BMMF Rep DNA**

In addition, laser microdissection of immunohistochemically stained peritumoral tissue regions of CRC patients (anti-Rep Ab 3-6) and following DNA preparation led to the identification of full-length MSBI1.176 BMMF DNA in all three patients tested (4799, 4808 and 4809). Therefore, total DNA was prepared from the dissected tissue

regions by over-night incubation of the dissected tissue with 25  $\mu$ l of a 5% suspension of Chelex beads (Bio-Rad Cat #142-1253), 5  $\mu$ g protease at 750 rpm on a thermomixer. The suspension was vortexed for 10 s at maximum speed, then placed in a thermomixer at 99°C for 8 min. The suspension was then centrifuged for 5 min at 12.000 g at room temperature to pellet the Chelex resin. 1  $\mu$ l of the supernatant was used for rolling circle amplification (RCA) with phi29 DNA Polymerase (New England Biolabs M0269S, 10 U) and the use of Exo-Resistant Random primers (Thermo Fisher scientific SO181, 25  $\mu$ M), dNTPs (0,75 mM each), 50 mM Tris-HCl, 10 mM MgCl<sub>2</sub>, 10 mM (NH<sub>4</sub>)<sub>2</sub> SO<sub>4</sub>, 4 mM DTT, 0,4 mg/ml BSA, pH 7.5.

| RCA setup                           |                                                                                                            |
|-------------------------------------|------------------------------------------------------------------------------------------------------------|
| 1 $\mu$ l DNA (50ng)                | 3 min 95 °C<br>then cool on ice                                                                            |
| 1 $\mu$ l primer 1,1 $\mu$ g        |                                                                                                            |
| 1 $\mu$ l 10x buffer                |                                                                                                            |
| 7 $\mu$ l H2O                       |                                                                                                            |
| 1 $\mu$ l 10x buffer                | prepare this solution (10 $\mu$ l) and add to the sample cooled on ice<br>then, 18 h 30 °C<br>10 min 65 °C |
| 0.4 $\mu$ l BSA 8 $\mu$ g           |                                                                                                            |
| 1,6 $\mu$ l H2O                     |                                                                                                            |
| 6 $\mu$ l d-NTP                     |                                                                                                            |
| 1 $\mu$ l phi29 10U                 |                                                                                                            |
| total Volume of reaction 20 $\mu$ l |                                                                                                            |

After RCA, 3  $\mu$ l of the RCA reaction was used for PCR with BMMF-specific back-to-back primers No (Gag gac gaa tta ata tta caa gtc) and Xo (Gtt ctc gct ttt ctt ggt aa) or Nn (gga tta atg cca atg atc c) and Xn (ctt tgc ctg ttt ctc tcg) with each 10 pmol per 50  $\mu$ l reaction together with 8  $\mu$ l dNTPs, 25  $\mu$ l buffer GC I, 0,5  $\mu$ l TAKARA Taq polymerase (Takara RR02AG) with the following PCR setup.

|           |        |          |        |           |        |       |        |
|-----------|--------|----------|--------|-----------|--------|-------|--------|
| NoXo/NnXn |        |          |        |           |        |       |        |
| 94° C     | 1 min  |          |        |           |        | 72° C | 10 min |
| 94° C     | 30 sec | 94° C    | 30 sec | 94° C     | 30 sec |       |        |
| 64° C     | 1 min  | 62° C    | 1 min  | 62° C     | 1 min  |       |        |
| 72° C     | 2 min  | 72° C    | 2 min  | 72° C     | 2 min  |       |        |
| 5 cycles  |        | 5 cycles |        | 30 cycles |        |       |        |

- If necessary, PCR targets were subjected to an additional cycle of PCR amplification after gel extraction of the PCR targets, whenever the target bands showed inadequate band intensity. After PCR, around 5 ng of DNA was used for cloning into a
- 5 pCR2.1 TA cloning vector (TA-cloning kit, Invitrogen K203040) according to the manufacturer's protocol followed by sequencing of the PCR insert.

# SEQUENCE LISTING

<110> Deutsches Krebsforschungszentrum

<120> Use of BMMF Rep Protein as Biomarker for Colon Cancer

<130> K 3647EP

<150> EP18154190

<151> 2018-01-30

<150> XX

<160> 19

<170> BiSSAP 1.3.6

<210> 1

<211> 324

<212> PRT

<213> artificial sequence

<220>

<223> MSBI1 Rep protein

<400> 1

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Ser | Asp | Leu | Ile | Val | Lys | Asp | Asn | Ala | Leu | Met | Asn | Ala | Ser | Tyr |
| 1   |     |     |     | 5   |     |     |     |     | 10  |     |     |     |     | 15  |     |
| Asn | Leu | Ala | Leu | Val | Glu | Gln | Arg | Leu | Ile | Leu | Leu | Ala | Ile | Ile | Glu |
|     |     |     | 20  |     |     |     |     | 25  |     |     |     |     | 30  |     |     |
| Ala | Arg | Glu | Thr | Gly | Lys | Gly | Ile | Asn | Ala | Asn | Asp | Pro | Leu | Thr | Val |
|     |     | 35  |     |     |     | 40  |     |     |     |     | 45  |     |     |     |     |
| His | Ala | Ser | Ser | Tyr | Ile | Asn | Gln | Phe | Asn | Val | Glu | Arg | His | Thr | Ala |
|     | 50  |     |     |     |     | 55  |     |     |     | 60  |     |     |     |     |     |
| Tyr | Gln | Ala | Leu | Lys | Asp | Ala | Cys | Lys | Asp | Leu | Phe | Ala | Arg | Gln | Phe |
| 65  |     |     |     |     | 70  |     |     |     |     | 75  |     |     |     | 80  |     |
| Ser | Tyr | Gln | Glu | Lys | Arg | Glu | Arg | Gly | Arg | Ile | Asn | Ile | Thr | Ser | Arg |
|     |     |     |     | 85  |     |     |     |     | 90  |     |     |     |     | 95  |     |
| Trp | Val | Ser | Gln | Ile | Gly | Tyr | Met | Asp | Asp | Thr | Ala | Thr | Val | Glu | Ile |
|     |     |     | 100 |     |     |     |     | 105 |     |     |     |     | 110 |     |     |
| Ile | Phe | Ala | Pro | Ala | Val | Val | Pro | Leu | Ile | Thr | Arg | Leu | Glu | Glu | Gln |
|     |     | 115 |     |     |     |     | 120 |     |     |     |     | 125 |     |     |     |
| Phe | Thr | Gln | Tyr | Asp | Ile | Glu | Gln | Ile | Ser | Gly | Leu | Ser | Ser | Ala | Tyr |
|     | 130 |     |     |     |     | 135 |     |     |     | 140 |     |     |     |     |     |
| Ala | Val | Arg | Met | Tyr | Glu | Leu | Leu | Ile | Cys | Trp | Arg | Ser | Thr | Gly | Lys |
| 145 |     |     |     |     | 150 |     |     |     |     | 155 |     |     |     | 160 |     |
| Thr | Pro | Ile | Ile | Glu | Leu | Asp | Glu | Phe | Arg | Lys | Arg | Ile | Gly | Val | Leu |
|     |     |     |     | 165 |     |     |     | 170 |     |     |     |     |     | 175 |     |
| Asp | Thr | Glu | Tyr | Thr | Arg | Thr | Asp | Asn | Leu | Lys | Met | Arg | Val | Ile | Glu |
|     |     |     | 180 |     |     |     | 185 |     |     |     |     |     | 190 |     |     |
| Leu | Ala | Leu | Lys | Gln | Ile | Asn | Glu | His | Thr | Asp | Ile | Thr | Ala | Ser | Tyr |
|     |     | 195 |     |     |     |     | 200 |     |     |     |     | 205 |     |     |     |
| Glu | Gln | His | Lys | Lys | Gly | Arg | Val | Ile | Thr | Gly | Phe | Ser | Phe | Lys | Phe |
|     | 210 |     |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     |

Lys His Lys Lys Gln Asn Ser Asp Lys Thr Pro Lys Asn Ser Asp Ser  
 225 230 235 240  
 Ser Pro Arg Ile Val Lys His Ser Gln Ile Pro Thr Asn Ile Val Lys  
 245 250 255  
 Gln Pro Glu Asn Ala Lys Met Ser Asp Leu Glu His Arg Ala Ser Arg  
 260 265 270  
 Val Thr Gly Glu Ile Met Arg Asn Arg Leu Ser Asp Arg Phe Lys Gln  
 275 280 285  
 Gly Asp Glu Ser Ala Ile Asp Met Met Lys Arg Ile Gln Ser Glu Ile  
 290 295 300  
 Ile Thr Asp Ala Ile Ala Asp Gln Trp Glu Ser Lys Leu Glu Glu Phe  
 305 310 315 320  
 Gly Val Val Phe

<210> 2  
 <211> 18  
 <212> PRT  
 <213> artificial sequence

<220>  
 <223> Rep peptide

<400> 2  
 Glu Ala Arg Glu Thr Gly Lys Gly Ile Asn Ala Asn Asp Pro Leu Thr  
 1 5 10 15  
 Val His

<210> 3  
 <211> 20  
 <212> PRT  
 <213> artificial sequence

<220>  
 <223> Rep peptide

<400> 3  
 Lys Gln Ile Asn Glu His Thr Asp Ile Thr Ala Ser Tyr Glu His Lys  
 1 5 10 15  
 Lys Gly Arg Thr  
 20

<210> 4  
 <211> 8  
 <212> PRT  
 <213> artificial sequence

<220>  
 <223> His tag

<400> 4  
Gly Ala His His His His His His  
1 5

<210> 5  
<211> 11  
<212> PRT  
<213> artificial sequence

<220>  
<223> T7 tag

<400> 5  
Met Ala Ser Met Thr Gly Gly Gln Gln Met Gly  
1 5 10

<210> 6  
<211> 8  
<212> PRT  
<213> artificial sequence

<220>  
<223> FLAG tag

<400> 6  
Asp Tyr Lys Asp Asp Asp Asp Lys  
1 5

<210> 7  
<211> 8  
<212> PRT  
<213> artificial sequence

<220>  
<223> Trep II tag

<400> 7  
Trp Ser His Pro Gln Phe Glu Lys  
1 5

<210> 8  
<211> 319  
<212> PRT  
<213> artificial sequence

<220>  
<223> MSBI2.176

<400> 8

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Ser | Lys | Leu | Val | Val | Lys | Asp | Asn | Ala | Leu | Met | Asn | Ala | Ser | Tyr |
| 1   |     |     |     | 5   |     |     |     |     | 10  |     |     |     |     | 15  |     |
| Asn | Leu | Asp | Leu | Val | Glu | Gln | Arg | Leu | Ile | Leu | Leu | Ala | Ile | Ile | Glu |
|     |     |     | 20  |     |     |     |     | 25  |     |     |     |     | 30  |     |     |
| Ala | Arg | Glu | Ser | Gly | Lys | Gly | Ile | Asn | Ala | Asn | Asp | Pro | Leu | Thr | Val |
|     |     | 35  |     |     |     |     | 40  |     |     |     |     | 45  |     |     |     |
| His | Ala | Glu | Ser | Tyr | Ile | Asn | Gln | Phe | Gly | Val | His | Arg | Val | Thr | Ala |
|     | 50  |     |     |     |     | 55  |     |     |     |     | 60  |     |     |     |     |
| Tyr | Gln | Ala | Leu | Lys | Asp | Ala | Cys | Asp | Asn | Leu | Phe | Ala | Arg | Gln | Phe |
| 65  |     |     |     |     | 70  |     |     |     |     | 75  |     |     |     |     | 80  |
| Ser | Tyr | Gln | Ser | Lys | Ser | Glu | Lys | Gly | Asn | Ile | Gln | Asn | His | Arg | Ser |
|     |     |     |     | 85  |     |     |     |     | 90  |     |     |     |     | 95  |     |
| Arg | Trp | Val | Ser | Glu | Ile | Ile | Tyr | Ile | Asp | Thr | Glu | Ala | Thr | Val | Lys |
|     |     |     | 100 |     |     |     |     | 105 |     |     |     |     |     | 110 |     |
| Ile | Ile | Phe | Ala | Pro | Ala | Ile | Val | Pro | Leu | Ile | Thr | Arg | Leu | Glu | Glu |
|     |     | 115 |     |     |     |     | 120 |     |     |     |     | 125 |     |     |     |
| Gln | Phe | Thr | Lys | Tyr | Asp | Ile | Glu | Gln | Ile | Ser | Asp | Leu | Ser | Ser | Ala |
|     | 130 |     |     |     |     | 135 |     |     |     |     | 140 |     |     |     |     |
| Tyr | Ala | Ile | Arg | Leu | Tyr | Glu | Leu | Leu | Ile | Cys | Trp | Arg | Ser | Thr | Gly |
| 145 |     |     |     |     | 150 |     |     |     |     | 155 |     |     |     |     | 160 |
| Lys | Thr | Pro | Ile | Ile | Gly | Leu | Gly | Glu | Phe | Arg | Asn | Arg | Val | Gly | Val |
|     |     |     |     | 165 |     |     |     | 170 |     |     |     |     |     | 175 |     |
| Leu | Asp | Ser | Glu | Tyr | His | Arg | Ile | Ala | His | Leu | Lys | Glu | Arg | Val | Ile |
|     |     |     | 180 |     |     |     |     | 185 |     |     |     |     |     | 190 |     |
| Glu | His | Ser | Ile | Lys | Gln | Ile | Asn | Glu | His | Thr | Asp | Ile | Thr | Ala | Thr |
|     |     | 195 |     |     |     |     | 200 |     |     |     |     | 205 |     |     |     |
| Tyr | Glu | Gln | His | Lys | Lys | Gly | Arg | Thr | Ile | Thr | Gly | Phe | Ser | Phe | Lys |
|     | 210 |     |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     |
| Phe | Lys | Gln | Lys | Lys | Pro | Lys | Gln | Ala | Glu | Ile | Ala | Thr | Glu | Thr | Pro |
| 225 |     |     |     |     | 230 |     |     |     |     | 235 |     |     |     |     | 240 |
| Lys | Thr | Ala | Thr | Asn | Asp | Pro | Asp | Thr | Thr | Lys | Pro | Leu | Thr | Glu | Pro |
|     |     |     |     | 245 |     |     |     |     | 250 |     |     |     |     | 255 |     |
| Gln | Ile | Ala | Lys | Tyr | Ser | Met | Ile | Leu | Cys | Lys | Leu | Gly | Ser | Ile | Ser |
|     |     |     | 260 |     |     |     |     | 265 |     |     |     |     | 270 |     |     |
| Asp | Leu | Ser | Asn | Phe | Pro | Asp | Tyr | Pro | Ala | Phe | Ala | Asn | Trp | Ile | Gly |
|     |     | 275 |     |     |     |     | 280 |     |     |     |     | 285 |     |     |     |
| Asn | Ile | Leu | Arg | Asn | Pro | Glu | Lys | Ala | Asp | Glu | Gln | Ile | Ala | Lys | Arg |
|     | 290 |     |     |     |     | 295 |     |     |     |     | 300 |     |     |     |     |
| Ile | Phe | Thr | Ala | Leu | Lys | Thr | Glu | Thr | Asp | Tyr | Ser | Lys | Lys | Asn |     |
| 305 |     |     |     |     | 310 |     |     |     |     | 315 |     |     |     |     |     |

<210> 9

<211> 7

<212> PRT

<213> artificial sequence

<220>

<223> MSBI1 epitope

<400> 9

|     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|
| Asn | Arg | Leu | Ser | Asp | Arg | Phe |
| 1   |     |     |     | 5   |     |     |



<210> 11  
 <211> 324  
 <212> PRT  
 <213> Artificial Sequence

<220>  
 <223> CMI2.214

<400> 11

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Ser | Asp | Leu | Ile | Val | Lys | Asp | Asn | Ala | Leu | Met | Asn | Ala | Ser | Tyr |
| 1   |     |     |     | 5   |     |     |     |     | 10  |     |     |     |     | 15  |     |
| Asn | Leu | Asp | Leu | Val | Glu | Gln | Arg | Leu | Ile | Leu | Leu | Ala | Ile | Leu | Glu |
|     |     |     | 20  |     |     |     |     | 25  |     |     |     |     | 30  |     |     |
| Ala | Arg | Glu | Thr | Gly | Lys | Gly | Ile | Asn | Ala | Asn | Asp | Pro | Leu | Thr | Val |
|     |     | 35  |     |     |     |     | 40  |     |     |     |     | 45  |     |     |     |
| His | Ala | Glu | Ser | Tyr | Ile | Asn | Gln | Phe | Gly | Val | Ala | Arg | Gln | Thr | Ala |
|     | 50  |     |     |     |     | 55  |     |     |     |     | 60  |     |     |     |     |
| Tyr | Gln | Ala | Leu | Lys | Asp | Ala | Cys | Lys | Asp | Leu | Phe | Ala | Arg | Gln | Phe |
| 65  |     |     |     |     | 70  |     |     |     |     | 75  |     |     |     |     | 80  |
| Ser | Tyr | Gln | Glu | Lys | Arg | Glu | Arg | Gly | Arg | Ala | Asn | Ile | Thr | Ser | Arg |
|     |     |     |     | 85  |     |     |     |     | 90  |     |     |     |     | 95  |     |
| Trp | Val | Ser | Gln | Ile | Ala | Tyr | Ile | Asp | Glu | Thr | Ala | Thr | Val | Glu | Val |
|     |     |     | 100 |     |     |     |     | 105 |     |     |     |     | 110 |     |     |
| Ile | Phe | Ala | Pro | Ala | Val | Val | Pro | Leu | Ile | Thr | Arg | Leu | Glu | Glu | Gln |
|     |     | 115 |     |     |     |     | 120 |     |     |     |     | 125 |     |     |     |
| Phe | Thr | Gln | Tyr | Asp | Ile | Glu | Gln | Ile | Ser | Gly | Leu | Ser | Ser | Ala | Tyr |
|     | 130 |     |     |     |     | 135 |     |     |     |     | 140 |     |     |     |     |
| Ala | Val | Arg | Leu | Tyr | Glu | Leu | Leu | Ile | Cys | Trp | Arg | Ser | Thr | Gly | Lys |
| 145 |     |     |     |     | 150 |     |     |     |     | 155 |     |     |     |     | 160 |
| Thr | Pro | Val | Ile | Glu | Leu | Ala | Glu | Phe | Arg | Lys | Arg | Leu | Gly | Val | Leu |
|     |     |     |     | 165 |     |     |     |     | 170 |     |     |     |     | 175 |     |
| Asn | Asp | Glu | Tyr | Thr | Arg | Ser | Asp | Asn | Phe | Lys | Lys | Trp | Ile | Ile | Glu |
|     |     | 180 |     |     |     |     |     | 185 |     |     |     |     | 190 |     |     |
| Asn | Pro | Ile | Lys | Gln | Ile | Asn | Glu | His | Thr | Asp | Ile | Thr | Ala | Ser | Tyr |
|     | 195 |     |     |     |     | 200 |     |     |     |     |     | 205 |     |     |     |
| Glu | Gln | His | Lys | Lys | Gly | Arg | Thr | Ile | Thr | Gly | Phe | Ser | Phe | Lys | Phe |
|     | 210 |     |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     |
| Lys | Gln | Lys | Lys | Lys | Thr | Glu | Pro | Glu | Thr | Pro | Lys | Asn | Ser | Asp | Ser |
| 225 |     |     |     |     | 230 |     |     |     |     | 235 |     |     |     |     | 240 |
| Ser | Gln | Arg | Ile | Glu | Lys | Pro | Ser | Gln | Ile | Pro | Ala | Asn | Ile | Val | Lys |
|     |     |     |     | 245 |     |     |     |     | 250 |     |     |     |     | 255 |     |
| Gln | Pro | Glu | Asn | Ala | Asn | Leu | Ser | Asp | Leu | Gln | His | Arg | Ala | Ser | Lys |
|     |     |     | 260 |     |     |     |     | 265 |     |     |     |     | 270 |     |     |
| Ile | Thr | Gly | Leu | Ile | Met | Ser | Asn | Arg | Leu | Ser | Asp | Arg | Phe | Lys | Gln |
|     | 275 |     |     |     |     |     | 280 |     |     |     |     | 285 |     |     |     |
| Gly | Asp | Glu | Ser | Ile | Met | Gln | Met | Met | Ala | Arg | Ile | Gln | Ser | Glu | Ile |
|     | 290 |     |     |     |     | 295 |     |     |     |     | 300 |     |     |     |     |
| Thr | Thr | Asp | Ser | Ile | Ala | Asp | Gln | Trp | Gln | Ser | Lys | Leu | Glu | Glu | Phe |
| 305 |     |     |     |     | 310 |     |     |     |     | 315 |     |     |     |     | 320 |
| Gly | Val | Val | Phe |     |     |     |     |     |     |     |     |     |     |     |     |

<210> 12  
 <211> 324  
 <212> PRT  
 <213> Artificial Sequence

<220>  
 <223> CMI3.168

<400> 12

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Ser | Asp | Leu | Ile | Val | Lys | Asp | Asn | Ala | Leu | Met | Asn | Ala | Ser | Tyr |
| 1   |     |     |     | 5   |     |     |     |     | 10  |     |     |     |     | 15  |     |
| Asn | Leu | Ala | Leu | Val | Glu | Gln | Arg | Leu | Ile | Leu | Leu | Ala | Ile | Leu | Glu |
|     |     |     | 20  |     |     |     |     | 25  |     |     |     |     | 30  |     |     |
| Ala | Arg | Glu | Thr | Gly | Lys | Gly | Ile | Asn | Ala | Asn | Asp | Pro | Leu | Thr | Val |
|     |     | 35  |     |     |     |     | 40  |     |     |     |     | 45  |     |     |     |
| His | Ala | Ser | Ser | Tyr | Ile | Asn | Gln | Phe | Asn | Val | Glu | Arg | His | Thr | Ala |
|     | 50  |     |     |     |     | 55  |     |     |     |     | 60  |     |     |     |     |
| Tyr | Gln | Ala | Leu | Lys | Asp | Ala | Cys | Lys | Asp | Leu | Phe | Ala | Arg | Gln | Phe |
| 65  |     |     |     |     | 70  |     |     |     |     | 75  |     |     |     | 80  |     |
| Ser | Tyr | Gln | Glu | Lys | Arg | Glu | Arg | Gly | Arg | Ala | Asn | Ile | Thr | Ser | Arg |
|     |     |     |     | 85  |     |     |     |     | 90  |     |     |     |     | 95  |     |
| Trp | Val | Ser | Gln | Ile | Ala | Tyr | Ile | Asp | Glu | Thr | Ala | Thr | Val | Glu | Val |
|     |     |     | 100 |     |     |     |     | 105 |     |     |     |     | 110 |     |     |
| Ile | Phe | Ala | Pro | Ala | Val | Val | Pro | Leu | Ile | Thr | Arg | Leu | Glu | Glu | Gln |
|     |     | 115 |     |     |     |     | 120 |     |     |     |     | 125 |     |     |     |
| Phe | Thr | Gln | Tyr | Asp | Ile | Glu | Gln | Ile | Ser | Gly | Leu | Ser | Ser | Ala | Tyr |
|     | 130 |     |     |     |     | 135 |     |     |     |     | 140 |     |     |     |     |
| Ala | Val | Arg | Leu | Tyr | Glu | Leu | Leu | Ile | Cys | Trp | Arg | Thr | Thr | Gly | Lys |
| 145 |     |     |     |     | 150 |     |     |     |     | 155 |     |     |     |     | 160 |
| Thr | Pro | Val | Leu | Asp | Leu | Thr | Glu | Phe | Arg | Lys | Arg | Leu | Gly | Val | Leu |
|     |     |     |     | 165 |     |     |     |     | 170 |     |     |     |     | 175 |     |
| Asp | Thr | Glu | Tyr | Thr | Arg | Thr | Asp | Asn | Leu | Lys | Met | Arg | Val | Ile | Glu |
|     |     |     | 180 |     |     |     |     | 185 |     |     |     |     | 190 |     |     |
| Gln | Ser | Leu | Lys | Gln | Ile | Asn | Lys | His | Thr | Asp | Ile | Thr | Ala | Ser | Tyr |
|     |     | 195 |     |     |     |     | 200 |     |     |     |     | 205 |     |     |     |
| Glu | Gln | His | Lys | Lys | Gly | Arg | Thr | Ile | Thr | Gly | Phe | Ser | Phe | Lys | Phe |
|     | 210 |     |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     |
| Lys | Gln | Lys | Lys | Lys | Thr | Glu | Pro | Glu | Thr | Pro | Lys | Asn | Asn | Asp | Ser |
| 225 |     |     |     |     | 230 |     |     |     |     | 235 |     |     |     |     | 240 |
| Gly | Val | Ser | Lys | Pro | Lys | Thr | Val | Glu | Ile | Pro | Ala | Glu | Val | Val | Lys |
|     |     |     | 245 |     |     |     |     |     | 250 |     |     |     |     | 255 |     |
| Gln | Pro | Lys | Asn | Thr | Asn | Leu | Ser | Asp | Leu | Glu | Lys | Arg | Val | Arg | Met |
|     |     |     | 260 |     |     |     |     | 265 |     |     |     |     | 270 |     |     |
| Ile | Thr | Gly | Ala | Ile | Ala | Lys | Asn | Asn | Leu | Ala | Ser | Arg | Phe | Gln | His |
|     |     | 275 |     |     |     |     | 280 |     |     |     |     |     | 285 |     |     |
| Gly | Asn | Glu | Ser | Pro | Leu | Asp | Met | Met | Lys | Arg | Ile | Gln | Ser | Glu | Ile |
|     | 290 |     |     |     |     | 295 |     |     |     |     | 300 |     |     |     |     |
| Thr | Ser | Asp | Glu | Thr | Ala | Asp | Leu | Trp | Gln | Asn | Lys | Leu | Glu | Ser | Met |
| 305 |     |     |     |     | 310 |     |     |     |     | 315 |     |     |     |     | 320 |
| Gly | Val | Val | Phe |     |     |     |     |     |     |     |     |     |     |     |     |

<210> 13  
 <211> 999  
 <212> DNA  
 <213> artificial sequence

<220>  
 <223> codon-optimized MSBI1

<400> 13  
 atgagcgacc tgatcgtgaa agacaatgcc ctgatgaacg cctcctacaa cctggcactg 60  
 gtcgaacaga gactgattct gctggctatc atcgaggcaa gggagaccgg caagggcatc 120  
 aacgccaatg accccctgac agtgcacgcc agctcctaca tcaaccagtt taatgtggag 180  
 cgccacaccg cctatcaggc cctgaaggac gcctgcaagg atctgtttgc cgggcagttc 240  
 agctaccagg agaagcggga gagaggcagg atcaacatca caagcagatg ggtgtcccag 300  
 atcggctata tggacgatac cgccacagtg gagatcatct ttgcaccagc agtgggtgcct 360  
 ctgatcacca ggctggagga gcagttcaca cagtacgaca tcgagcagat ctccggactg 420  
 tctagcgect acgccgtgcg catgtatgag ctgctgatct gttggcggtc taccggcaag 480  
 acacctatca tcgagctgga tgagttccgc aagcggatcg gcgtgctgga caccgagtac 540  
 accagaacag ataacctgaa gatgagagtg atcgagctgg ccctgaagca gatcaatgag 600  
 cacaccgata tcacagcctc ttatgagcag cacaagaagg gccgctgat caccggcttc 660  
 agctttaagt tcaagcacia gaagcagaac tctgacaaga caccaaagaa tagcgattcc 720  
 tctccccgga tcgtgaagca cagccagatc cctaccaaca tcgtgaagca gccagagaat 780  
 gccaaagatgt ccgacctgga gcacagggca tctagggatga caggcgagat catgagaaat 840  
 aggctgagcg atcggttcaa gcagggcgac gagtccgcca tcgatatgat gaagagaatc 900  
 cagtccgaga tcatcaccga cgccatcgcc gatcagtggg aatctaaact ggaagagttt 960  
 ggagtcgtgt ttggagcaca tcaccatcat catcactga 999

<210> 14  
 <211> 326  
 <212> PRT  
 <213> artificial sequence

<220>

<223> MSBI1 protein

<400> 14

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Ser | Asp | Leu | Ile | Val | Lys | Asp | Asn | Ala | Leu | Met | Asn | Ala | Ser | Tyr |
| 1   |     |     |     | 5   |     |     |     |     | 10  |     |     |     |     | 15  |     |
| Asn | Leu | Ala | Leu | Val | Glu | Gln | Arg | Leu | Ile | Leu | Leu | Ala | Ile | Ile | Glu |
|     |     |     | 20  |     |     |     |     | 25  |     |     |     |     | 30  |     |     |
| Ala | Arg | Glu | Thr | Gly | Lys | Gly | Ile | Asn | Ala | Asn | Asp | Pro | Leu | Thr | Val |
|     |     | 35  |     |     |     |     | 40  |     |     |     |     | 45  |     |     |     |
| His | Ala | Ser | Ser | Tyr | Ile | Asn | Gln | Phe | Asn | Val | Glu | Arg | His | Thr | Ala |
|     | 50  |     |     |     |     | 55  |     |     |     |     | 60  |     |     |     |     |
| Tyr | Gln | Ala | Leu | Lys | Asp | Ala | Cys | Lys | Asp | Leu | Phe | Ala | Arg | Gln | Phe |
| 65  |     |     |     |     | 70  |     |     |     |     | 75  |     |     |     | 80  |     |
| Ser | Tyr | Gln | Glu | Lys | Arg | Glu | Arg | Gly | Arg | Ile | Asn | Ile | Thr | Ser | Arg |
|     |     |     |     | 85  |     |     |     |     | 90  |     |     |     |     | 95  |     |
| Trp | Val | Ser | Gln | Ile | Gly | Tyr | Met | Asp | Asp | Thr | Ala | Thr | Val | Glu | Ile |
|     |     |     | 100 |     |     |     |     | 105 |     |     |     |     | 110 |     |     |
| Ile | Phe | Ala | Pro | Ala | Val | Val | Pro | Leu | Ile | Thr | Arg | Leu | Glu | Glu | Gln |
|     |     | 115 |     |     |     |     | 120 |     |     |     |     | 125 |     |     |     |
| Phe | Thr | Gln | Tyr | Asp | Ile | Glu | Gln | Ile | Ser | Gly | Leu | Ser | Ser | Ala | Tyr |
|     | 130 |     |     |     |     | 135 |     |     |     |     | 140 |     |     |     |     |
| Ala | Val | Arg | Met | Tyr | Glu | Leu | Leu | Ile | Cys | Trp | Arg | Ser | Thr | Gly | Lys |
| 145 |     |     |     |     | 150 |     |     |     |     | 155 |     |     |     |     | 160 |
| Thr | Pro | Ile | Ile | Glu | Leu | Asp | Glu | Phe | Arg | Lys | Arg | Ile | Gly | Val | Leu |
|     |     |     |     | 165 |     |     |     |     | 170 |     |     |     |     | 175 |     |
| Asp | Thr | Glu | Tyr | Thr | Arg | Thr | Asp | Asn | Leu | Lys | Met | Arg | Val | Ile | Glu |
|     |     |     | 180 |     |     |     |     | 185 |     |     |     |     | 190 |     |     |
| Leu | Ala | Leu | Lys | Gln | Ile | Asn | Glu | His | Thr | Asp | Ile | Thr | Ala | Ser | Tyr |
|     |     | 195 |     |     |     |     | 200 |     |     |     |     | 205 |     |     |     |
| Glu | Gln | His | Lys | Lys | Gly | Arg | Val | Ile | Thr | Gly | Phe | Ser | Phe | Lys | Phe |
|     | 210 |     |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     |
| Lys | His | Lys | Lys | Gln | Asn | Ser | Asp | Lys | Thr | Pro | Lys | Asn | Ser | Asp | Ser |
| 225 |     |     |     |     | 230 |     |     |     |     | 235 |     |     |     | 240 |     |
| Ser | Pro | Arg | Ile | Val | Lys | His | Ser | Gln | Ile | Pro | Thr | Asn | Ile | Val | Lys |
|     |     |     |     | 245 |     |     |     |     | 250 |     |     |     |     | 255 |     |
| Gln | Pro | Glu | Asn | Ala | Lys | Met | Ser | Asp | Leu | Glu | His | Arg | Ala | Ser | Arg |
|     |     |     | 260 |     |     |     |     | 265 |     |     |     |     | 270 |     |     |
| Val | Thr | Gly | Glu | Ile | Met | Arg | Asn | Arg | Leu | Ser | Asp | Arg | Phe | Lys | Gln |
|     |     | 275 |     |     |     |     | 280 |     |     |     |     | 285 |     |     |     |
| Gly | Asp | Glu | Ser | Ala | Ile | Asp | Met | Met | Lys | Arg | Ile | Gln | Ser | Glu | Ile |
|     | 290 |     |     |     |     | 295 |     |     |     |     | 300 |     |     |     |     |
| Ile | Thr | Asp | Ala | Ile | Ala | Asp | Gln | Trp | Glu | Ser | Lys | Leu | Glu | Glu | Phe |
| 305 |     |     |     |     | 310 |     |     |     |     | 315 |     |     |     |     | 320 |
| Gly | Val | Val | Phe | Gly | Ala |     |     |     |     |     |     |     |     |     |     |
|     |     |     |     | 325 |     |     |     |     |     |     |     |     |     |     |     |

<210> 15

<211> 974

<212> DNA

<213> artificial sequence

<220>

<223> MSBI1 wild type

<400> 15

|                                                                    |     |
|--------------------------------------------------------------------|-----|
| atgagcgaatt taatagtaaa agataacgcc ctaatgaatg ctagttataa cttagctttg | 60  |
| gttgaacaga ggttaattct attagcaatc atagaagcga gagaaacagg caaagggatt  | 120 |
| aatgccaatg atcctttaac agttcatgca agtagctata tcaatcaatt taacgtagaa  | 180 |
| aggcatacgg catatcaagc cctcaaagat gcttgtaaag acttgtttgc ccgtcaattc  | 240 |
| agttaccaag aaaagcgaga acgaggacga attaatatta caagtcgatg ggtttcgcaa  | 300 |
| attggctata tggacgatac agcaaccgtt gagattatct ttgccctgc ggttgctcct   | 360 |
| ctgattacac ggctagagga acagttcacc cagtagcata ttgagcaaatt tagcggttta | 420 |
| tcgagtgcatt atgctgttcg tatgtacgaa ctgctgattt gttggcgtag cacaggcaaa | 480 |
| acaccaatta ttgagctaga cgagtttaga aagcgaatag gtgttttaga tactgaatac  | 540 |
| actagaacag ataatttaaa gatgcgagtt attgaattag ccctaaaaca aatcaacgaa  | 600 |
| catacagaca tcacagcaag ctatgaacaa cacaaaaaag ggcgagtgat tacaggattc  | 660 |
| tcattcaagt ttaagcacia gaaacaaaac agcgataaaa cgccaaaaaa tagcgattct  | 720 |
| agcccacgta tcgtaaaaca tagtcaaatt cctccaacat tgtaaaacag cctgaaaacg  | 780 |
| ccaaaatgag cgatttagaa catagagcga gccgtgttac aggggaaata atgcgaaatc  | 840 |
| gtctgtcaga tcggttttaa caaggcgatg aatcagcaat cgacatgatg aaacgtattc  | 900 |
| aaagtgaaat aataaccgat gcaatagcag accagtggga aagcaaactg gaggagtttg  | 960 |
| gcgtgggtttt ttag                                                   | 974 |

<210> 16

<211> 24

<212> DNA

<213> Artificial Sequence

<220>

<223> primer

<400> 16

gaggacgaat taatattaca agtc

24

<210> 17  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> primer

<400> 17  
gttctcgctt ttcttggtaa 20

<210> 18  
<211> 19  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> primer

<400> 18  
ggattaatgc caatgatcc 19

<210> 19  
<211> 18  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> primer

<400> 19  
ctttgcctgt ttctctcg 18

## REFERENCES:

- Fung, K. Y.C. et al. (2015). "Blood-based protein biomarker panel for the detection of colorectal cancer", PLOS ONE, Vol. 10, No. 3, p. e0120425
- 5 Funk, M., et al. (2014). "Isolation of protein-associated circular DNA from healthy cattle serum". Genome Announc 2(4)
- Giraldo, R., et al. (2011). "RepA-WH1 prionoid: a synthetic amyloid proteinopathy in a minimalist host." Prion 5(2):60-64
- Gunst, K., et al. (2014). "Isolation of bacterial plasmid-related replication-associated circular DNA from a serum sample of a multiple sclerosis patient." Genome Announc  
10 2(4).
- Lamberto, I., et al. (2014). "Mycovirus-like DNA virus sequences from cattle serum and human brain and serum samples from multiple sclerosis patients." Genome Announc 2(4).
- Manuelidis L., 2011. "Nuclease resistant circular DNAs co-purify with infectivity in  
15 scrapie and CJD". J. Neurovirol. 17:131-145.
- Torreira, E., et al. (2015). "Amyloidogenesis of bacterial prionoid RepA-WH1 recaptulates dimer to monomer transitions of RepA in DNA replication initiation." Structure 23(1):183-189
- Whitley, C., et al. (2014). "Novel replication-competent circular DNA molecules from  
20 healthy cattle serum and milk and multiple sclerosis-affected human brain tissue." Genome Announc 2(4).
- zur Hausen, H., Bund, T., de Villiers, E.-M. (2017). „Infectious agents in bovine red meat and milk and their potential role in cancer and other chronic diseases." Curr. Top. Microbiol. Immunol., Volume 407, 83-116.

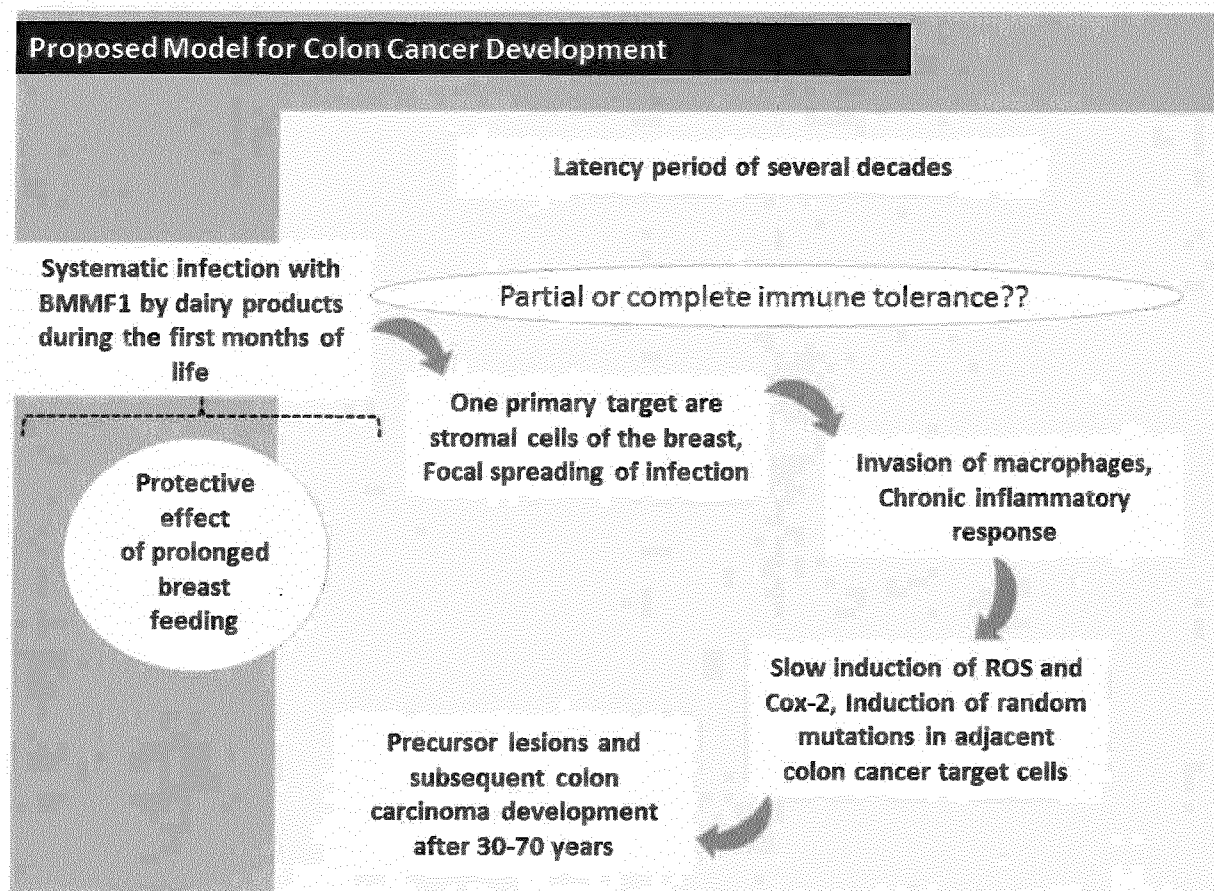
## CLAIMS

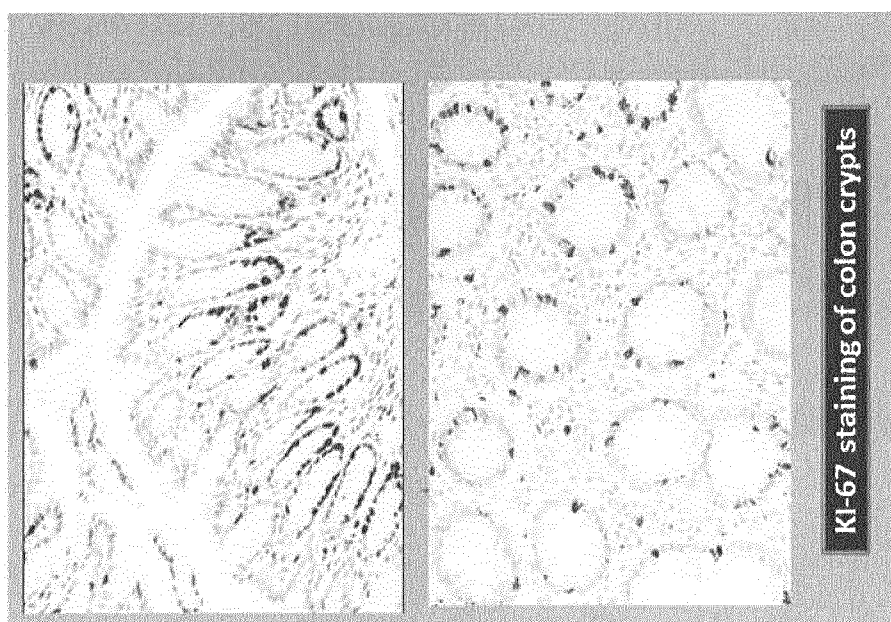
- 1) Use of Bovine Meat and Milk Factor (BMMF) of Group 1 Rep Protein as a biomarker for colorectal cancer.
- 2) The use of claim 1 wherein the Rep protein is a MSBI1 genome-encoded Rep protein (MSBI1 Rep), a MSBI2 genome-encoded Rep protein (MSBI2 Rep), a CMI1 genome-encoded Rep protein (CMI1 Rep), a CMI2 genome-encoded Rep protein (CMI2 Rep) or CMI3 genome-encoded Rep protein (CMI3 Rep).
- 3) A method for providing a diagnosis or predisposition for colorectal cancer (CRC) in a subject, comprising the step of detecting Rep protein in a sample from a subject by anti-Rep antibodies that bind to an epitope comprised by SEQ ID NO:2 or SEQ ID NO:3.
- 4) The method of claim 3, wherein the antibody specific for Rep protein binds to an epitope that is within an amino acid sequence selected from the group consisting of amino acids from 1 to 136, from 137 to 229 and from 230 to 324 of SEQ ID NO:1.
- 5) The method of claim 3 or 4, wherein the sample from a subject is selected from the group consisting of a cancerous colon tissue, peripheral tissue surrounding the cancerous colon tissue and colon polyps.
- 6) The method of any of claims 3 to 5, wherein additionally CD68 positive cells are detected in the sample by an anti-CD68 antibody.
- 7) A method for providing a prognostic score on a CRC patient's survival time, comprising the steps of

detecting Rep protein in a sample from peritumoral regions of CRC patients by anti-Rep antibodies that bind to an epitope comprised by SEQ ID NO:2 or SEQ ID NO:3,

determining for each antibody the percentage of stained cells and intensity proposing a immunohistological score which serves as general indicator of CRC-specific patient survival time.

Fig. 1



**Fig. 2**

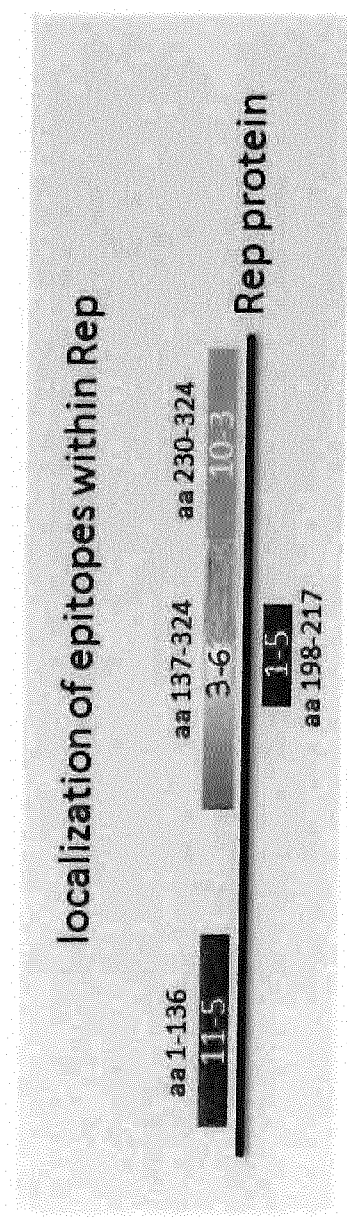
**Fig. 3**

Fig. 4

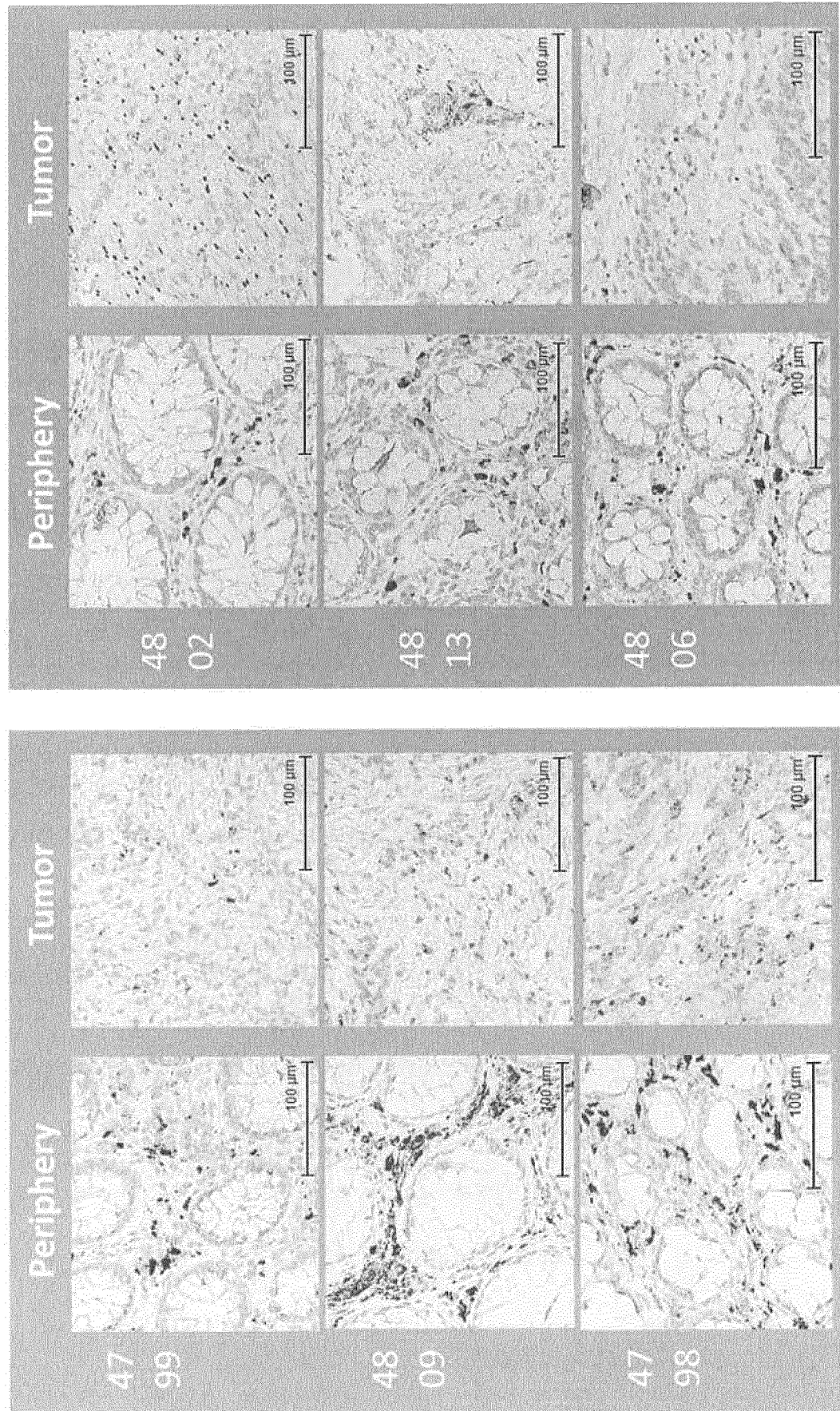
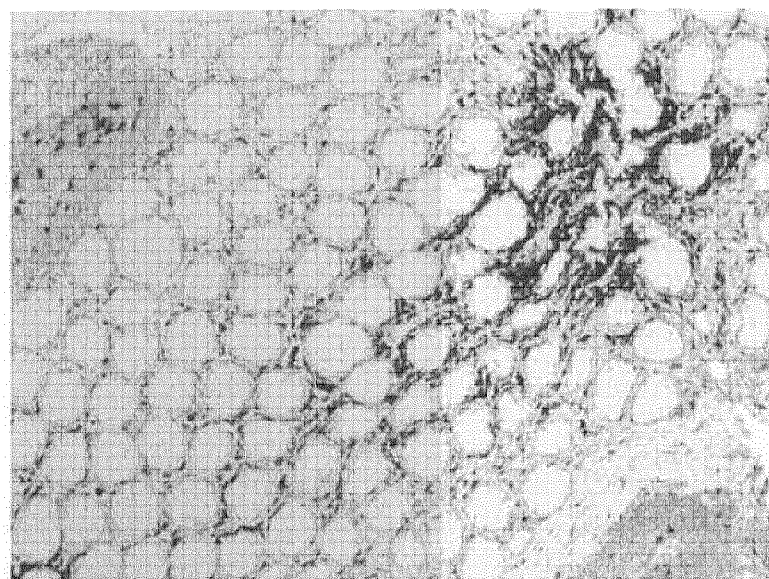
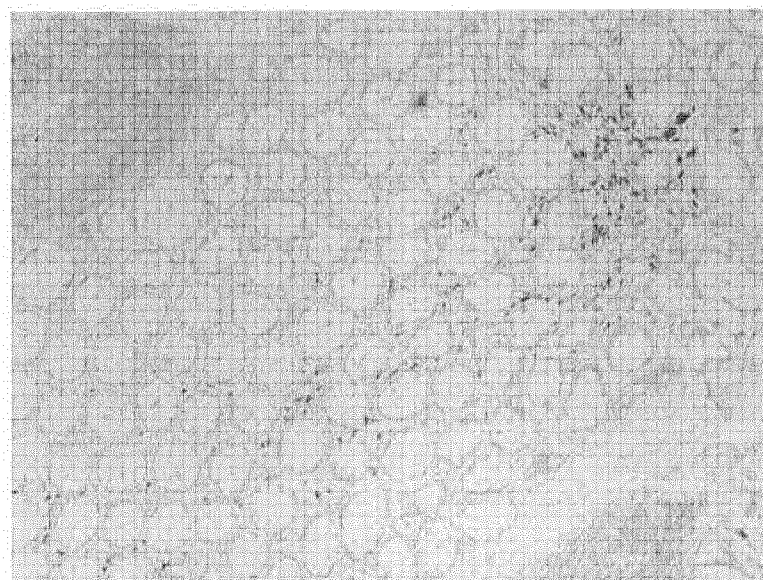


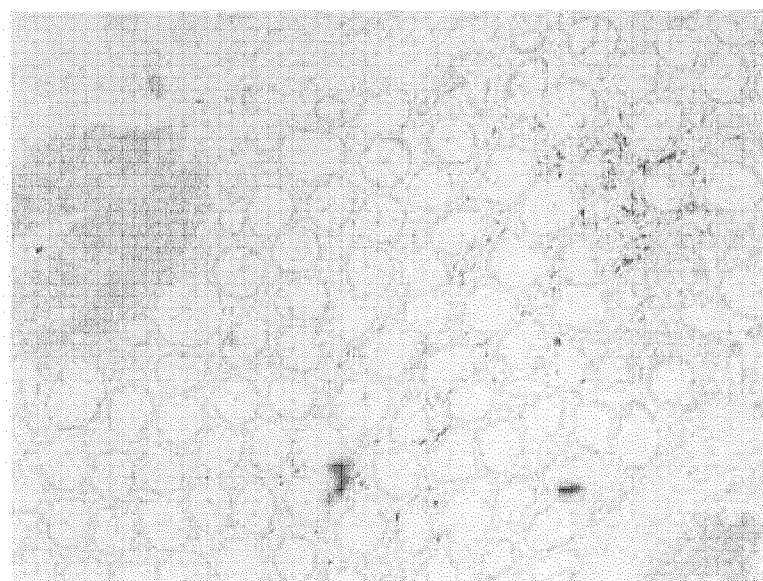
Fig. 5



anti-  
CD68

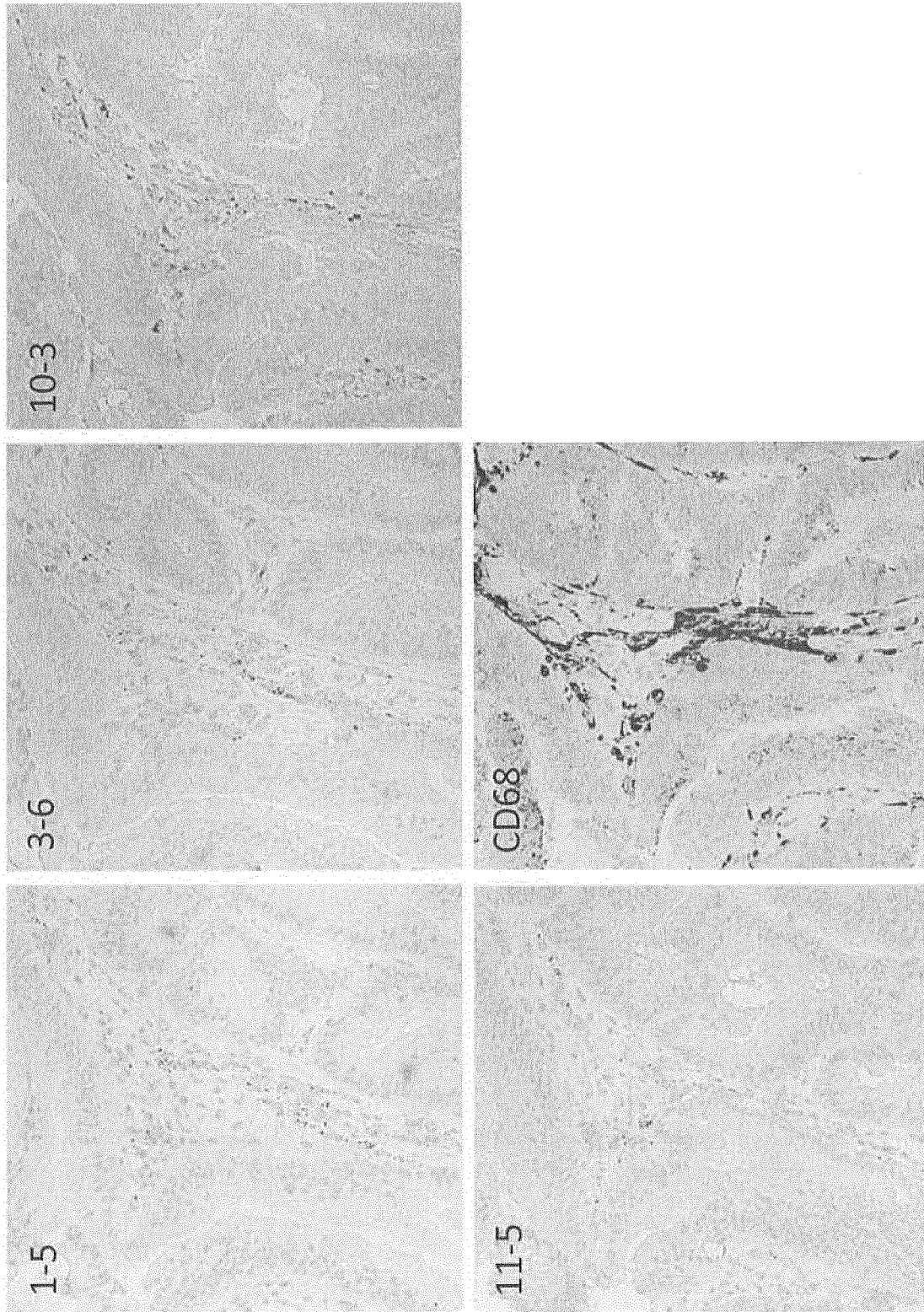


anti-Rep  
mAb 10-3



anti-Rep  
mAb 3-6

Fig. 6

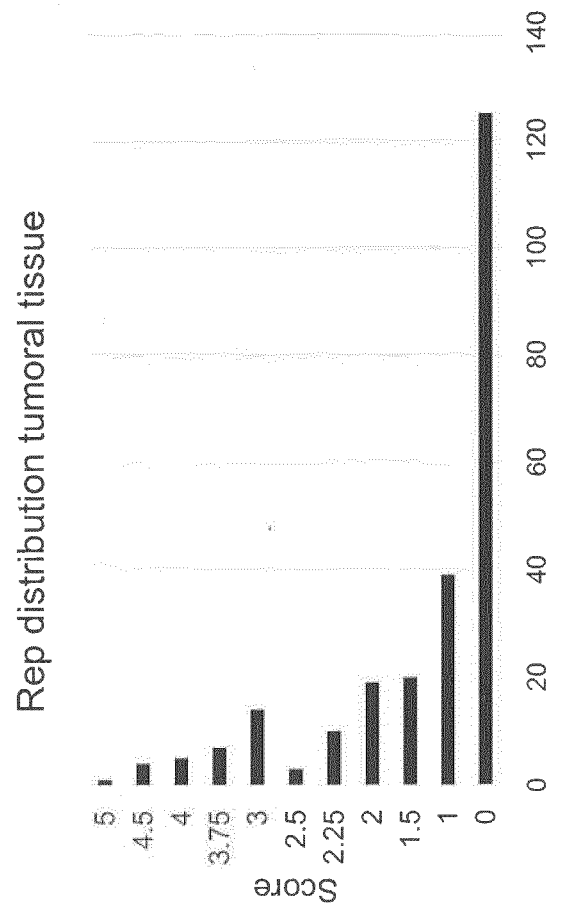


5

10

15

Fig. 7



5

Fig. 8

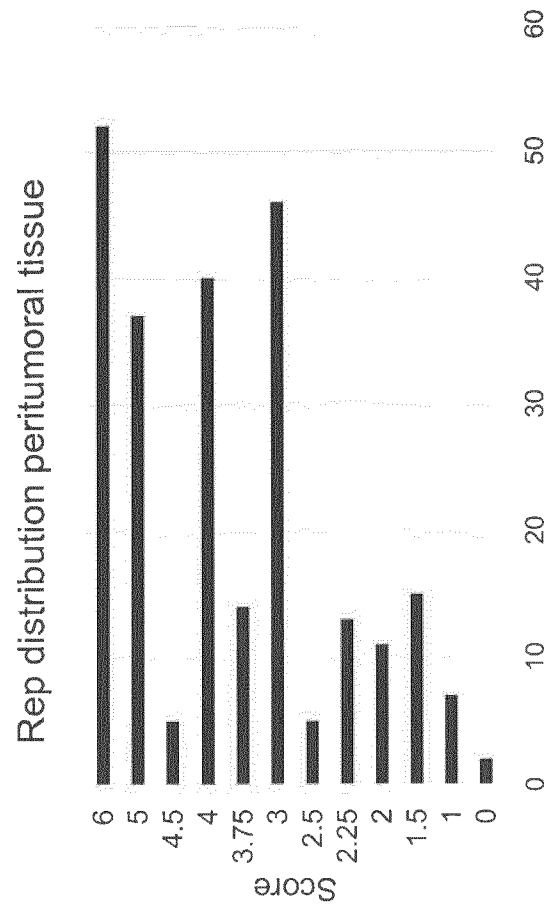


Fig. 9:

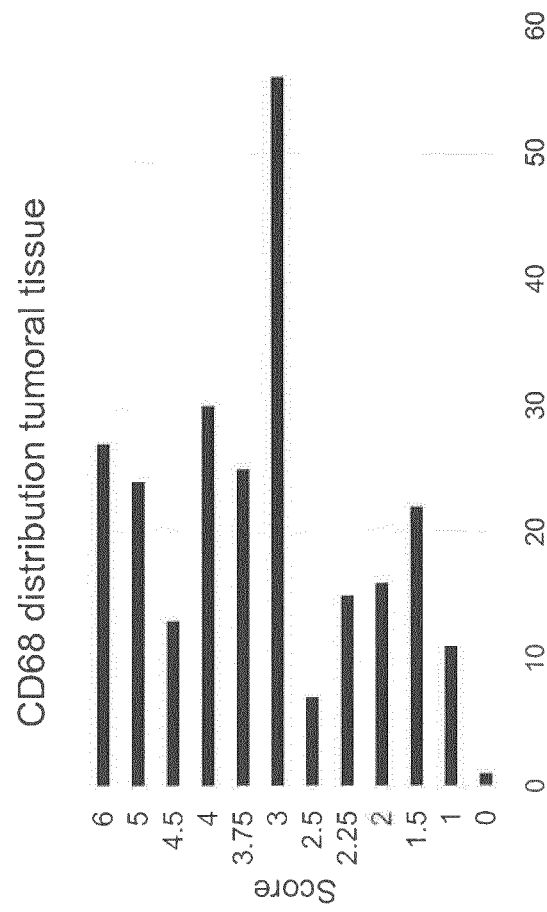


Fig. 10

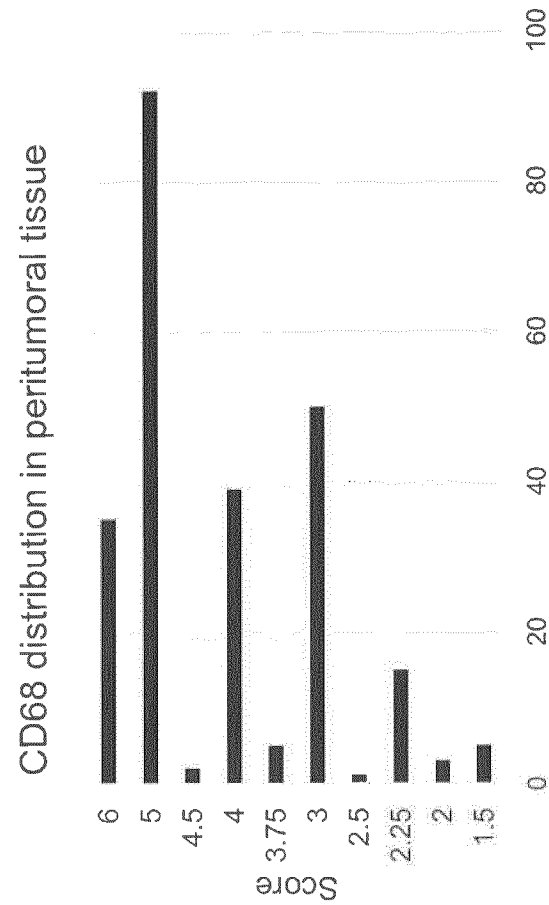


Fig. 11

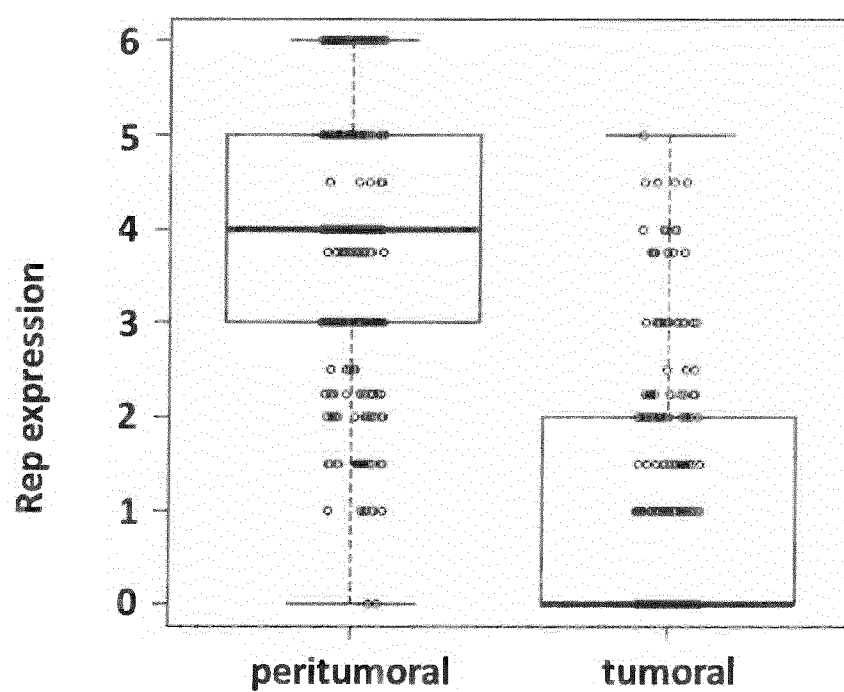


Fig. 12

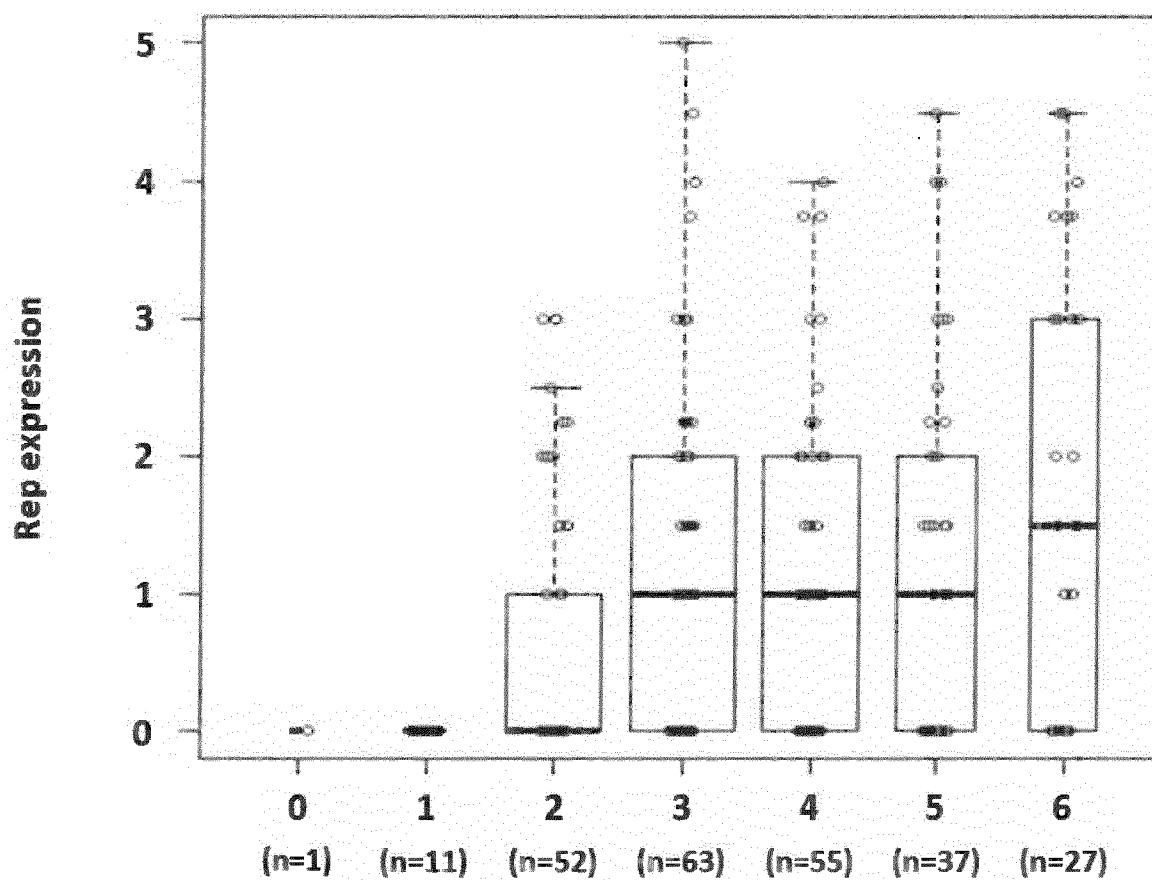


Fig. 13

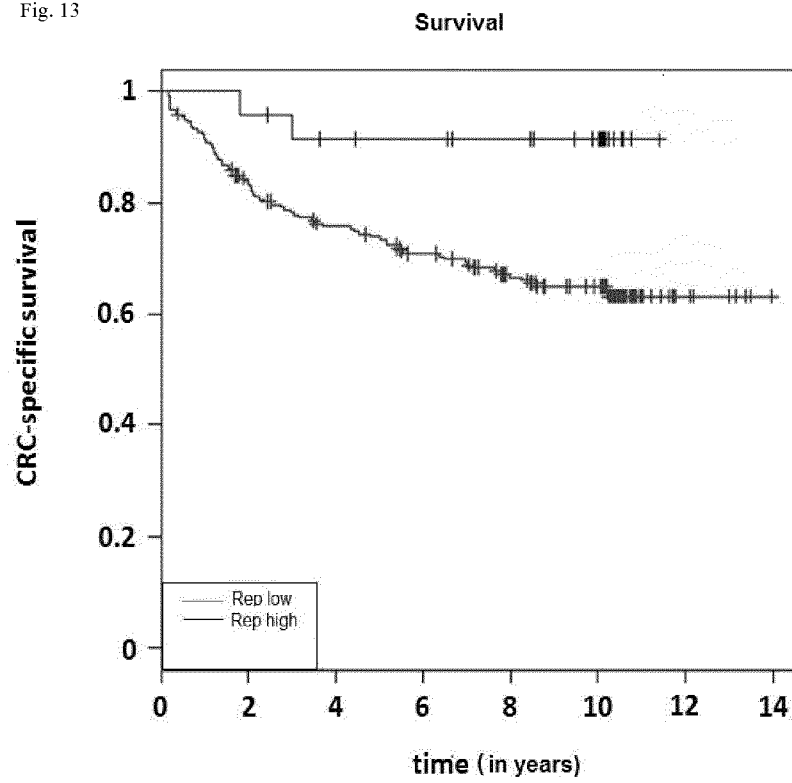


Fig. 14

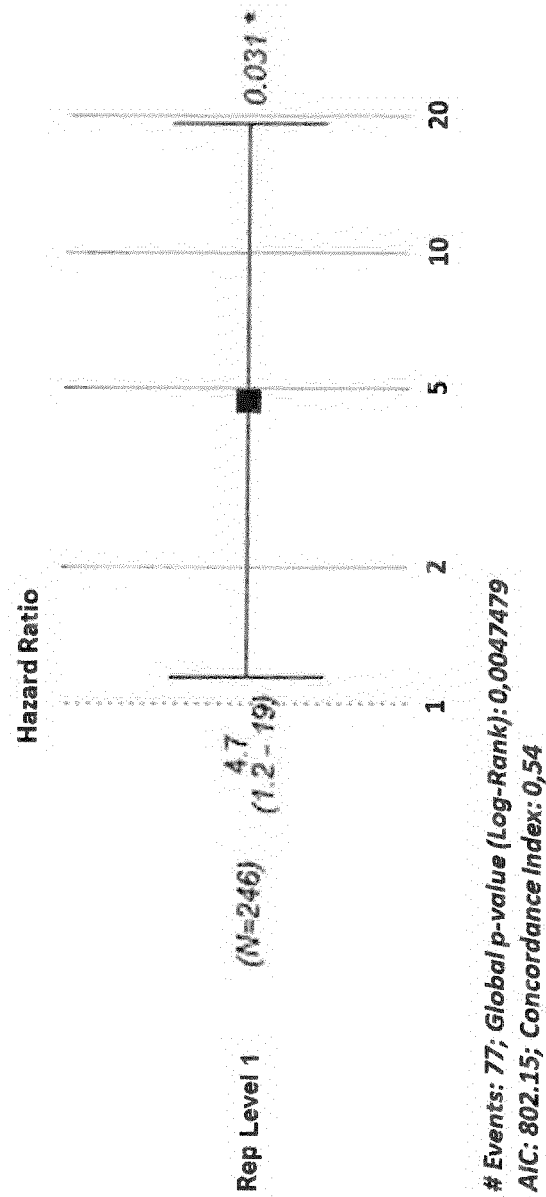


Fig. 15

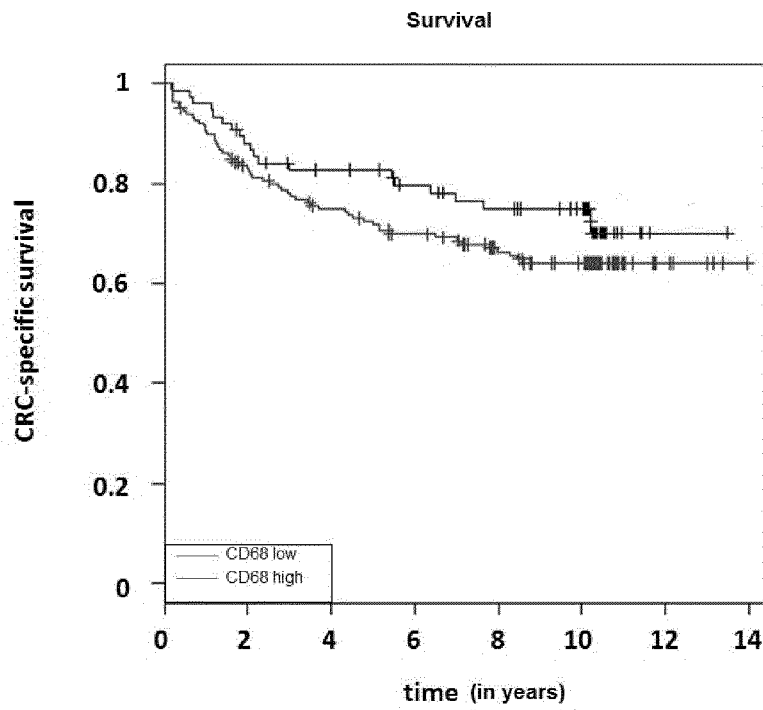


Fig. 16

