portion of frameshift peptides covered will be different (i.e. each cancer sample contains a different subset of frameshift mutations and, consequently, a different subset of frameshift peptides out of the optimal collection will be present), the optimal collection is selected in such a way that the fraction of frameshift peptides covered in each tumor sample will have a total length of at least 400 amino acids. Incorporating this 400 amino acid rule into the disclosed selection method thus ensures induction of T cells agaist FSP present in tumors of the vaccine target population.

SUMMARY ~~OF THE INVENTION~~

In a first aspect, the present ~~invention~~disclosure relates to a method of selecting a collection of frame-shift peptides (CFSP) for producing a universal micro-satellite instability (MSI) cancer vaccine peptide collection (CVP) for prophylaxis or treatment of a patient with a cancer comprising hereditary and sporadic MSI cancer or being at risk of developing such a cancer, comprising the steps of:

(i) selecting a collection of nucleic acids (CFSM), each comprising a frame-shift mutation (FSM), each FSM being present in one or more of at least M cancer samples (CS), each of a different patient, wherein the cancer of the patient comprises cancer cells with an MSI phenotype;

wherein at least 50% of the FSM that are selected fulfill criteria (a), (b), (c) and/or (d):

(a) the FSM is present in a mononucleotide repeat (MNR) of coding genes with a length equal to or longer than 6 nucleotides;

(b) the FSM corresponds to a deletion of 1 nucleotide;

(c) the number of DNA sequencing reads harboring the FSM is significantly higher in the tumor sample as compared to the matched normal sample (FDR-corrected Fisher test p-value equal to or lower than 0.05);

(d) the FSM is present in the matched normal samples with an allele frequency lower than 25%,

(ii) selecting X different frame-shift peptides (FSPs), wherein each selected FSP is the complete translation product of the protein-encoding segment of a nucleic acid comprising a FSM of the CFSM of at least 4 amino acid length, starting from the codon encoding the first amino acid that is different relative to the translation product of the corresponding wild type (wt) nucleic acid without the FSM,

wherein X is at least 20 more preferably at least 35 and M is at least 5.

In another aspect, the present disclosure relates to a method of modifying the amino acid sequences of peptides comprised in a cancer vaccine peptide collection (CVP) or of the nucleic acid sequences encoding the peptides comprised in the CVP, comprising the steps of:

(a) selecting at least Y frame-shift peptides (FSPs) or at least 8 amino acid long fragments thereof from the collection of frame-shift peptides (CFSP) selected according to the method as described herein;

(b) modifying the amino acid sequence of one or more or of all of those FSPs which fulfill the following criteria: (i) the FSP has a length of between 4 to 9 amino acids, and/or (ii) the FSP contains one or more identical contiguous stretches of 8 or more amino acids present in more than one FSP encoded by the same frame-shift mutation (FSM), and/or (iii) the FSP contains one or more contiguous stretches of 8 or more amino acids also present in wt human proteins,

wherein the amino acid sequence of a FSP according to (i) is modified by adding to the N-terminus of the FSP between 1 to 4 amino acids of the wild type (wt) amino acid sequence present immediately upstream of the FSP and wherein the modified FSP (mFSP) has a length of at least 8 amino acids; the amino acid sequence of a FSP according to (ii) is modified by removing these contiguous stretches from all but the longest FSP with the proviso that FSPs with a length of less than 4 amino acids after removal of the contiguous stretch are excluded from the CVP; and/or the amino acid sequence of a FSP according to (iii) is modified by removing these stretches; modified FSPs with a length of less than 4 amino acids after removal of the contiguous stretch are excluded from the CVP,

wherein the amino acid sequences of the CVP comprise the amino acid sequences of the FSPs or fragments thereof selected in step (a) and/or modified in step (b);

and wherein Y is at least 20, more preferably at least 35.

In a second ~~aspect~~example, the present ~~invention~~disclosure relates to a method of determining the amino acid sequences of peptides comprised in a CVP or of the nucleic acid sequences encoding the peptides comprised in the CVP, comprising the steps of:

(a) selecting at least Y FSPs or at least 8 amino acid long fragments thereof from the CFSP selected according to the first ~~aspect~~example of the ~~invention~~disclosure;

(b) modifying the amino acid sequence of one or more or of all of those FSPs which fulfill the following criteria: (i) the FSP has a length of between 4 to 9 amino acids, and/or (ii) the FSP contains one or more identical contiguous stretches of 8 or more amino acids present in more than one FSP encoded by the same FSM and/or (iii) the FSP contains one or more contiguous stretches of 8 or more amino acids also present in wt human proteins,

wherein the amino acid sequence of an FSP according to (i) is modified by adding to the N-terminus of the FSP between 1 to 4 amino acids of the wild type (wt) amino acid sequence present immediately upstream of the FSP and wherein the modified FSP (mFSP) has a length of at least 8 amino acids; the amino acid sequence of a FSP according to (ii) is modified by removing these contiguous stretches from all but the longest FSP with the proviso that FSPs with a length of less than 4 amino acids after removal of the contiguous stretch are excluded from the CVP; and/or the amino acid sequence of a FSP according to (iii) is modified by removing these stretches; modified FSPs with a length of less than 4 amino acids after removal of the contiguous stretch are excluded from the CVP, and

wherein the amino acid sequences of the CVP comprises the amino acid sequences of the FSPs or fragments thereof selected in step a) and/or modified in step (b);

and wherein Y is at least 20 more preferably at least 35.

In another aspect, the present disclosure relates to a method of producing a CVP or a collection of nucleic acids encoding the CVP comprising the steps of

(i) obtaining the amino acid or nucleic acid sequence modified in the method as described herein; and

(ii) synthesizing the amino acid sequence of the CVP in one or more polypeptides or a collection of nucleic acids with that sequence and optionally inserting the collection of nucleic acids into one or more expression cassettes and/or a collection of expression vectors.

In a third ~~aspect~~example, the present ~~invention~~disclosure relates to a method of producing a CVP or a collection of nucleic acids encoding the CVP comprising the steps of

(i) obtaining the amino acid or nucleic acid sequence information determined in the method of the second ~~aspect~~example of the ~~invention~~disclosure; and

(ii) synthesizing the amino acid sequence of the CVP in one or more polypeptides or a collection of nucleic acids with that sequence and optionally inserting the collection of nucleic acids into one or more expression cassettes and/or a collection of expression vectors.

In a fourth ~~aspect~~example, the present ~~invention~~disclosure relates to a CVP or a collection of nucleic acids encoding the peptides of said CVP producible by the method of the third ~~aspect~~example of the ~~invention~~disclosure.

In a fifth ~~aspect~~example, the CVP comprising or consisting of Y different FSPs and/or mFSPs, wherein each FSP or the FSP which is modified to mFSP is a fragment of or the complete

translation product of the protein-encoding segment of a FSM containing nucleic acid starting from the codon encoding the first amino acid that is different relative to the translation of the corresponding wt nucleic acid without the FSM of in either case at least 4 amino acid length and wherein at least 50% of the FSPs or the FSPs which are modified to mFSPs fulfill one or more of the following criteria:

(a) the FSP is encoded by a FSM that is observed with a cancer type-specific frequency (CF) observed for a subset of CS of a specific type of cancer that are part of a collection of M different CS that is at least 5% for at least one of the cancer types present in the CS; and/or

(b) the average mRNA expression level of the gene with the FSM encoding the FSP is in the top 80th percentile of the distribution describing the average mRNA expression values of each protein-encoding gene across the CS; and/or

(c) the FSM generating the FSP is observed in less than 2% in normal tissues in a cohort of subjects without a cancer

wherein the amino acid sequence of one or more or all those FSPs which fulfill the following criteria: (i) the FSP has a length of between 4 to 9 amino acids, and/or (ii) the FSP contains one or more identical contiguous stretches of 8 or more amino acids present in more than one FSP encoded by the same FSM and/or (iii) the FSP contains one or more contiguous stretches of 8 or more amino acids also present in wt human proteins,

is modified for a FSP according to (i) by adding to the N-terminus of the FSP between 1 to 4 amino acids of the wild type (wt) amino acid sequence present immediately upstream of the FSP and wherein the modified FSP (mFSP) has a length of at least 8 amino acids; for a FSP according to (ii) is modified by removing these contiguous stretches from all but the longest FSP with the proviso that FSPs with a length of less than 4 amino acids after removal of the contiguous stretch are excluded from the CVP; and/or for a FSP according to (iii) is modified by removing these stretches; modified FSPs with a length of less than 4 amino acids after removal of the contiguous stretch are excluded from the CVP;

and wherein Y is at least 20 and more preferably at least 35 and M is at least 5.

In another aspect, the present disclosure relates to a cancer vaccine peptide collection (CVP) comprising the 209 frame-shift peptides (FSPs) and modified FSPs (mFSPs) according to SEQ ID NO: 1 to 209, preferably comprising four polypeptides with amino acid sequences according to

A) SEQ ID NO: 1088 to 1091,

B) SEQ ID NO: 1092 to 1095,

C) SEQ ID NO: 1155 to 1158, or

D) SEQ ID NO: 1159 to 1162.

In yet another aspect, the present disclosure relates to a collection of nucleic acids encoding the cancer vaccine peptide collection (CVP) as described herein. In yet another aspect, the present disclosure relates to a collection of expression vectors each comprising all or part of the collection of nucleic acids as described herein, wherein the entirety of the collection of expression vectors comprise all of the collection of nucleic acids as described herein, wherein preferably at least one of the expression vectors comprises one or more elements that enhance immunogenicity of the expression vector.

In a sixth ~~aspect~~example, the present ~~invention~~disclosure relates to a nucleic acid collection encoding the CVP of the fifth ~~aspect~~example of the present ~~invention~~disclosure.

In a seventh ~~aspect~~example, the present ~~invention~~disclosure relates to a collection of one or more expression vectors, each comprising all or part of the nucleic acid collection of the fourth or sixth ~~aspect~~example of the present ~~invention~~disclosure, wherein the entirety of the collection of expression vectors comprise all of the nucleic acid collection of the fourth or sixth ~~aspect~~example of the present ~~invention~~disclosure.

In a further aspect, the present disclosure relates to a use of a cancer vaccine peptide collection (CVP) as described herein, a collection of nucleic acids as described herein, or an expression vector collection as described herein, in the manufacture of a medicament for the prophylaxis or treatment of a cancer in a patient, wherein the cancer comprises cancer cells with a micro-satellite instability (MSI) phenotype or being at risk of developing such cancer, wherein prophylaxis is preferably for patients known to be at risk of developing MSI cancers, including patients with germline mutations in genes involved in the mismatch repair system (MMR) such as MLH-1, MSH-2, MSH-6, PMS2 and TACSTD1/EPCAM and wherein treatment is for patients with cancers at all stages (I-IV), arising in any tissue after diagnosis of MSI status and wherein the use of such a vaccine is intended for treatment of cancers with an MSI status that can be either spontaneous or pharmacologically induced and wherein the cancer is preferably selected from the group consisting of colorectal cancer, gastric cancer, endometrial cancer, small intestine cancer, hepatobiliary tract cancer, liver cancer, neuroendocrine cancers, cervical cancer, ovarian cancer, uterine sarcomas, brain cancer and skin cancer.

In a further aspect, the present disclosure relates to a use of a collection of nucleic acids as described herein and/or the expression vector collection as described herein in the manufacture of a medicament for the prophylaxis or treatment of a cancer in a patient, wherein the cancer comprises cancer cells with a micro-satellite instability (MSI) phenotype or being at risk of developing such cancer, wherein the collection of nucleic acids and/or the expression vector collection is to be administered in a heterologous prime-boost vaccination scheme, preferably the prime is with an adenovirus vector and the one or more boosts with a poxviral vector, preferably an MVA vector and wherein prophylaxis is preferably for patients known to be at risk of developing MSI cancers, including patients with germline mutations in genes involved in the mismatch repair system (MMR) such as MLH-1, MSH-2, MSH-6, PMS2 and TACSTD1/EPCAM and wherein treatment is for patients with cancers at all stages (I-IV), arising in any tissue after diagnosis of MSI status and wherein the use of such a vaccine is intended for treatment of cancers with an MSI status that can be either spontaneous or pharmacologically induced and wherein the cancer is preferably selected from the group consisting of colorectal cancer, gastric cancer, endometrial cancer, small intestine cancer, hepatobiliary tract cancer, liver cancer, neuroendocrine cancers, cervical cancer, ovarian cancer, uterine sarcomas, brain cancer and skin cancer.

In an eighth ~~aspect~~example, the present ~~invention~~disclosure relates to a CVP of the fourth or fifth ~~aspect~~example of the present ~~invention~~disclosure, a nucleic acid collection of the fourth or sixth ~~aspect~~example of

the present ~~invention~~disclosure, an expression vector collection of the seventh ~~aspect~~example of the present ~~invention~~disclosure, for use in prophylaxis or treatment of a patient with a cancer comprising cancer cells with an MSI phenotype or being at risk of developing such cancer, wherein the cancer is preferably selected from the group consisting of colorectal cancer, gastric cancer, endometrial cancer, small intestine cancer, hepatobiliary tract cancer, liver cancer, neuroendocrine cancers, cervical cancer, ovarian cancer, uterine sarcomas, brain cancer and skin cancer

In an eight ~~aspect~~example, the present ~~invention~~disclosure relates to a nucleic acid collection of the fourth or sixth ~~aspect~~example of the present ~~invention~~disclosure and/or a expression vector collection of the seventh ~~aspect~~example of the present ~~invention~~disclosure for use in prophylaxis or treatment of a patient with a cancer comprising cancer cells with an MSI phenotype or being at risk of developing such cancer, wherein the nucleic acid collection and/or the expression vector collection is administered in a heterologous prime-boost vaccination scheme, preferably the prime is with an adenovirus vector and one or more boosts are with an MVA vector.

FIGURE LEGENDS

Fig. 1: Antigen length determines the number of vaccine-induced immunogenic epitopes.

Fig. 2: Immunogenic coverage of Nous-209 in MSI cell lines.

Fig. 3: Immunogenic coverage of Nous-209 in MSI CRC biopsies.

Fig. 4: Number of predicted MHC-I binding epitopes arising from Nous-209 in biopsies of MSI patients.

Fig. 5: The FSP corresponding to SEQ ID NO: 123 included in Nous-209 is immunogenic *in vivo* in HLA.A02 transgenic mice. A representative CD8 T cell response in 5 animals to a HLA-A02 nonamer present in FSP SEQ ID NO: 123, as measured by intracellular staining (ICS) for interferon γ (IFN-γ) (panel A). FACS plot of gating strategy for IFN-γ+ CD8 T cells from one of those mice showing a significant percentage (5.6%) of FSP-reactive T cells (panel B).

Fig. 6: Disorder profiles predicted by IUPRED for the eight artificial polypeptides containing Nous-209 FSPs in layout A and B.

Fig. 7: Schematic view of the expression cassettes.

Fig. 8: Immunogenicity of GAd20-209-FSPs/MVA-209-FSPs prime/boost regimen in mice. IFNγ ELISpot responses are measured 2 weeks post GAd20-209-FSPs prime(GAd) and one-week post MVA-209-FSPs boost (GAd/MVA). Shown are the responses

Claims

1. A method~~Method~~ of selecting a collection of frame-shift peptides (CFSP) for producing a universal micro-satellite instability (MSI) cancer vaccine peptide collection (CVP) for prophylaxis or treatment of a patient with a cancer comprising hereditary and sporadic MSI cancer or being at risk of developing such a cancer, comprising the steps of:

(i) selecting a collection of nucleic acids (CFSM) each comprising a frame-shift mutation (FSM) each FSM being present in one or more of at least M cancer samples (CS) each of a different patient, wherein the cancer of the patient comprises cancer cells with an MSI phenotype;

wherein at least 50% of the FSM that are selected fulfill criteria (a), (b), (c) and (d):

(a) the FSM is present in a mononucleotide repeat (MNR) of coding genes with a length equal to or longer than 6 nucleotides;

(b) the FSM corresponds to a deletion of 1 nucleotide;

(c) the number of DNA sequencing reads harboring the FSM is significantly higher in the tumor sample as compared to the matched normal sample (FDR-corrected Fisher test p-value equal to or lower than 0.05);

(d) the FSM is present in the matched normal samples with an allele frequency lower than 25%,

(ii) selecting X different frame-shift peptides (FSPs), wherein each selected FSP is the complete translation product of the protein-encoding segment of a nucleic acid comprising a FSM of the CFSM of at least 4 amino acid length, starting from the codon encoding the first amino acid that is different relative to the translation product of the corresponding wild type (wt) nucleic acid without the FSM,

wherein X is at least 20 more preferably at least 35 and M is at least 5.

2. The method of claim 1 wherein at least 50% of the FSPs that are selected fulfill one or more of the following criteria:

(a) the FSP is encoded by a FSM that is observed with a cancer type specific frequency (CF) observed for a subset of CS of a specific type of cancer that are part of the collection of M different CSs that is at least 5% for at least one of the cancer types present in the CS; and/or

(b) the average mRNA expression level of the gene with the FSM encoding the FSP is in the top 80th percentile of the distribution describing the average mRNA expression values of each protein-encoding gene across the CS; and/or

(c) the FSM generating the FSP is observed in less than 2% in normal tissues in a cohort of subjects without a cancer;.

and/or

3. The method of claim 1 or 2, wherein:

(i) the CS are derived from patients having a MSI tumor, preferably colorectal cancer and/or gastric cancer and/or endometrial cancer, more preferably colorectal cancer, gastric cancer and endometrial cancer; and/or

(ii) M is at least 10, preferably at least 10, preferably at least 30, preferably at least 50, more preferably at least 100, more preferably at least 200, most preferably at least 300; and/or

(iii) X is at least 35, preferably at least 50, preferably at least 100, more preferably at least 200.

34. A method of modifyingdetermining the amino acid sequences of peptides comprised in a cancer vaccine peptide collection (CVP) or of the nucleic acid sequences encoding the peptides comprised in the CVP, comprising the steps of:

(a) selecting at least Y frame-shift peptides (FSPs) or at least 8 amino acid long fragments thereof from the collection of frame-shift peptides (CFSP) selected according to any of claims 1 to 23;

(b) modifying the amino acid sequence of one or more or of all of those FSPs which fulfill the following criteria: (i) the FSP has a length of between 4 to 9 amino acids, and/or (ii) the FSP contains one or more identical contiguous stretches of 8 or more amino acids present in more than one FSP encoded by the same frame-shift mutation (FSM), and/or (iii) the FSP contains one or more contiguous stretches of 8 or more amino acids also present in wt human proteins,

wherein the amino acid sequence of a FSP according to (i) is modified by adding to the N-terminus of the FSP between 1 to 4 amino acids of the wild type (wt) amino acid sequence present immediately upstream of the FSP and wherein the modified FSP (mFSP) has a length of at least 8 amino acids; the amino acid sequence of a FSP according to (ii) is modified by removing these contiguous stretches from all but the longest FSP with the proviso that FSPs with a length of less than 4 amino acids after removal of the contiguous stretch are excluded from the CVP; and/or the amino acid sequence of a FSP according to (iii) is modified by removing these stretches;

modified FSPs with a length of less than 4 amino acids after removal of the contiguous stretch are excluded from the CVP,

wherein the amino acid sequences of the CVP comprise~~s~~ the amino acid sequences of the FSPs or fragments thereof selected in step (a) and/or modified in step (b);

~~-~~and wherein Y is at least 20, more preferably at least 35.

~~4~~5. The method of claim ~~3~~4, wherein

~~-~~in step (a) the FSPs of the CVP are successively selected from the CFSP and in each selection step a new FSP is selected from the CFSP that increases the cumulative amount of total FSP amino acid length (CAFSPL) to reach a threshold value (TV) in the maximal number of cancer samples which are still below that threshold value and, optionally, if more than one FSP increases the CAFSPL for the maximal number of cancer samples for which the CAFSPL is still below the threshold value, the FSP with the highest score is selected;~~.~~

preferably wherein

(a) the TV is at least 400 amino acids, preferably at least 600 amino acids, more preferably at least 800 amino acids; or

(b) the TV is:

(i) at least 400 amino acids, preferably at least 600 amino acids, more preferably at least 800 amino acids for colorectal and gastric cancer; and

(ii) at least 200 amino acids, preferably at least 300 amino acids, more preferably at least 400 amino acids for endometrial cancer.

~~5~~6. The method of claim ~~3~~4 or ~~4~~5, wherein

(i) the CAFSPL for each cancer sample is determined by summing up the amino acid length of the FSPs that are already part of the CVP and the new FSP from the CFSP for which the corresponding FSMs are present in the cancer sample; and/or

(ii) wherein the threshold value is defined separately for each subset of samples out of the cancer sample (CS) that belong to a particular cancer type; and/or

(iii) wherein the score is defined as the product between the amino acid length of the FSP and the overall frequency with which the FSM generating the FSP is observed in the CS, optionally without counting the FSM in the CS from a cancer type for which the cancer type specific frequency (CF) is below 5%, and/or

(iv) wherein the subset of cancer samples comprises all cancer samples from tumor types where the FSM is present with a CF greater than or equal to 5%; and/or

(v) wherein FSPs generated by a FSM with an overall frequency lower than 5% are excluded from the selection; and/or

(vi) wherein addition of new FSPs and/or modified FSPs proceeds until no further FSP is available whose inclusion in the CVP would increase the CAFSPL of any cancer sample whose CAFSPL is still below a threshold value or the cumulative length of all FSPs present in the CVP has reached a maximal value of V amino acids; and/or

(vii) wherein FSPs originating from the same FSM are treated as one FSP with a combined score calculated as the sum of the scores of the individual FSPs.

67. The method of claim 45 or 56, wherein

addition of new FSPs proceeds until no further FSP is available whose inclusion in the CVP would increase the CAFSPL of any cancer sample whose CAFSPL is still below TV; and/or

wherein FSPs originating from the same FSM are treated as one FSP with a combined score calculated as the sum of the scores of the individual FSPs.

~~8. The method of claim 6 or 7, wherein:~~

~~(a) the TV is at least 400 amino acids, preferably at least 600 amino acids, more preferably at least 800 amino acids; or~~

~~(b) the TV is:~~

~~(i) at least 400 amino acids, preferably at least 600 amino acids, more preferably at least 800 amino acids for colorectal and gastric cancer; and~~

~~(ii) at least 200 amino acids, preferably at least 300 amino acids, more preferably at least 400 amino acids for endometrial cancer.~~

79. The method of any one of claims 34 to 68, wherein:

(i) the CVP comprises at least 4 amino acids of each FSP selected from the CFSP; and/or

(ii) Y is selected in such that the cumulative amino acid length V of all peptides that are part of the CVP is at least 280 amino acids, preferably at least 6000;

(iii) Y is at least 35, preferably at least 50, preferably at least 100, more preferably at least 200, and/or

(iv) the CVP comprises FSPs and/or mFSPs selected from the group of FSPs and/or mFSPs according to SEQ ID NO: 1 to 1087, preferably selected from the group of FSPs and/or mFSPs according to SEQ ID NO: 1 to 209.

8~~10~~. A method of producing a CVP or a collection of nucleic acids encoding the CVP comprising the steps of

(i) obtaining the amino acid or nucleic acid sequence modified~~information determined~~ in the method of any one of claims 3~~4~~ to 7~~10~~; and

(ii) synthesizing the amino acid sequence of the CVP in one or more polypeptides or a collection of nucleic acids with that sequence and optionally inserting the collection of nucleic acids into one or more expression cassettes and/or a collection of expression vectors.

~~11. A CVP or a collection of nucleic acids encoding the peptides of the CVP producible by the method of claim 10.~~

~~12. A CVP comprising or consisting of Y different FSPs and/or mFSPs or fragments thereof with a length of at least 8 amino acids, wherein each FSP or the FSPs which are modified to mFSP is a fragment of or the complete translation product of the protein-encoding segment of a FSM containing nucleic acid starting from the codon encoding the first amino acid that is different relative to the translation of the corresponding wt nucleic acid without the FSM of in either case at least 4 amino acid length and wherein at least 50% of the FSPs or the FSPs which are modified to mFSPs fulfill one or more of the following criteria:~~

~~(a) the FSP is encoded by a FSM that is observed with a cancer type specific frequency (CF) observed for a subset of CS of a specific type of cancer that are part of a collection of M different CSs that is at least 5% for at least one of the cancer types present in the CS; and/or~~

~~(b) the average mRNA expression level of the gene with the FSM encoding the FSP is in the top 80th percentile of the distribution describing the average mRNA expression values of each protein-encoding gene across the CS; and/or~~

(c) the FSM generating the FSP is observed in less than 2% in normal tissues in a cohort of subjects without a cancer

wherein the amino acid sequence of one or more or all those FSPs which fulfill the following criteria: (i) the FSP encoded by a nucleic acid comprising a FSM starting from the amino acid encoded by the FC has a length between 4 to 9 amino acids, and/or (ii) the FSP contains one or more identical contiguous stretches of 8 or more amino acids present in more than one FSP encoded by the same FSM and/or (iii) the FSP contains one or more contiguous stretches of 8 or more amino acids also present in wt human proteins,

are modified for a FSP according to (i) by adding to the N-terminus of the FSP between 1 to 4 amino acids of the wild type (wt) amino acid sequence present immediately upstream of the FSP and wherein the modified FSP (mFSP) has a length of at least 8 amino acids; for a FSP according to (ii) is modified by removing these contiguous stretches from all but the longest FSP with the proviso that FSPs with a length of less than 4 amino acids after removal of the contiguous stretch are excluded from the CVP; and/or for a FSP according to (iii) is modified by removing these stretches; modified FSPs with a length of less than 4 amino acids after removal of the contiguous stretch are excluded from the CVP;

and wherein Y is at least 20, more preferably at least 35 and M is at least 5.

13. The CVP of claim 12, wherein:

(i) the CS are derived from patients having a MSI tumor, preferably colorectal cancer and/or gastric cancer and/or endometrial cancer, more preferably colorectal cancer, gastric cancer and endometrial cancer; and/or

(ii) M is at least 10, preferably at least 20, preferably at least 30, preferably at least 50, more preferably at least 100, more preferably at least 200, most preferably at least 300; and/or

(iii) Y is at least 35, preferably 50, preferably at least 100, more preferably at least 200; and/or

(iv) the CVP comprises at least 4 amino acids of each FSP; and/or

(v) Y is selected in such that the cumulative amino acid length V of all peptides that are part of the CVP is at least 280 amino acids, preferably at least 6000; and/or

(vi) the FSPs and/or mFSPs of the CVP are selected from the group of FSPs and/or mFSPs according to SEQ ID NO: 1 to 1087, preferably selected from the group of FSPs and/or mFSPs according to SEQ ID NO: 1 to 209.

~~14. The CVP of claim 12 or 13, wherein the FSPs of the CVP have been successively selected from a CFSP and wherein in each selection step a new FSP is selected from the CFSP to increase the CAFSPL to reach a threshold value in the maximal number of cancer samples which are below that threshold value and optionally, if more than one FSP increases the CAFSPL for the maximal number of cancer samples for which the CAFSPL is still below the threshold value, the FSP with the highest score is selected.~~

~~15. The CVP of claim 14, wherein:~~

~~(i) the CAFSPL for each cancer sample is determined by summing up the amino acid length of the FSPs that are already part of the CVP and the new FSP from the CFSPP for which the corresponding FSMs are present in the cancer sample; and/or~~

~~(ii) the threshold value is defined separately for each subset of samples out of the CS that belong to a particular cancer type; and/or~~

~~(iii) the score is defined as the product between the amino acid length of the FSP and the overall frequency with which the FSM generating the FSP is observed in the CS; and/or~~

~~(iv) addition of new FSPs proceeds until no further FSP is available whose inclusion in the CVP would increase the CAFSPL of any cancer sample whose CAFSPL is still below a threshold value (TV); and/or~~

~~(v) wherein FSPs originating from the same FSM are treated as one FSP with a combined score calculated as the sum of the scores of the individual FSPs.~~

~~16. The CVP of any of claims 12 to 15,~~

~~(i) wherein the CAFSPL for each cancer sample is determined by summing up the amino acid length of the FSPs that are already part of the CVP and the new FSP from the CFSP for the corresponding FSM are present in the cancer sample; and/or~~

~~(ii) wherein the threshold value is defined separately for each subset of samples out of the CS that belong to a particular cancer type; and/or~~

~~(iii) wherein the score is defined as the product between the amino acid length of the FSP and the overall frequency of the FSM generating the FSP without counting the FSM in cancer samples from a cancer type for which CF is below 5%; and/or~~

~~(iv) wherein the subset of cancer samples comprises all cancer samples from tumor types where the FSM is present with a CF greater than or equal to 5%; and/or~~

~~(v) wherein FSPs generated by a FSM with an overall frequency lower than 5% are excluded from the selection; and/or~~

~~(vi) wherein addition of new FSPs proceeds until no further FSP is available whose inclusion in the CVP would increase the CAFSPL of any cancer sample whose CAFSPL is still below a TV or the cumulative length of all FSPs present in the CVP has reached a maximal value of V amino acids; and/or~~

~~(vii) and wherein FSPs originating from the same FSM are treated as one FSP with a combined score calculated as the sum of the scores of the individual FSPs.~~

~~17. The CVP of any of claims 12 to 16, wherein the TV is:~~

~~(a) the TV is at least 400 amino acids, preferably at least 600 amino acids, more preferably at least 800 amino acids; or~~

~~(b) the TV is:~~

~~(i) at least 400 amino acids, preferably at least 600 amino acids, more preferably at least 800 amino acids for colorectal and gastric cancer; and~~

~~(ii) at least 200 amino acids, preferably at least 300 amino acids, more preferably at least 400 amino acids for endometrial cancer.~~

~~18. The CVP of any of claims 12 to 17, wherein the peptides of the CVP are separate or at least two are comprised in one or more polypeptides.~~

9~~19~~. ~~The~~ A cancer vaccine peptide collection (CVP) ~~of any of claims 12 to 18, wherein the CVP consists of or comprises the~~ comprising the 209 frame-shift peptides (FSPs) and modified FSPs (mFSPs) according to SEQ ID NO: 1 to 209, preferably comprising four polypeptides with amino acid sequences according to

A) SEQ ID NO: 1088 to 1091,

B) ~~(layout A),~~ SEQ ID NO: 1092 to 1095,

C) ~~(layout B),~~ SEQ ID NO: 1155 to 1158, ~~(layout C)~~ or

D) SEQ ID NO: 1159 to 1162. ~~(layout D).~~

10. A collection of nucleic acids encoding the cancer vaccine peptide collection (CVP) of claim 9.

11. A collection of expression vectors each comprising all or part of the collection of nucleic acids of claim 10, wherein the entirety of the collection of expression vectors comprise all of the collection of nucleic acids of claim 10, wherein preferably at least one of the expression vectors comprises one or more elements that enhance immunogenicity of the expression vector.

~~20. The CVP of any of claims 12 to 19, wherein one or more peptides of the CVP are linked, preferably by peptide bonds, to one or more of the following elements that enhance immunogenicity of the~~ 12. The collection of expression vectors of claim 11, wherein at least one of the expression vector comprises one or more nucleic acids encoding an element that enhances immunogenicity of the cancer vaccine peptide collection (CVP): an invariant chain sequence or fragment thereof; a tissue-type plasminogen activator; a PEST sequence; a cyclin destruction box; an ubiquitination signal; a SUMOylation signal; an Interleukin, preferably an Interleukin 2, Interleukin 12, or Interleukin 15; a checkpoint protein specific ligand, preferably an anti-PD1 antibody or PD1-binding fragment thereof, an anti-CTLA4 antibody or an anti-CTLA4-binding fragment thereof, an anti-LAG3 antibody or an anti-LAG3-binding fragment, an anti-TIM3 antibody or an anti-TIM3-binding fragment thereof.

~~21. A collection of nucleic acids encoding the CVP of any of claims 12 to 20.~~

~~22. A collection of one or more expression vectors each comprising all or part of the collection of nucleic acids of claim 10 or 20, wherein the entirety of the collection of expression vectors comprise all of the collection of nucleic acids of claim 10 or 20.~~

~~23. The collection of expression vectors of claim 21 wherein at least one of the expression vector comprises one or more elements that enhance immunogenicity of the expression vector.~~

~~24. The collection of expression vectors of claim 22 or 23, wherein at least one of the expression vector comprises one or more nucleic acids encoding an element that enhances immunogenicity of the CVP: an invariant chain sequence or fragment thereof; a tissue-type plasminogen activator; a PEST sequence; a cyclin destruction box; an ubiquitination signal; a SUMOylation signal; an Interleukin, preferably an Interleukin 2, Interleukin 12, or Interleukin 15; a checkpoint protein specific ligand, preferably an anti-PD1 antibody~~

~~or PD1-binding fragment thereof, an anti-CTLA4 antibody or an anti-CTLA4-binding fragment thereof, an anti-LAG3 antibody or an anti-LAG3-binding fragment, an anti-TIM3 antibody or an anti-TIM3-binding fragment thereof.~~

1325. The collection of expression vectors of ~~any of~~ claim~~s~~ ~~22~~ 11 or ~~to 24~~12, wherein each expression vector is selected from the group consisting of a plasmid; a cosmid; an RNA; an RNA-formulated with an adjuvant; an RNA formulated in liposomal particles; a self-amplifying RNA (SAM); a SAM formulated with an adjuvant; a SAM formulated in liposomal particles; a viral vector; preferably an alphavirus vector, a venezuelan equine encephalitis (VEE) virus vector, a sindbis (SIN) virus vector, a semliki forest virus (SFV) virus vector, also preferably a replication competent or incompetent adenoviral vector preferably derived from chimpanzee or bonobo or gorilla, a poxvirus vector, a vaccinia virus vector or a modified vaccinia ankara (MVA) vector, a simian or human cytomegalovirus (CMV) vector, a Lymphocyte choriomeningitis virus (LCMV) vector, a retroviral or lentiviral vector.

~~26~~14. ~~A~~Use of a cancer vaccine peptide collection (CVP) of claim 9~~any of claims 11 to 20~~, a collection of nucleic acids of claim 10,~~11~~ or ~~21,~~ an expression vector collection of any of claims 11 to 13, in the manufacture of a medicament~~22 to 25,~~ for the~~use in~~ prophylaxis or treatment of a cancer in a patient, wherein the ~~with a~~ cancer compris~~es~~~~ing~~ cancer cells with a micro-satellite instability (~~an~~ MSI) phenotype or being at risk of developing such cancer, wherein prophylaxis is preferably for patients known to be at risk of developing MSI cancers ~~according to clinical guidelines~~, including patients with germline mutations in genes involved in the mismatch repair system (MMR) such as MLH-1, MSH-2, MSH-6, PMS2 and TACSTD1/EPCAM and wherein treatment is for patients with cancers at all stages (I-IV), arising in any tissue after diagnosis of MSI status ~~according most recent clinical guidelines~~ and wherein the use of such a vaccine is intended for treatment of cancers with an MSI status that can be either spontaneous or pharmacologically induced and wherein the cancer is preferably selected from the group consisting of colorectal cancer, gastric cancer, endometrial cancer, small intestine cancer, hepatobiliary tract cancer, liver cancer, neuroendocrine cancers, cervical cancer, ovarian cancer, uterine sarcomas, brain cancer and skin cancer.

~~27~~15. Use of a~~A~~ collection of nucleic acids of claim 10~~11 or 21~~ and/or the expression vector collection of any of claims 11~~22~~ to 13 in the manufacture of a medicament~~25~~ for the~~use in~~ prophylaxis or treatment of a cancer in a patient, wherein the ~~with a~~ cancer compris~~es~~~~ing~~ cancer cells with a micro-satellite instability (~~an~~ MSI) phenotype or being at risk of developing such cancer, wherein the collection of nucleic acids and/or the expression vector collection is to be administered in a heterologous prime-boost vaccination scheme, preferably the prime is with an adenovirus vector and the one or more boosts with a poxviral vector, preferably an MVA vector and wherein prophylaxis is preferably for patients known to be at risk of developing MSI cancers ~~according to clinical guidelines~~, including patients with germline mutations in genes involved in the mismatch repair system (MMR) such as MLH-1, MSH-2, MSH-6, PMS2 and TACSTD1/EPCAM and wherein treatment is for patients with cancers at all stages (I-IV), arising in any tissue after diagnosis of MSI status ~~according most recent clinical guidelines~~ and wherein the use of such a vaccine is intended for treatment of cancers with an MSI status that can be either spontaneous or pharmacologically induced and wherein the cancer is preferably selected from the group consisting of colorectal cancer, gastric cancer, endometrial cancer, small intestine cancer, hepatobiliary tract cancer, liver cancer, neuroendocrine cancers, cervical cancer, ovarian cancer, uterine sarcomas, brain cancer and skin cancer.