

As provided herein multiple HLAs expressed by an individual need to present the same peptide in order to trigger a T cell response. The fragments of a polypeptide antigen that are immunogenic for a specific individual are those that can bind to multiple class I (activate cytotoxic T cells) or class II (activate helper T cells) HLAs expressed by that individual. For example, the inventors have discovered that the presence of a T cell epitope that binds to at least three HLA type I of a subject predicts an immune response in the subject to a polypeptide.

Based on this discovery the inventors have identified the T cell epitopes from certain breast, ovarian and/or colorectal cancer associated-polypeptide antigens (cancer testis antigens (CTA)) that are capable of binding to at least three class I HLA in a high proportion of individuals. These T cell epitopes, or fragments of the antigens comprising the T cell epitopes, are useful for inducing specific immune responses against tumor cells expressing these antigens and for treating or preventing cancer.

In one aspect, the disclosure refers to a polypeptide that

(a) comprises a fragment of up to 50 consecutive amino acids of a colorectal cancer-associated antigen selected from TSP50, EpCAM, SPAG9, CAGE1, FBXO39, SURVIVIN, LEMD1, MAGE-A8, MAGE-A6 and MAGE-A3, wherein the fragment comprises an amino acid sequence selected from any one of SEQ ID NOs: 21 to 34 and 234 to 243;

(b) comprises a fragment of up to 50 consecutive amino acids of an ovarian cancer-associated antigen selected from PIWIL-4, WT1, EpCAM, BORIS, AKAP-4, OY-TES-1, SP17, PIWIL-2, PIWIL-3, SPAG9, PRAME, HIWI, SURVIVIN, and AKAP-3 wherein the fragment comprises the amino acid sequence of any one of SEQ ID NOs: 272 to 301;

(c) comprises a fragment of up to 50 consecutive amino acids of a breast cancer associated antigen selected from PIWIL-2, AKAP-4, EpCAM, BORIS, HIWI, SPAG9, PLU-1, TSGA10, ODF-4, SP17, RHOXF-2, PRAME, NY-SAR-35, MAGE-A9, NY-BR-1, SURVIVIN, MAGE-A11, HOM-TES-85 and NY-ESO-1 wherein the fragment comprises an amino acid sequence selected from any one of SEQ ID NOs: 1 to 17, 24 and 172 to 192; and/or

(d) comprises or consists of two or more fragments of one or more cancer-associated antigens selected from

(i) TSP50, EpCAM, SPAG9, CAGE1, FBXO39, SURVIVIN, MAGE-A8, MAGE-A6, MAGE-A3 and LEMD1;

(ii) PIWIL-4, WT1, EpCAM, BORIS, AKAP-4, OY-TES-1, SP17, PIWIL-2, PIWIL-3, SPAG9, PRAME, HIWI, SURVIVIN and AKAP-3; and/or

(iii) SPAG9, AKAP-4, BORIS, NY-SAR-35, NY-BR-1, SURVIVIN, MAGE-A11, PRAME, MAGE-A9, HOM-TES-8, PIWIL-2, EpCAM, HIWI, PLU-1, TSGA10, ODF-4, SP17, RHOXF-2;

wherein each fragment comprises a different amino acid sequence selected from SEQ ID NOs: 21 to 38 and 234 to 244; SEQ ID NOs: 272 to 301; and/or SEQ ID NOs: 1 to 20, 24 and 172 to 192.

In another aspect, the disclosure refers to a polypeptide or a panel of polypeptides consisting of one or more amino acid sequences selected from SEQ ID NOs: 41-80, 251 to 271, 302 to 331 and 195 to 233.

In a further aspect, the disclosure refers to a panel of two or more polypeptides, wherein each polypeptide comprises a fragment of up to 50 consecutive amino acids of a cancer-associated antigen selected from:

(a) TSP50, EpCAM, SPAG9, CAGE1, FBXO39, SURVIVIN, MAGE-A8, MAGE-A6, MAGE-A3 and LEMD1;

(b) PIWIL-4, WT1, EpCAM, BORIS, AKAP-4, OY-TES-1, SP17, PIWIL-2, PIWIL-3, SPAG9, PRAME, HIWI, SURVIVIN and AKAP-3; and/or

(c) SPAG9, AKAP-4, BORIS, NY-SAR-35, NY-BR-1, SURVIVIN, MAGE-A11, PRAME, MAGE-A9, HOM-TES-8, PIWIL-2, EpCAM, HIWI, PLU-1, TSGA10, ODF-4, SP17, RHOXF-2;

and wherein each fragment comprises a different amino acid sequence selected from:

(i) SEQ ID NOs: 21 to 38 and 234 to 244;

(ii) SEQ ID NOs: 272 to 301; and/or

(iii) SEQ ID NOs: 1 to 20, 24 and 172 to 192

optionally wherein the fragments are flanked at the N and/or C terminus by additional amino acids that are not part of the sequence of the breast, ovarian or colorectal cancer-associated antigens.

5 In yet another aspect, the disclosure refers to a pharmaceutical composition or kit having one or more polypeptides as described herein, or a panel of polypeptides as described herein, or a polypeptide comprising at least two amino acid sequences selected SEQ ID NOs: 21 to 38 and 234 to 244; SEQ ID NOs: 272 to 301; and/or SEQ ID NOs: 1 to 20, 24 and 172 to 192 as an active ingredient.

10 In yet another aspect, the disclosure refers to a pharmaceutical composition as described herein for use in therapy for vaccination, providing immunotherapy or inducing a cytotoxic T cell response in a subject, wherein the pharmaceutical composition is to be administered to the subject, optionally wherein the therapy is prophylactic.

15 In yet another aspect, the disclosure refers to the use of the pharmaceutical composition as described herein in the manufacture of a medicament for treating cancer, optionally colorectal cancer, ovarian cancer or breast cancer.

20 In yet another aspect, the disclosure refers to a method of identifying a human subject who is predicted to have a cytotoxic T cell response to administration of a pharmaceutical composition as disclosed herein, the method comprising (i) determining that the active ingredient polypeptide(s) of the pharmaceutical composition comprise a sequence that is a T cell epitope capable of binding to at least three HLA class I molecules of the subject; and (ii) identifying the subject as being predicted to have a cytotoxic T cell response to administration of the pharmaceutical composition.

25 In yet another aspect, the disclosure refers to a method of identifying a subject who is predicted to have a clinical response to the use as described herein, the method comprising (i) determining that the active ingredient polypeptide(s) of the pharmaceutical composition comprise two or more different amino acid sequences each of which is

- a. a T cell epitope capable of binding to at least three HLA class I molecules of the subject; and

- b. a fragment of a cancer-associated antigen expressed by cancer cells of the subject, optionally wherein the cancer-associated antigen is present in a sample obtained from the subject; and

(ii) identifying the subject as predicted to have a clinical response to the use.

5 In yet another aspect, the disclosure refers to a method of determining the probability that a specific human subject will have a clinical response to the use as described herein, wherein one or more of the following factors corresponds to a higher probability of a clinical response:

(a) presence in the active ingredient polypeptide(s) of a higher number of amino acid sequences and/or different amino acid sequences that are each a T cell epitope capable of
10 binding to at least three HLA class I of the subject;

(b) a higher number of target polypeptide antigens, comprising at least one amino acid sequence that is both

A. comprised in an active ingredient polypeptide; and

B. a T cell epitope capable of binding to at least three HLA class I of the subject;
15 optionally wherein the target polypeptide antigens are expressed in the subject, further optionally wherein the target polypeptides antigens are in one or more samples obtained from the subject;

(c) a higher probability that the subject expresses target polypeptide antigens, optionally a threshold number of the target polypeptide antigens and/or optionally target polypeptide
20 antigens that have been determined to comprise at least one amino acid sequence that is both

A. comprised in in an active ingredient polypeptide; and

B. a T cell epitope capable of binding to at least three HLA class I of the subject;
and/or

(d) a higher number of target polypeptide antigens that the subject is predicted to express, optionally a higher number of target polypeptide antigens that the subject expresses with a threshold probability, and/or optionally the target polypeptide antigens that have been
25 determined to comprise at least one amino acid sequence that is both

A. comprised in in an active ingredient polypeptide; and

B. a T cell epitope capable of binding to at least three HLA class I of the subject.
30

In yet another aspect, the disclosure refers to a method of identifying a human subject who is predicted not have a clinical response to the use as described herein, the method comprising

(i) determining that the active ingredient peptide(s) of the pharmaceutical composition do
35 not comprise two or more different amino acid sequences each of which is a T cell epitope capable of binding to at least three HLA class I molecules of the subject; and

(ii) identifying the subject as predicted not to have a clinical response to the use.

In a first example the disclosure provides a polypeptide that comprises a fragment of up to 50 consecutive amino acids of

(a) a colorectal cancer-associated antigen selected from TSP50, EpCAM, SPAG9, CAGE1, FBXO39, SURVIVIN, LEMD1, MAGE-A8, MAGE-A6 and MAGE-A3, wherein the fragment comprises an amino acid sequence selected from any one of SEQ ID NOs: 21 to 40 and 234 to 250;

(b) an ovarian cancer-associated antigen selected from PIWIL-4, WT1, EpCAM, BORIS, AKAP-4, OY-TES-1, SP17, PIWIL-2, PIWIL-3, SPAG9, PRAME, HIWI, SURVIVIN, and AKAP-3 wherein the fragment comprises the amino acid sequence of any one of SEQ ID NOs: 272 to 301; and/or

(c) a breast cancer associated antigen selected from PIWIL-2, AKAP-4, EpCAM, BORIS, HIWI, SPAG9, PLU-1, TSGA10, ODF-4, SP17, RHOXF-2, PRAME, NY-SAR-35, MAGE-A9, NY-BR-1, SURVIVIN, MAGE-A11, HOM-TES-85 and NY-ESO-1 wherein the fragment comprises an amino acid sequence selected from any one of SEQ ID NOs: 1 to 20, 24 and 172 to 194.

In some specific cases the disclosure provides a polypeptide that

(a) is a fragment of a colorectal cancer-associated antigen selected from TSP50, EpCAM, SPAG9, CAGE1, FBXO39, SURVIVIN, MAGE-A8, MAGE-A6, MAGE-A3 and LEMD1, wherein the fragment comprises an amino acid sequence selected from any one of SEQ ID NOs: 21 to 40 and 234 to 250; or

(b) comprises or consists of two or more fragments of one or more colorectal cancer associated antigens selected from TSP50, EpCAM, SPAG9, CAGE1, FBXO39, SURVIVIN, MAGE-A8, MAGE-A6, MAGE-A3 and LEMD1, wherein each fragment comprises a different amino acid sequence selected from any one of SEQ ID NOs: 21 to 40 and 234 to 250, optionally wherein the fragments overlap or are arranged end to end in the polypeptide; or

(c) is a fragment of a ovarian cancer-associated antigen selected from PIWIL-4, WT1, EpCAM, BORIS, AKAP-4, OY-TES-1, SP17, PIWIL-2, PIWIL-3, SPAG9, PRAME, HIWI, SURVIVIN and AKAP-3, wherein the fragment comprises an amino acid sequence selected from any one of SEQ ID NOs: 272 to 301; or

CLAIMS

1. A polypeptide that

(a) comprises a fragment of up to 50 consecutive amino acids of a colorectal cancer-associated antigen selected from TSP50, EpCAM, SPAG9, CAGE1, FBXO39, SURVIVIN, LEMD1, MAGE-A8, MAGE-A6 and MAGE-A3, wherein the fragment comprises an amino acid sequence selected from any one of SEQ ID NOs: 21 to 34 and 234 to 243;

(b) comprises a fragment of up to 50 consecutive amino acids of an ovarian cancer-associated antigen selected from PIWIL-4, WT1, EpCAM, BORIS, AKAP-4, OY-TES-1, SP17, PIWIL-2, PIWIL-3, SPAG9, PRAME, HIWI, SURVIVIN, and AKAP-3 wherein the fragment comprises the amino acid sequence of any one of SEQ ID NOs: 272 to 301;

(c) comprises a fragment of up to 50 consecutive amino acids of a breast cancer associated antigen selected from PIWIL-2, AKAP-4, EpCAM, BORIS, HIWI, SPAG9, PLU-1, TSGA10, ODF-4, SP17, RHOXF-2, PRAME, NY-SAR-35, MAGE-A9, NY-BR-1, SURVIVIN, MAGE-A11, HOM-TES-85 and NY-ESO-1 wherein the fragment comprises an amino acid sequence selected from any one of SEQ ID NOs: 1 to 17, 24 and 172 to 192; and/or

(d) comprises or consists of two or more fragments of one or more cancer-associated antigens selected from

(i) TSP50, EpCAM, SPAG9, CAGE1, FBXO39, SURVIVIN, MAGE-A8, MAGE-A6, MAGE-A3 and LEMD1;

(ii) PIWIL-4, WT1, EpCAM, BORIS, AKAP-4, OY-TES-1, SP17, PIWIL-2, PIWIL-3, SPAG9, PRAME, HIWI, SURVIVIN and AKAP-3; and/or

(iii) SPAG9, AKAP-4, BORIS, NY-SAR-35, NY-BR-1, SURVIVIN, MAGE-A11, PRAME, MAGE-A9, HOM-TES-8, PIWIL-2, EpCAM, HIWI, PLU-1, TSGA10, ODF-4, SP17, RHOXF-2;

wherein each fragment comprises a different amino acid sequence selected from SEQ ID NOs: 21 to 38 and 234 to 244; SEQ ID NOs: 272 to 301; and/or SEQ ID NOs: 1 to 20, 24 and 172 to 192.

2. The polypeptide of claim 1, wherein the polypeptide
 - a. is a fragment of a colorectal cancer-associated antigen selected from TSP50, EpCAM, SPAG9, CAGE1, FBXO39, SURVIVIN, MAGE-A8, MAGE-A6, MAGE-A3 and LEMD1, wherein the fragment comprises an amino acid sequence selected from any one of SEQ ID NOs: 21 to 34 and 234 to 243; or
 - b. is a fragment of an ovarian cancer-associated antigen selected from PIWIL-4, WT1, EpCAM, BORIS, AKAP-4, OY-TES-1, SP17, PIWIL-2, PIWIL-3, SPAG9, PRAME, HIWI, SURVIVIN and AKAP-3, wherein the fragment comprises an amino acid sequence selected from any one of SEQ ID NOs: 272 to 301; or
 - c. is a fragment of a breast cancer associated antigen selected from SPAG9, AKAP-4, BORIS, NY-SAR-35, NY-BR-1, SURVIVIN, MAGE-A11, PRAME, MAGE-A9, HOM-TES-85, PIWIL-2, EpCAM, HIWI, PLU-1, TSGA10, ODF-4, SP17, RHOXF-2, wherein the fragment comprises the amino acid sequence from any one of SEQ ID NOs: 1 to 17, 24 and 172 to 192.

3. The polypeptide according to claim 1 or claim 2, wherein the polypeptide comprises or consists of fragments of at least two different cancer-associated antigens, wherein the cancer-associated antigens are selected from
 - (a) TSP50, EpCAM, SPAG9, CAGE1, FBXO39, SURVIVIN, MAGE-A8, MAGE-A6, MAGE-A3 and LEMD1;
 - (b) PIWIL-4, WT1, EpCAM, BORIS, AKAP-4, OY-TES-1, SP17, PIWIL-2, PIWIL-3, SPAG9, PRAME, HIWI, SURVIVIN and AKAP-3; and/or
 - (c) SPAG9, AKAP-4, BORIS, NY-SAR-35, NY-BR-1, SURVIVIN, MAGE-A11, PRAME, MAGE-A9, HOM-TES-8, PIWIL-2, EpCAM, HIWI, PLU-1, TSGA10, ODF-4, SP17, RHOXF-2;
 wherein each fragment comprises a different amino acid sequence selected from SEQ ID NOs: 21 to 38 and 234 to 244; SEQ ID NOs: 272 to 301; and/or SEQ ID NOs: 1 to 20, 24 and 172 to 192.

4. A polypeptide or a panel of polypeptides consisting of one or more amino acid sequences selected from SEQ ID NOs: 41-80, 251 to 271, 302 to 331 and 195 to 233.
5. The polypeptide according to any one of claims 1 to 4 comprising or consisting of the amino acid sequence of any one of SEQ ID NOs: 81 to 142, 332 to 346 and 435-449.
6. A panel of two or more polypeptides, wherein each polypeptide comprises a fragment of up to 50 consecutive amino acids of a cancer-associated antigen selected from:
- (a) TSP50, EpCAM, SPAG9, CAGE1, FBXO39, SURVIVIN, MAGE-A8, MAGE-A6, MAGE-A3 and LEMD1;
 - (b) PIWIL-4, WT1, EpCAM, BORIS, AKAP-4, OY-TES-1, SP17, PIWIL-2, PIWIL-3, SPAG9, PRAME, HIWI, SURVIVIN and AKAP-3; and/or
 - (c) SPAG9, AKAP-4, BORIS, NY-SAR-35, NY-BR-1, SURVIVIN, MAGE-A11, PRAME, MAGE-A9, HOM-TES-8, PIWIL-2, EpCAM, HIWI, PLU-1, TSGA10, ODF-4, SP17, RHOXF-2;
- and wherein each fragment comprises a different amino acid sequence selected from:
- (i) SEQ ID NOs: 21 to 38 and 234 to 244;
 - (ii) SEQ ID NOs: 272 to 301; and/or
 - (iii) SEQ ID NOs: 1 to 20, 24 and 172 to 192
- optionally wherein the fragments are flanked at the N and/or C terminus by additional amino acids that are not part of the sequence of the breast, ovarian or colorectal cancer-associated antigens.
7. The panel of polypeptides according to claim 6 comprising six peptides having the amino acid sequences of SEQ ID NOs: 130, 121, 131, 124, 134 and 126.
8. A pharmaceutical composition or kit having one or more polypeptides according to any one of claims 1 to 5, or a panel of polypeptides according to claim 6 or claim 7, or a polypeptide comprising at least two amino acid sequences selected SEQ ID NOs: 21 to 38

and 234 to 244; SEQ ID NOs: 272 to 301; and/or SEQ ID NOs: 1 to 20, 24 and 172 to 192 as an active ingredient.

9. A pharmaceutical composition according to claim 8 for use in therapy for vaccination, providing immunotherapy or inducing a cytotoxic T cell response in a subject, wherein the pharmaceutical composition is to be administered to the subject, optionally wherein the therapy is prophylactic.

10. Use of the pharmaceutical composition of claim 8 in the manufacture of a medicament for treating cancer, optionally colorectal cancer, ovarian cancer or breast cancer.

11. A method of identifying a human subject who is predicted to have a cytotoxic T cell response to administration of a pharmaceutical composition according to claim 8, the method comprising

- (i) determining that the active ingredient polypeptide(s) of the pharmaceutical composition comprise a sequence that is a T cell epitope capable of binding to at least three HLA class I molecules of the subject; and
- (ii) identifying the subject as being predicted to have a cytotoxic T cell response to administration of the pharmaceutical composition.

12. The method of claim 11 further comprising using population expression data for each antigen that

- (a) is selected from TSP50, EpCAM, SPAG9, CAGE1, FBXO39, SURVIVIN, LEMD1, MAGE-A8, MAGE-A6, MAGE-A3, PIWIL-4, WT1, BORIS, AKAP-4, OY-TES-1, SP17, PIWIL-2, PIWIL-3, PRAME, HIWI, PLU-1, TSGA10, ODF-4, RHOXF-2, NY-SAR-35, MAGE-A9, NY-BR-1, MAGE-A11, HOM-TES-85, NY-ESO-1 and AKAP-3; and
- (b) comprises an amino acid sequence that is

- i. a fragment of an active ingredient peptide of the pharmaceutical composition; and
- ii. a T cell epitope capable of binding to at least three HLA class I molecules of the subject;

to determine the probability that the subject will have a cytotoxic T cell response that targets one or more polypeptide antigens that are expressed by cancer cells of the subject.

13. A method of identifying a subject who is predicted to have a clinical response to the use according to claim 10, the method comprising

- (i) determining that the active ingredient polypeptide(s) of the pharmaceutical composition comprise two or more different amino acid sequences each of which is
 - a. a T cell epitope capable of binding to at least three HLA class I molecules of the subject; and
 - b. a fragment of a cancer-associated antigen expressed by cancer cells of the subject, optionally wherein the cancer-associated antigen is present in a sample obtained from the subject; and
- (ii) identifying the subject as predicted to have a clinical response to the use.

14. A method of determining the probability that a specific human subject will have a clinical response to the use according to claim 10, wherein one or more of the following factors corresponds to a higher probability of a clinical response:

- (a) presence in the active ingredient polypeptide(s) of a higher number of amino acid sequences and/or different amino acid sequences that are each a T cell epitope capable of binding to at least three HLA class I of the subject;
- (b) a higher number of target polypeptide antigens, comprising at least one amino acid sequence that is both
 - A. comprised in an active ingredient polypeptide; and
 - B. a T cell epitope capable of binding to at least three HLA class I of the subject;

optionally wherein the target polypeptide antigens are expressed in the subject, further optionally wherein the target polypeptides antigens are in one or more samples obtained from the subject;

(c) a higher probability that the subject expresses target polypeptide antigens, optionally a threshold number of the target polypeptide antigens and/or optionally target polypeptide antigens that have been determined to comprise at least one amino acid sequence that is both

A. comprised in in an active ingredient polypeptide; and

B. a T cell epitope capable of binding to at least three HLA class I of the subject;

and/or

(d) a higher number of target polypeptide antigens that the subject is predicted to express, optionally a higher number of target polypeptide antigens that the subject expresses with a threshold probability, and/or optionally the target polypeptide antigens that have been determined to comprise at least one amino acid sequence that is both

A. comprised in in an active ingredient polypeptide; and

B. a T cell epitope capable of binding to at least three HLA class I of the subject.

15. The method of claim 14, wherein the method comprises

(i) identifying which polypeptide antigens targeted by the active ingredient polypeptide(s) comprise an amino acid sequence that is both

A. comprised in an active ingredient polypeptide; and

B. a T cell epitope capable of binding to at least three HLA class I of the subject;

(ii) using population expression data for each antigen identified in step (i) to determine the probability that the subject expresses one or more of the antigens identified in step (i) that together comprise at least two different amino acid sequences of step (i); and

- (iii) determining the probability that the subject will have a clinical response to administration of the pharmaceutical composition, kit or panel of polypeptides, wherein a higher probability determined in step (ii) corresponds to a higher probability of a clinical response.

16. The method of claim 15, wherein the at least two different amino acid sequences are comprised in the amino acid sequence of two different polypeptide antigens targeted by the active ingredient polypeptide(s).

17. The method of any one of claims 13 to 16 further comprising selecting or recommending administration of the pharmaceutical composition as a method of treatment for the subject.

18. The use according to claim 10, wherein the subject has been identified as predicted to have a clinical response or as having above a threshold minimum likelihood of having a clinical response to the use according to any one of claims 13 to 16.

19. The pharmaceutical composition of claim 9 or the use of claim 10 or 18, wherein the pharmaceutical composition is to be administered in combination with chemotherapy, targeted therapy or a checkpoint inhibitor.

20. A method of identifying a human subject who is predicted not have a clinical response to the use according to claim 10, the method comprising

- (i) determining that the active ingredient peptide(s) of the pharmaceutical composition do not comprise two or more different amino acid sequences each of which is a T cell epitope capable of binding to at least three HLA class I molecules of the subject; and
- (ii) identifying the subject as predicted not to have a clinical response to the use.