



(86) Date de dépôt PCT/PCT Filing Date: 2013/07/25
(87) Date publication PCT/PCT Publication Date: 2014/02/20
(85) Entrée phase nationale/National Entry: 2015/01/27
(86) N° demande PCT/PCT Application No.: CA 2013/050580
(87) N° publication PCT/PCT Publication No.: 2014/026277
(30) Priorités/Priorities: 2012/08/15 (US61/683,361);
2013/05/01 (US61/818,040)

(51) Cl.Int./Int.Cl. *C07K 14/74* (2006.01),
A61K 35/14 (2015.01), *A61P 35/00* (2006.01),
C07K 7/06 (2006.01), *C12N 15/12* (2006.01),
C12Q 1/68 (2006.01), *G01N 33/48* (2006.01),
G06F 19/22 (2011.01)

(71) Demandeur/Applicant:
UNIVERSITE DE MONTREAL, CA

(72) Inventeurs/Inventors:
PERREAULT, CLAUDE, CA;
THIBAUT, PIERRE, CA;
LEMIEUX, SEBASTIEN M., CA;
GRANADOS, DIANA POALA, CA;
DEV, SRIRANGANADANE, CA; ...

(54) Titre : PROCEDE POUR L'IDENTIFICATION DE NOUVEAUX ANTIGENES D'HISTOCOMPATIBILITE MINEURE
(54) Title: METHOD FOR IDENTIFYING NOVEL MINOR HISTOCOMPATIBILITY ANTIGENS

(57) **Abrégé/Abstract:**

A novel method for human minor histocompatibility antigen (MiHA) discovery, novel MiHAs identified using this method, as well as uses of the novel MiHAs, are described. One of the features of the novel method is the inclusion of personalized translated transcriptome and/or exome in the database used for peptide identification by mass spectroscopy (MS). Candidate MiHAs are identified by comparing the personalized transcriptome and/or exome to a reference genome and/or to the transcriptome and/or exome of an HLA-matched subject.



(72) Inventeurs(suite)/Inventors(continued): DAOUDA, TARIQ, CA; CARON-LOZOTTE, OLIVIER, CA

(74) Agent: GOUDREAU GAGE DUBUC

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

(19) World Intellectual Property
Organization
International Bureau



WIPO | PCT



(10) International Publication Number

WO 2014/026277 A1

(43) International Publication Date
20 February 2014 (20.02.2014)

(51) International Patent Classification:

C07K 14/74 (2006.01) C12N 15/12 (2006.01)
A61K 35/14 (2006.01) C12Q 1/68 (2006.01)
A61P 35/00 (2006.01) G01N 33/48 (2006.01)
C07K 7/06 (2006.01) G06F 19/22 (2011.01)

(21) International Application Number:

PCT/CA2013/050580

(22) International Filing Date:

25 July 2013 (25.07.2013)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

61/683,361 15 August 2012 (15.08.2012) US
61/818,040 1 May 2013 (01.05.2013) US

(71) Applicant: UNIVERSITÉ DE MONTRÉAL [CA/CA];
2900 Édouard-Montpetit, Montreal, Québec H3T 1J4
(CA).

(72) Inventors: PERREAULT, Claude; 70 Claude-Cham-
pagne, Outremont, Québec H2V 2X1 (CA). THIBAUT,
Pierre; 288 des Érables, Ile Bizzard, Québec H9C 1N8
(CA). LEMIEUX, Sébastien; 271 rue de Fort-Rémy, Las-
alle, Québec H8R 4C6 (CA). GRANADOS, Diana Poala;
rue Philippe-Lalande, Apt. #3, Montréal, Québec H4C 0B3
(CA). DEV, Sriranganadane; 5138 5ème Avenue, Mon-
tréal, Québec H1Y 2S5 (CA). DAOUDA, Tariq; 140 rue
Sherbrooke est, Apt. # 106, Montréal, Québec H2L 1J6
(CA). CARON-LOZOTTE, Olivier; 1195 Berthier, Lon-
gueuil, Québec J4M 1G2 (CA).

(74) Agent: LECLERC, Alain, M.; Goudreau Gage Dubuc,
2000 McGill College, Suite 2200, Montréal, Québec H3A
3H3 (CA).

(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, DK, DM,
DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,
HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KN, KP, KR,
KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME,
MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ,
OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, 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, SZ, TZ,
UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ,
TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK,
EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV,
MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM,
TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW,
KM, ML, MR, NE, SN, TD, TG).

Published:

- with international search report (Art. 21(3))
- with sequence listing part of description (Rule 5.2(a))

(54) Title: METHOD FOR IDENTIFYING NOVEL MINOR HISTOCOMPATIBILITY ANTIGENS

(57) Abstract: A novel method for human minor histocompatibility antigen (MiHA) discovery, novel MiHAs identified using this method, as well as uses of the novel MiHAs, are described. One of the features of the novel method is the inclusion of personalized translated transcriptome and/or exome in the database used for peptide identification by mass spectroscopy (MS). Candidate MiHAs are identified by comparing the personalized transcriptome and/or exome to a reference genome and/or to the transcriptome and/or exome of an HLA-matched subject.



WO 2014/026277 A1

METHOD FOR IDENTIFYING NOVEL MINOR HISTOCOMPATIBILITY ANTIGENS

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of U.S. provisional application serial No. 61/683,361, filed on August 15, 2012, and of U.S. provisional application serial No. 61/818,040, filed on May 1, 2013, which are incorporated herein by reference in their entirety.

TECHNICAL FIELD

The present invention generally relates to histocompatibility antigens, and more specifically to minor histocompatibility antigens (MiHAs), identification and use thereof.

BACKGROUND ART

Histocompatibility antigens are a group of cell membrane alloantigens that are recognized by T lymphocytes and thereby initiate graft rejection or graft-versus-host disease (GVHD) following transplantation (1). In the early days of immunogenetics, the identification of major histocompatibility complex (MHC) antigens was based on their strong immunogenicity in skin transplant experiments between congenic-resistant strains of mice. Other less potent antigens were called minor histocompatibility antigens (MiHA). It soon became obvious that a distinction between major and minor antigens based solely on their immunogenicity was imprecise, as some MHC antigens are weak immunogens while some MiHA appear “neither weak nor minor” (2,3). It is now known that MHC antigens (also referred to as HLA antigens) are transmembrane glycoproteins encoded by closely linked polymorphic loci located on chromosome 6 in humans. Their primary role is to bind endogenous and exogenous peptides that are scrutinized by T cells. MHC (or HLA) molecules present thousands of peptides at the surface of human cells (4,5). These MHC-associated peptides (MAPs) are referred to as the immunopeptidome and they originate from proteasomal processing and further processing of endogenous proteins (6-8). The immunopeptidome of identical twins (also referred to as syngeneic subjects) is identical. By contrast, MAPs present on cells from HLA-identical non-syngeneic subjects are classified into two categories: i) invariant MAPs which are present in all subjects with a given HLA type, and ii) MiHAs which are MAPs present in some but absent in other subjects (9). When T cells are transplanted into an MHC-identical host, they react promptly and specifically to what they see as non-self: host-specific MiHAs. MiHAs are essentially genetic polymorphisms that are immunogenic for T cells. MiHAs are a consequence of any form of accumulated genetic variation that translates to differential MAP display (3,9-13).

Two main strategies can be used for cancer immunotherapy: vaccination and adoptive T-cell immunotherapy (ATCI). The term “ATCI” refers to transfusion of T lymphocytes that may come from different types of donors: the patient (autologous), a genetically-identical twin

(syngeneic), or a non-identical donor (allogeneic). To date, ATCI has yielded much higher cancer remission and cure rates than vaccines, and the most widely used form of cancer ATCI is allogeneic hematopoietic cell transplantation (AHCT) (17-22). The graft-versus-tumor (GVT) effect induced by allogeneic AHCT is due mainly to T-cell responses against host MiHAs: GVT is abrogated if the donor is an identical twin (no MiHA differences with the recipient) or if the graft is T-cell depleted (20-23). More than 200,000 individuals treated for hematological malignancies owe their life to the MiHA-dependent GVT effect which represents the most striking evidence of the ability of the human immune system to eradicate neoplasias (18-24-28). Though the allogeneic GVT effect is being used essentially to treat patients with hematologic malignancies, preliminary evidence suggests that it may be also effective for the treatment of solid tumors (29-33). Nonetheless, the considerable potential of MiHA-targeted cancer immunotherapy has not been properly exploited in medicine. In current medical practice, MiHA-based ATCI is limited to "conventional" AHCT, that is, injection of hematopoietic cells from an allogeneic HLA-matched donor. Such unselective injection of allogeneic lymphocytes is a very rudimentary form of MiHA-targeted therapy. First, it lacks specificity and is therefore highly toxic: unselected allogeneic T cells react against a multitude of host MiHAs and thereby induce GVHD in 60% of recipients. GVHD is always incapacitating and frequently lethal (34-38). Second, conventional AHCT induces only an attenuated form of GVT reaction because donor T cells are not being primed (pre-activated) against specific MiHAs expressed on cancer cells prior to injection into the patient. While primed T cells are resistant to tolerance induction, naïve T cells can be tolerized by tumor cells (39-42).

It has been demonstrated in mice models of AHCT that, by replacing unselected donor lymphocytes with CD8 T cells primed against a single MiHA, it was possible to cure leukemia and solid tumors without causing GVHD or any untoward effect (33-43-44). Success depends on two main points: selection of an immunodominant (highly immunogenic) MiHA expressed on neoplastic cells, and priming of donor CD8 T cells against the target MiHA prior to AHCT. A recent article (20) describes why MiHA-targeted ATCI is so effective and how translation of this approach in the clinic could have a significant impact on cancer immunotherapy. Implementation of MiHA-targeted ATCI in humans has been limited mainly by the paucity of molecularly defined human MiHAs. Thus, only 33% of patients with leukemia would be eligible for MiHA-based ATCI (15).

Human MiHAs have been discovered using reductionist T-cell based methods. Starting with cytotoxic T lymphocytes (CTLs) from an individual reactive against cells of another HLA-identical subject, investigators have tried and identified MiHAs recognized by these T cells. Different methods have used to do so. First, CTLs were tested on MiHA-negative cells coated with MAPs eluted from MiHA-positive cells. The MAP eluates were fractionated and ultimately the MiHA recognized by CTLs was sequenced by mass spectrometry (MS) (48-53). Second,

CTLs were used to screen MiHA-negative cells transfected with cDNA libraries to identify MiHA-coding transcripts (16:54-59). Finally, CTLs have been tested on lymphoblastoid cell lines from many subjects and linkage analyses were performed (based for instance on whole genome association scans or HapMap resources) on lines recognized or not by CTLs (60-67).

5 The various methods used to discover MiHAs present significant caveats. Firstly, they are not really suitable for high-throughput MiHA discovery: MiHA discovery is made one by one and depends on the availability of a CTL line. Secondly, only MiHAs that have been eluted from living cells and identified by MS can be considered to be validated (direct identification). In the other cases (indirect identification), uncertainty remains as to the exact structure of MiHAs
10 naturally presented at the cell surface (an important criterion for MiHA-targeted immunotherapy). The ambiguity stems mainly from two factors: i) T cells are eminently cross-reactive and can recognize more than one peptide (68); ii) bioinformatic tools used for identification of MAPs in general and MiHAs in particular do not have sufficient reliability to replace direct proteomic identification (69-71).

15 Thus, there is a need for novel approaches for the identification of MiHAs.

 The present description refers to a number of documents, the content of which is herein incorporated by reference in their entirety.

SUMMARY OF THE INVENTION

 In a first aspect, the present invention provides a method of identifying a minor
20 histocompatibility antigen (MiHA) candidate, the method comprising: (a) isolating and sequencing MHC-associated peptides (MAPs) in a first cell sample from a first subject; (b) performing a whole transcriptome and/or exome sequencing on a second cell sample obtained from said first subject; (c) comparing the sequenced whole transcriptome and/or exome to a reference genome to identify single nucleotide variations (SNVs) between the transcriptome
25 and/or exome of said first subject and the reference genome; (d) *in silico* translating the sequences containing the identified SNVs to identify peptide sequences comprising at least one non-synonymous mutation caused by said SNVs; (e) comparing the sequences of the MAPs isolated in (a) with the peptide sequences identified in (d); and (f) identifying a MiHA candidate based on said comparison.

30 In another aspect, the present invention provides a method of identifying a minor histocompatibility antigen (MiHA) candidate, the method comprising: (a) isolating and sequencing MHC-associated peptides (MAPs) in a first cell sample from a first and second subjects, wherein said first and second subjects are human leukocyte antigen (HLA)-matched; (b) performing a whole transcriptome and/or exome sequencing on a second cell sample
35 obtained from said first and second subjects; (c) comparing the sequenced whole transcriptomes and/or exomes to identify single nucleotide variations (SNVs) between the

transcriptomes and/or exomes of said first and second subjects; (d) *in silico* translating the sequences containing the identified SNVs to identify peptide sequences comprising at least one non-synonymous mutation caused by said SNVs; (e) comparing the sequences of the MAPs isolated in (a) with the peptide sequences identified in (d); and (f) identifying a MiHA candidate
 5 based on said comparison.

In an embodiment, the above-mentioned MiHA candidate is a MAP whose sequence comprises at least one mutation relative to the corresponding sequence translated from the reference genome.

In an embodiment, the above-mentioned MiHA candidate is a MAP present in the first
 10 cell sample from said first subject but absent from the first cell sample from said second subject.

In an embodiment, the above-mentioned reference genome is the Genome Reference Consortium Human Build 37 (GRCh37).

In an embodiment, the above-mentioned first and/or second cell sample is a peripheral blood cell sample. In a further embodiment, the above-mentioned peripheral blood cell sample
 15 an immortalized peripheral blood cell sample. In a further embodiment, the above-mentioned immortalized peripheral blood cell sample is an Epstein-Barr virus (EBV)-transformed B lymphoblastoid cell line.

In an embodiment, the above-mentioned isolating MAPs comprises (i) releasing said MAPs from said cell sample by mild acid treatment; and (ii) subjecting the released MAPs to
 20 chromatography.

In an embodiment, the above-mentioned method further comprises filtering the released peptides with a size exclusion column prior to said chromatography. In a further embodiment, the above-mentioned size exclusion column has a cut-off of about 3000 Da. In an embodiment, the above-mentioned chromatography is cation exchange chromatography.

25 In an embodiment, the above-mentioned peptide sequences of (d) have a length of 12 amino acids or less. In a further embodiment, the above-mentioned peptide sequences of (d) have a length of 8 to 11 amino acids.

In an embodiment, the above-mentioned comparing comprises subjecting the MAPs isolated in (a) to mass spectrometry and comparing the MS spectra obtained with the peptide
 30 sequences identified in (d).

In an embodiment, the above-mentioned method further comprises determining the binding of the MiHA candidate identified in (f) to a major histocompatibility complex (MHC) class I molecule.

In another aspect, the present invention provides a peptide of 50 amino acids or less
 35 comprising the sequence (I)



wherein Z^1 is an amino terminal modifying group or is absent; X^1 is a sequence of 1 to 43 amino acids or is absent; X^2 is L or S; X^3 is a sequence of 1 to 43 amino acids or is absent; and Z^2 is a carboxy terminal modifying group or is absent. In an embodiment, X^1 is an acidic amino acid, in a further embodiment glutamic acid (E). In an embodiment, X^3 is an amino acid, in a further embodiment a hydrophobic amino acid, more particularly leucine (L). In an embodiment, X^2 is L. In another embodiment, X^2 is S. In an embodiment, the peptide comprises the sequence ELQEKFLSL (SEQ ID NO: 15), in a further embodiment the peptide is ELQEKFLSL (SEQ ID NO: 15). In another embodiment, the peptide comprises the sequence ELQEKFSSL (SEQ ID NO: 16), in a further embodiment the peptide is ELQEKFSSL (SEQ ID NO: 16).

In another aspect, the present invention provides a method of treating cancer, said method comprising administering to a subject in need thereof an effective amount of CD8 T lymphocytes recognizing a MHC class I molecule of the HLA-B*0801 allele loaded with the above-mentioned peptide (I). In an embodiment, the above-mentioned method further comprises determining whether said subject expresses a CENPF nucleic acid comprising a T or a C at a position corresponding to nucleotide 4409 in the nucleic acid sequence of human CENPF (FIGs. 1A to 1D, NCBI Reference Sequence: NM_016343.3), and/or a CENPF polypeptide comprising a leucine or serine residue at a position corresponding to residue 1412 in the protein sequence of human CENPF (FIG. 1E, NCBI Reference Sequence: NP_057427.3), wherein (a) if said subject expresses a CENPF nucleic acid comprising a T at a position corresponding to nucleotide 4409 in the nucleic acid sequence of human CENPF and/or a CENPF polypeptide comprising a leucine residue at a position corresponding to residue 1412 in the protein sequence of human CENPF, X^2 is L in said peptide; (b) if said subject expresses a CENPF nucleic acid comprising a C at a position corresponding to nucleotide 4409 in the nucleic acid sequence of human CENPF and/or a CENPF polypeptide comprising a serine residue at a position corresponding to residue 1412 in the protein sequence of human CENPF, X^2 is S in said peptide.

In an embodiment, the above-mentioned determining comprises sequencing a CENPF nucleic acid. In an embodiment, the above-mentioned CD8 T lymphocytes are *in vitro* expanded CD8 T lymphocytes.

In an embodiment, the above-mentioned method further comprises: (i) if said subject is the subject of (a), culturing CD8 T lymphocytes from a second subject comprising a C at a position corresponding to nucleotide 4409 in the nucleic acid sequence of human CENPF and/or a CENPF polypeptide comprising a serine residue at a position corresponding to residue 1412 in the protein sequence of human CENPF in the presence of cells expressing a MHC class I molecule of the HLA-B*0801 allele loaded with the above-mentioned peptide (I), wherein X^2 is L in said peptide, under conditions suitable for CD8 T lymphocyte expansion; or (ii) if said subject is the subject of (b), culturing CD8 T lymphocytes from a second subject comprising a T

at a position corresponding to nucleotide 4409 in the nucleic acid sequence of human CENPF and/or a CENPF polypeptide comprising a leucine residue at a position corresponding to residue 1412 in the protein sequence of human CENPF in the presence of cells expressing a MHC class I molecule of the HLA-B*0801 allele loaded with the above-mentioned peptide (I),
 5 wherein X² is S in said peptide, under conditions suitable for CD8 T lymphocyte expansion.

In an embodiment, the above-mentioned subject is an allogeneic stem cell transplantation (ASCT) recipient.

In another aspect, the present invention provides a method of expanding CD8 T lymphocytes for adoptive T-cell immunotherapy, said method comprising: (a) determining
 10 whether a candidate donor expresses a CENPF nucleic acid comprising a T or a C at a position corresponding to nucleotide 4409 in the nucleic acid sequence of human CENPF (FIGs. 1A to 1D, NCBI Reference Sequence: NM_016343.3), and/or a CENPF polypeptide comprising a leucine or serine residue at a position corresponding to residue 1412 in the protein sequence of human CENPF (FIG. 1E, NCBI Reference Sequence: NP_057427.3) and (b)(i) if said candidate
 15 donor expresses a CENPF nucleic acid comprising a T at a position corresponding to nucleotide 4409 in the nucleic acid sequence of human CENPF and/or a CENPF polypeptide comprising a leucine residue at a position corresponding to residue 1412 in the protein sequence of human CENPF, culturing CD8 T lymphocytes from said candidate donor in the presence of cells expressing a MHC class I molecule of the HLA-B*0801 allele loaded with the above-mentioned
 20 peptide (I), wherein X² is S in said peptide, under conditions suitable for CD8 T lymphocyte expansion; or (b)(ii) if said candidate donor expresses a CENPF nucleic acid comprising a C at a position corresponding to nucleotide 4409 in the nucleic acid sequence of human CENPF and/or a CENPF polypeptide comprising a serine residue at a position corresponding to residue 1412 in the protein sequence of human CENPF, culturing CD8 T lymphocytes from said
 25 candidate donor in the presence of cells expressing a MHC class I molecule of the HLA-B*0801 allele loaded with the above-mentioned peptide (I), wherein X² is L in said peptide, under conditions suitable for CD8 T lymphocyte expansion.

In another aspect, the present invention provides a peptide of 50 amino acids or less comprising the sequence (II)



wherein Z¹ is an amino terminal modifying group or is absent; X⁴ is a sequence of 1 to 43 amino acids or is absent; X⁵ is G or R; X⁶ is an amino acid or is absent; X⁷ is a sequence of 1 to 43 amino acids or is absent; and Z² is a carboxy terminal modifying group or is absent.

In an embodiment, X⁴ is glutamine (Q). In an embodiment, X⁶ is an amino acid, in a
 35 further embodiment a basic amino acid, more particularly lysine (K). In an embodiment, X⁷ is an amino acid, in a further embodiment leucine (L). In an embodiment, X⁵ is G. In another embodiment, X⁵ is R. In an embodiment, the peptide comprises the sequence QELDGVFQKL

(SEQ ID NO:17). In a further embodiment, the peptide is QELDGVFQKL (SEQ ID NO:17). In another embodiment, the peptide comprises the sequence QELDRVVFQKL (SEQ ID NO:18). In a further embodiment, the peptide is QELDRVVFQKL (SEQ ID NO:18).

In another aspect, the present invention provides a method of treating cancer, said method comprising administering to a subject in need thereof an effective amount of CD8 T lymphocytes recognizing a MHC class I molecule of the HLA-B*4403 allele loaded with the above-mentioned peptide (II).

In an embodiment, the above-mentioned method further comprises determining whether said subject expresses a ZWINT nucleic acid comprising an A or a G at a position corresponding to nucleotide 596 in the nucleic acid sequence of human ZWINT (FIG. 2A, NCBI Reference Sequence: NM_007057.3), and/or a ZWINT polypeptide comprising a arginine or glycine residue at a position corresponding to residue 187 in the protein sequence of human ZWINT (FIG. 2B, NCBI Reference Sequence: NP_008988.2), wherein (a) if said subject expresses a ZWINT nucleic acid comprising an A at a position corresponding to nucleotide 596 in the nucleic acid sequence of human ZWINT and/or a ZWINT polypeptide comprising an arginine residue at a position corresponding to residue 187 in the protein sequence of human ZWINT, X⁵ is R in said peptide; (b) if said subject expresses a ZWINT nucleic acid comprising a G at a position corresponding to nucleotide 596 in the nucleic acid sequence of human ZWINT and/or a ZWINT polypeptide comprising a glycine residue at a position corresponding to residue 187 in the protein sequence of human ZWINT, X⁵ is G in said peptide.

In an embodiment, the above-mentioned determining comprises sequencing a human ZWINT nucleic acid. In an embodiment, the above-mentioned CD8 T lymphocytes are *in vitro* expanded CD8 T lymphocytes.

In an embodiment, the above-mentioned method further comprises: (i) if said subject is the subject of (a), culturing CD8 T lymphocytes from a second subject comprising a G at a position corresponding to nucleotide 596 in the nucleic acid sequence of human ZWINT and/or a ZWINT polypeptide comprising a glycine residue at a position corresponding to residue 187 in the protein sequence of human ZWINT in the presence of cells expressing a MHC class I molecule of the HLA-B*4403 allele loaded with the above-mentioned peptide (II), wherein X⁵ is R in said peptide, under conditions suitable for CD8 T lymphocyte expansion; or (ii) if said subject is the subject of (b), culturing CD8 T lymphocytes from a second subject comprising an A at a position corresponding to nucleotide 596 in the nucleic acid sequence of human ZWINT and/or a ZWINT polypeptide comprising an arginine residue at a position corresponding to residue 187 in the protein sequence of human ZWINT in the presence of cells expressing a MHC class I molecule of the HLA-B*4403 allele loaded with the above-mentioned peptide (II), wherein X⁵ is G in said peptide, under conditions suitable for CD8 T lymphocyte expansion. In

an embodiment, the above-mentioned subject is an allogeneic stem cell transplantation (ASCT) recipient.

In another aspect, the present invention provides a method of expanding CD8 T lymphocytes for adoptive T-cell immunotherapy, said method comprising: (a) determining
 5 whether a candidate donor expresses a ZWINT nucleic acid comprising an A or a G at a position corresponding to nucleotide 596 in the nucleic acid sequence of human ZWINT (FIG. 2A, NCBI Reference Sequence: NM_007057.3), and/or a ZWINT polypeptide comprising a arginine or glycine residue at a position corresponding to residue 187 in the protein sequence of human ZWINT (FIG. 2B, NCBI Reference Sequence: NP_008988.2) and (b)(i) if said candidate
 10 donor expresses a ZWINT nucleic acid comprising an A at a position corresponding to nucleotide 596 in the nucleic acid sequence of human ZWINT and/or a ZWINT polypeptide comprising an arginine residue at a position corresponding to residue 187 in the protein sequence of human ZWINT, culturing CD8 T lymphocytes from said candidate donor in the presence of cells expressing a MHC class I molecule of the HLA-B*4403 allele loaded with the
 15 above-mentioned peptide (II), wherein X⁵ is G in said peptide, under conditions suitable for CD8 T lymphocyte expansion; or (b)(ii) if said candidate donor expresses a ZWINT nucleic acid comprising a G at a position corresponding to nucleotide 596 in the nucleic acid sequence of human ZWINT and/or a ZWINT polypeptide comprising a glycine residue at a position corresponding to residue 187 in the protein sequence of human ZWINT, culturing CD8 T
 20 lymphocytes from said candidate donor in the presence of cells expressing a MHC class I molecule of the HLA-B*4403 allele loaded with the above-mentioned peptide (II), wherein X⁵ is R in said peptide, under conditions suitable for CD8 T lymphocyte expansion.

In another aspect, the present invention provides a peptide of 50 amino acids or less comprising the sequence (III)

25
$$Z^1-X^8\text{LFFRKVX}^9\text{X}^{10}-Z^2 \quad (\text{III})$$

wherein Z¹ is an amino terminal modifying group or is absent; X⁸ is a sequence of 1 to 43 amino acids or is absent; X⁹ is P or A; X¹⁰ is a sequence of 1 to 43 amino acids or is absent; and Z² is a carboxy terminal modifying group or is absent. In an embodiment, X⁸ is serine (S). In an embodiment, X¹⁰ is an amino acid, in a further embodiment an aromatic amino acid, more
 30 particularly phenylalanine (F). In an embodiment, X⁹ is P. In another embodiment, X⁹ is A. In an embodiment, the peptide comprises the sequence SLFFRKVPF (SEQ ID NO:19). In a further embodiment, the peptide is SLFFRKVPF (SEQ ID NO:19). In another embodiment, the peptide comprises the sequence SLFFRKVAF (SEQ ID NO:20). In a further embodiment, the peptide is SLFFRKVAF (SEQ ID NO:20).

35 In another aspect, the present invention provides a method of treating cancer, said method comprising administering to a subject in need thereof an effective amount of CD8 T

lymphocytes recognizing a MHC class I molecule of the HLA-B*0801 allele loaded with the above-mentioned peptide.

In an embodiment, the above-mentioned method further comprises determining whether said subject expresses a MTCH2 nucleic acid comprising a C or a G at a position
5 corresponding to nucleotide 1057 in the nucleic acid sequence of human MTCH2 (FIG. 3A, NCBI Reference Sequence: NM_014342.3), and/or a MTCH2 polypeptide comprising a proline or alanine residue at a position corresponding to residue 290 in the protein sequence of human MTCH2 (FIG. 3B, NCBI Reference Sequence: NP_055157.1), wherein (a) if said subject expresses a MTCH2 nucleic acid comprising a C at a position corresponding to nucleotide 1057
10 in the nucleic acid sequence of human MTCH2 and/or a MTCH2 polypeptide comprising a proline residue at a position corresponding to residue 290 in the protein sequence of human MTCH2, X⁹ is P in said peptide; (b) if said subject expresses a MTCH2 nucleic acid comprising a G at a position corresponding to nucleotide 1057 in the nucleic acid sequence of human MTCH2 and/or a MTCH2 polypeptide comprising an alanine residue at a position corresponding
15 to residue 290 in the protein sequence of human MTCH2, X⁵ is A in said peptide.

In an embodiment, the above-mentioned determining comprises sequencing a human MTCH2 nucleic acid. In an embodiment, the above-mentioned CD8 T lymphocytes are *in vitro* expanded CD8 T lymphocytes.

In an embodiment, the above-mentioned method further comprises: (i) if said subject is
20 the subject of (a), culturing CD8 T lymphocytes from a second subject comprising a G at a position corresponding to nucleotide 1057 in the nucleic acid sequence of human MTCH2 and/or a MTCH2 polypeptide comprising an alanine residue at a position corresponding to residue 290 in the protein sequence of human MTCH2 in the presence of cells expressing a MHC class I molecule of the HLA-B*0801 allele loaded with the above-mentioned peptide (III),
25 wherein X⁹ is P in said peptide, under conditions suitable for CD8 T lymphocyte expansion; or (ii) if said subject is the subject of (b), culturing CD8 T lymphocytes from a second subject comprising a C at a position corresponding to nucleotide 1057 in the nucleic acid sequence of human MTCH2 and/or a MTCH2 polypeptide comprising a proline residue at a position corresponding to residue 290 in the protein sequence of human MTCH2 in the presence of cells
30 expressing a MHC class I molecule of the HLA-B*0801 allele loaded with the above-mentioned peptide (III), wherein X⁹ is A in said peptide, under conditions suitable for CD8 T lymphocyte expansion.

In an embodiment, the above-mentioned subject is an allogeneic stem cell transplantation (ASCT) recipient.

35 In another aspect, the present invention provides a method of expanding CD8 T lymphocytes for adoptive T-cell immunotherapy, said method comprising: (a) determining whether a candidate donor expresses a MTCH2 nucleic acid comprising a C or a G at a position

corresponding to nucleotide 1057 in the nucleic acid sequence of human MTCH2 (FIG. 3A, NCBI Reference Sequence: NM_014342.3), and/or a MTCH2 polypeptide comprising a proline or alanine residue at a position corresponding to residue 290 in the protein sequence of human MTCH2 (FIG. 3B, NCBI Reference Sequence: NP_055157.1) and (b)(i) if said candidate donor
 5 expresses a MTCH2 nucleic acid comprising a C at a position corresponding to nucleotide 1057 in the nucleic acid sequence of human MTCH2 and/or a MTCH2 polypeptide comprising a proline residue at a position corresponding to residue 290 in the protein sequence of human MTCH2, culturing CD8 T lymphocytes from said candidate donor in the presence of cells expressing a MHC class I molecule of the HLA-B*0801 allele loaded with the above-mentioned
 10 peptide (III), wherein X^9 is A in said peptide, under conditions suitable for CD8 T lymphocyte expansion; or (b)(ii) if said candidate donor expresses a MTCH2 nucleic acid comprising a G at a position corresponding to nucleotide 1057 in the nucleic acid sequence of human MTCH2 and/or a MTCH2 polypeptide comprising an alanine residue at a position corresponding to residue 290 in the protein sequence of human MTCH2, culturing CD8 T lymphocytes from said
 15 candidate donor in the presence of cells expressing a MHC class I molecule of the HLA-B*0801 allele loaded with the above-mentioned peptide (III), wherein X^9 is P in said peptide, under conditions suitable for CD8 T lymphocyte expansion.

In another aspect, the present invention provides a peptide of 50 amino acids or less comprising the sequence (IV)

20 $Z^1-X^{11}X^{12}VLKPGNX^{13}X^{14}-Z^2$ (IV)

wherein Z^1 is an amino terminal modifying group or is absent; X^{11} is a sequence of 1 to 43 amino acids or is absent; X^{12} is S or T; X^{13} is an amino acid or is absent; X^{14} is a sequence of 1 to 43 amino acids or is absent; and Z^2 is a carboxy terminal modifying group or is absent. In an embodiment, X^{11} is absent. In an embodiment, the above-mentioned X^{13} is an amino acid, in a
 25 further embodiment serine (S). In an embodiment, X^{14} is an amino acid, in a further embodiment a basic amino acid, more particularly lysine (K). In an embodiment, X^{12} is S. In another embodiment, X^{12} is T. In an embodiment, the above-mentioned peptide comprises the sequence SVLKPGNSK (SEQ ID NO:21). In a further embodiment, the peptide is SVLKPGNSK (SEQ ID NO:21). In another embodiment, the above-mentioned peptide comprises the sequence
 30 TVLKPGNSK (SEQ ID NO:22). In a further embodiment, the peptide is TVLKPGNSK (SEQ ID NO:22).

In another aspect, the present invention provides a method of treating cancer, said method comprising administering to a subject in need thereof an effective amount of CD8 T lymphocytes recognizing a MHC class I molecule of the HLA-A*0301 allele loaded with the
 35 above-mentioned peptide (IV).

In an embodiment, the above-mentioned method further comprises determining whether said subject expresses an ELF1 nucleic acid comprising an A or T at a position

corresponding to nucleotide 1400 in the nucleic acid sequence of human ELF1 (FIGs. 4A and 4B, NCBI Reference Sequence: NM_172373.3) and/or an ELF1 polypeptide having a threonine or a serine at a position corresponding to residue 343 in the ELF1 protein sequence (FIG. 4C, NCBI Reference Sequence: NP_758961.1), wherein (a) if said subject expresses an ELF1

5 nucleic acid comprising an A at a position corresponding to nucleotide 1400 in the nucleic acid sequence of human ELF1 and/or an ELF1 polypeptide comprising a threonine residue at a position corresponding to residue 343 in the protein sequence of human ELF1, X¹² is T in said peptide; (b) if said subject expresses an ELF1 nucleic acid comprising a T at a position corresponding to nucleotide 1400 in the nucleic acid sequence of human ELF1 and/or an ELF1

10 polypeptide comprising a serine residue at a position corresponding to residue 343 in the protein sequence of human ELF1, X¹² is S in said peptide.

In an embodiment, the above-mentioned determining comprises sequencing a human ELF1 nucleic acid. In an embodiment, the above-mentioned CD8 T lymphocytes are *in vitro* expanded CD8 T lymphocytes.

15 In an embodiment, the above-mentioned method further comprises: (i) if said subject is the subject of (a), culturing CD8 T lymphocytes from a second subject comprising a T at a position corresponding to nucleotide 1400 in the nucleic acid sequence of human ELF1 and/or an ELF1 polypeptide comprising a serine residue at a position corresponding to residue 343 in the protein sequence of human ELF1 in the presence of cells expressing a MHC class I

20 molecule of the HLA-A*0301 allele loaded with the peptide of any one of claims 86 to 99, wherein X¹² is T in said peptide, under conditions suitable for CD8 T lymphocyte expansion; or (ii) if said subject is the subject of (b), culturing CD8 T lymphocytes from a second subject comprising an A at a position corresponding to nucleotide 1400 in the nucleic acid sequence of human ELF1 and/or an ELF1 polypeptide comprising a threonine residue at a position

25 corresponding to residue 343 in the protein sequence of human ELF1 in the presence of cells expressing a MHC class I molecule of the HLA-A*0301 allele loaded with the peptide of any one of claims 86 to 99, wherein X¹² is S in said peptide, under conditions suitable for CD8 T lymphocyte expansion.

In an embodiment, the above-mentioned subject is an allogeneic stem cell

30 transplantation (ASCT) recipient.

In another aspect, the present invention provides a method of expanding CD8 T lymphocytes for adoptive T-cell immunotherapy, said method comprising: (a) determining whether a candidate donor expresses an ELF1 nucleic acid comprising an A or T at a position corresponding to nucleotide 1400 in the nucleic acid sequence of human ELF1 (FIGs. 4A and

35 4B, NCBI Reference Sequence: NM_172373.3) and/or an ELF1 polypeptide having a threonine or a serine at a position corresponding to residue 343 in the ELF1 protein sequence (FIG. 4C, NCBI Reference Sequence: NP_758961.1) and (b)(i) if said candidate donor expresses an

ELF1 nucleic acid comprising an A at a position corresponding to nucleotide 1400 in the nucleic acid sequence of human ELF1 and/or an ELF1 polypeptide having a threonine at a position corresponding to residue 343 in the ELF1 protein sequence, culturing CD8 T lymphocytes from said candidate donor in the presence of cells expressing a MHC class I molecule of the HLA-A*0301 allele loaded with the above-mentioned peptide (IV), wherein X¹² is S in said peptide, under conditions suitable for CD8 T lymphocyte expansion; or (b)(ii) if said candidate donor expresses an ELF1 nucleic acid comprising a T at a position corresponding to nucleotide 1400 in the nucleic acid sequence of human ELF1 and/or an ELF1 polypeptide having a serine at a position corresponding to residue 343 in the ELF1 protein sequence, culturing CD8 T lymphocytes from said candidate donor in the presence of cells expressing a MHC class I molecule of the HLA-A*0301 allele loaded with the above-mentioned peptide (IV), wherein X¹² is T in said peptide, under conditions suitable for CD8 T lymphocyte expansion.

In another aspect, the present invention provides a peptide of 50 amino acids or less comprising the sequence V

15 $Z^1-X^{15}X^{16}YDKGPFX^{17}X^{18}X^{19}-Z^2$ (V)

wherein Z¹ is an amino terminal modifying group or is absent; X¹⁵ is a sequence of 1 to 43 amino acids or is absent; X¹⁶ is an amino acid or is absent; X¹⁷ is R or W; X¹⁸ is an amino acid or is absent; X¹⁹ is a sequence of 1 to 43 amino acids or is absent; Z² is a carboxy terminal modifying group or is absent. In an embodiment, the above-mentioned X¹⁶ is an amino acid, in a further embodiment a methionine (M). In an embodiment, X¹⁵ is an amino acid, in a further embodiment an alanine (A). In an embodiment, X¹⁸ is an amino acid, in a further embodiment a serine (S). In an embodiment, X¹⁹ is an amino acid, in a further embodiment a basic amino acid, more particularly lysine (K). In an embodiment, X¹⁷ is R. In another embodiment, X¹⁷ is W. In an embodiment, the above-mentioned peptide comprises the sequence AMYDKGPFRSK (SEQ ID NO:23). In an embodiment, the above-mentioned peptide is AMYDKGPFRSK (SEQ ID NO:23). In another embodiment, the above-mentioned peptide comprises the sequence AMYDKGPFWSK (SEQ ID NO:24). In an embodiment, the above-mentioned peptide is AMYDKGPFWSK (SEQ ID NO:24).

In another aspect, the present invention provides a method of treating cancer, said method comprising administering to a subject in need thereof an effective amount of CD8 T lymphocytes recognizing a MHC class I molecule of the HLA-A*0301 allele loaded with the above-mentioned peptide (V).

In an embodiment, the above-mentioned method further comprises determining whether said subject expresses an NQO1 nucleic acid comprising an C or T at a position corresponding to nucleotide 615 in the nucleic acid sequence of human NQO1 (FIG. 5A, NCBI Reference Sequence: NM_000903.2) and/or an NQO1 polypeptide having an arginine or a tryptophan at a position corresponding to residue 139 in the NQO1 protein sequence (FIG. 5B,

NCBI Reference Sequence: NP_000894.1), wherein (a) if said subject expresses a NQO1 nucleic acid comprising an C at a position corresponding to nucleotide 615 in the nucleic acid sequence of human NQO1 and/or a NQO1 polypeptide comprising an arginine residue at a position corresponding to residue 139 in the protein sequence of human NQO1, X¹⁷ is R in said peptide; (b) if said subject expresses an NQO1 nucleic acid comprising a T at a position corresponding to nucleotide 615 in the nucleic acid sequence of human NQO1 and/or a NQO1 polypeptide comprising a tryptophan residue at a position corresponding to residue 139 in the protein sequence of human NQO1, X¹⁷ is W in said peptide.

In an embodiment, the above-mentioned determining comprises sequencing a human NQO1 nucleic acid. In an embodiment, the above-mentioned CD8 T lymphocytes are *in vitro* expanded CD8 T lymphocytes.

In an embodiment, the above-mentioned method further comprises: (i) if said subject is the subject of (a), culturing CD8 T lymphocytes from a second subject comprising a T at a position corresponding to nucleotide 615 in the nucleic acid sequence of human NQO1 and/or a NQO1 polypeptide comprising a tryptophan residue at a position corresponding to residue 139 in the protein sequence of human NQO1 in the presence of cells expressing a MHC class I molecule of the HLA-A*0301 allele loaded with the above-mentioned peptide (V), wherein X¹⁷ is R in said peptide, under conditions suitable for CD8 T lymphocyte expansion; or (ii) if said subject is the subject of (b), culturing CD8 T lymphocytes from a second subject comprising a C at a position corresponding to nucleotide 615 in the nucleic acid sequence of human NQO1 and/or a NQO1 polypeptide comprising an arginine residue at a position corresponding to residue 139 in the protein sequence of human NQO1 in the presence of cells expressing a MHC class I molecule of the HLA-A*0301 allele loaded with the above-mentioned peptide (V), wherein X¹⁷ is W in said peptide, under conditions suitable for CD8 T lymphocyte expansion.

In an embodiment, the above-mentioned subject is an allogeneic stem cell transplantation (ASCT) recipient.

In another aspect, the present invention provides a method of expanding CD8 T lymphocytes for adoptive T-cell immunotherapy, said method comprising: (a) determining whether said subject expresses an NQO1 nucleic acid comprising an C or T at a position corresponding to nucleotide 615 in the nucleic acid sequence of human NQO1 (FIG. 5A, NCBI Reference Sequence: NM_000903.2) and/or an NQO1 polypeptide having an arginine or a tryptophan at a position corresponding to residue 139 in the NQO1 protein sequence (FIG. 5B, NCBI Reference Sequence: NP_000894.1) and (b)(i) if said candidate donor expresses a NQO1 nucleic acid comprising an C at a position corresponding to nucleotide 615 in the nucleic acid sequence of human NQO1 and/or a NQO1 polypeptide comprising an arginine residue at a position corresponding to residue 139 in the protein sequence of human NQO1, culturing CD8 T lymphocytes from said candidate donor in the presence of cells expressing a MHC class I

molecule of the HLA-A*0301 allele loaded with the above-mentioned peptide (V), wherein X^{17} is W in said peptide, under conditions suitable for CD8 T lymphocyte expansion; or (b)(ii) if said candidate donor expresses a NQO1 nucleic acid comprising a T at a position corresponding to nucleotide 615 in the nucleic acid sequence of human NQO1 and/or a NQO1 polypeptide comprising a tryptophan residue at a position corresponding to residue 139 in the protein sequence of human NQO1, culturing CD8 T lymphocytes from said candidate donor in the presence of cells expressing a MHC class I molecule of the HLA-A*0301 allele loaded with the above-mentioned peptide (V), wherein X^{17} is R in said peptide, under conditions suitable for CD8 T lymphocyte expansion.

In another aspect, the present invention provides a peptide of 50 amino acids or less comprising the sequence VI



wherein Z^1 is an amino terminal modifying group or is absent; X^{20} is a sequence of 1 to 43 amino acids or is absent; X^{21} is an amino acid or is absent; X^{22} is an amino acid or is absent; X^{23} is G or R; X^{24} is a sequence of 1 to 43 amino acids or is absent; Z^2 is a carboxy terminal modifying group or is absent. In an embodiment, the peptide has a length of 8 to 12 amino acids. In an embodiment, X^{22} is an amino acid, in a further embodiment valine (V). In an embodiment, X^{21} is an amino acid, in a further embodiment arginine (R). In an embodiment, X^{23} is glycine (G), in another embodiment X^{23} is arginine (R). In an embodiment, the above-mentioned peptide comprises the sequence RVSLPTSPG (SEQ ID NO:25). In an embodiment, the above-mentioned peptide is RVSLPTSPG (SEQ ID NO:25). In another embodiment, the above-mentioned peptide comprises the sequence RVSLPTSPR (SEQ ID NO:26). In an embodiment, the above-mentioned peptide is RVSLPTSPR (SEQ ID NO:26).

In another aspect, the present invention provides a method of treating cancer, said method comprising administering to a subject in need thereof an effective amount of CD8 T lymphocytes recognizing a MHC class I molecule of the HLA-A*0301 allele loaded with the above-mentioned peptide of sequence VI.

In an embodiment, the above-mentioned method further comprising determining whether said subject expresses a KIAA0226L nucleic acid comprising a G or A at a position corresponding to nucleotide 1059 in the nucleic acid sequence of human KIAA0226L (FIGs. 6A and 6B, NCBI Reference Sequence: NM_025113.2) and/or an KIAA0226L polypeptide having a glycine or an arginine at a position corresponding to residue 152 in the KIAA0226L protein sequence (FIG. 6C, NCBI Reference Sequence: NP_079389.2), wherein (a) if said subject expresses a KIAA0226L nucleic acid comprising a G at a position corresponding to nucleotide 1059 in the nucleic acid sequence of human KIAA0226L and/or a KIAA0226L polypeptide comprising a glycine residue at a position corresponding to residue 152 in the protein sequence of human KIAA0226L, X^{23} is G in said peptide of sequence VI; (b) if said subject expresses a

KIAA0226L nucleic acid comprising an A at a position corresponding to nucleotide 1059 in the nucleic acid sequence of human a KIAA0226L and/or a KIAA0226L polypeptide comprising an arginine residue at a position corresponding to residue 152 in the protein sequence of human KIAA0226L, X²³ is R in said peptide of sequence VI.

5 In another embodiment, the determining comprises sequencing a human KIAA0226L nucleic acid. In an embodiment, the above-mentioned CD8 T lymphocytes are *in vitro* expanded CD8 T lymphocytes.

In an embodiment, the above-mentioned method further comprises: (i) if said subject is the subject of (a) above, culturing CD8 T lymphocytes from a second subject comprising an A at
10 a position corresponding to nucleotide to nucleotide 1059 in the nucleic acid sequence of human KIAA0226L and/or a KIAA0226L polypeptide comprising an arginine residue at a position corresponding to residue 152 in the protein sequence of human KIAA0226L in the presence of cells expressing a MHC class I molecule of the HLA-A*0301 allele loaded with the peptide of sequence VI, wherein X²³ is G in said peptide, under conditions suitable for CD8 T
15 lymphocyte expansion; or (ii) if said subject is the subject of (b) above, culturing CD8 T lymphocytes from a second subject comprising a G at a position corresponding to nucleotide 1059 in the nucleic acid sequence of human KIAA0226L and/or a KIAA0226L polypeptide comprising a glycine residue at a position corresponding to residue 152 in the protein sequence of human KIAA0226L in the presence of cells expressing a MHC class I molecule of the HLA-
20 A*0301 allele loaded with the peptide of sequence VI, wherein X²³ is R in said peptide, under conditions suitable for CD8 T lymphocyte expansion.

In an embodiment, the above-mentioned subject is an allogeneic stem cell transplantation (ASCT) recipient.

In another aspect, the present invention provides a method of expanding CD8 T
25 lymphocytes for adoptive T-cell immunotherapy, said method comprising: (a) determining whether a candidate donor expresses a KIAA0226L nucleic acid comprising a G or A at a position corresponding to nucleotide 1059 in the nucleic acid sequence of human KIAA0226L (FIGs. 6A and 6B, NCBI Reference Sequence: NM_025113.2) and/or a KIAA0226L polypeptide having a glycine or an arginine at a position corresponding to residue 152 in the KIAA0226L protein sequence (FIG. 6C, NCBI Reference Sequence: NP_079389.2) and (b)(i) if
30 said candidate donor expresses a KIAA0226L nucleic acid comprising G at a position corresponding to nucleotide 1059 in the nucleic acid sequence of human KIAA0226L and/or a KIAA0226L polypeptide comprising a glycine residue at a position corresponding to residue 152 in the protein sequence of human KIAA0226L, culturing CD8 T lymphocytes from said
35 candidate donor in the presence of cells expressing a MHC class I molecule of the HLA-A*0301 allele loaded with the peptide of sequence VI, wherein X²³ is R in said peptide, under conditions suitable for CD8 T lymphocyte expansion; or (b)(ii) if said subject expresses a KIAA0226L

nucleic acid comprising an A at a position corresponding to nucleotide 1059 in the nucleic acid sequence of human a KIAA0226L and/or a KIAA0226L polypeptide comprising an arginine residue at a position corresponding to residue 152 in the protein sequence of human KIAA0226L, culturing CD8 T lymphocytes from said candidate donor in the presence of cells
 5 expressing a MHC class I molecule of the HLA-A*0301 allele loaded with the peptide of sequence VI, wherein X^{23} is G in said peptide, under conditions suitable for CD8 T lymphocyte expansion.

In another aspect, the present invention provides a peptide of 50 amino acids or less comprising the sequence VII

10 $Z^1-X^{25}X^{26}X^{27}GNPGTFX^{28}X^{29}-Z^2$ (VII)

wherein Z^1 is an amino terminal modifying group or is absent; X^{25} is a sequence of 1 to 43 amino acids or is absent; X^{26} is an amino acid or is absent; X^{27} is an amino acid or is absent; X^{28} is K or N; X^{29} is a sequence of 1 to 43 amino acids or is absent; Z^2 is a carboxy terminal modifying group or is absent. In an embodiment, the peptide has a length of 8 to 12 amino
 15 acids. In an embodiment, X^{27} is an amino acid, more specifically methionine (M). In an embodiment, X^{26} is an amino acid, more specifically valine (V). In an embodiment, X^{28} is K. In another embodiment, X^{28} is N. In an embodiment, the peptide comprises the sequence VMGNPGTFK (SEQ ID NO: 27). In a further embodiment, the peptide is VMGNPGTFK (SEQ ID NO: 27). In another embodiment, the peptide comprises the sequence VMGNPGTFN (SEQ ID
 20 NO: 28). In an embodiment, the peptide is VMGNPGTFN (SEQ ID NO: 28).

In another aspect, the present invention provides a method of treating cancer, said method comprising administering to a subject in need thereof an effective amount of CD8 T lymphocytes recognizing a MHC class I molecule of the HLA-A*0301 allele loaded with the above-mentioned peptide of sequence VII.

25 In an embodiment, the above-mentioned method further comprises determining whether said subject expresses an RMDN1 nucleic acid comprising an A or C at a position corresponding to nucleotide 316 in the nucleic acid sequence of human RMDN1 (FIG. 7A, NCBI Reference Sequence: NM_016033.2) and/or an RMDN1 polypeptide having a lysine or an asparagine at a position corresponding to residue 52 in the RMDN1 protein sequence (FIG. 7B,
 30 NCBI Reference Sequence: NP_057117.2), wherein (a) if said subject expresses an RMDN1 nucleic acid comprising an A at a position corresponding to nucleotide 316 in the nucleic acid sequence of human RMDN1 and/or an RMDN1 polypeptide comprising a lysine residue at a position corresponding to residue 52 in the protein sequence of human RMDN1, X^{28} is K in said peptide of sequence VII; (b) if said subject expresses an RMDN1 nucleic acid comprising a C at
 35 a position corresponding to nucleotide 316 in the nucleic acid sequence of human RMDN1 and/or a RMDN1 polypeptide comprising an asparagine residue at a position corresponding to residue 52 in the protein sequence of human RMDN1, X^{28} is N in said peptide of sequence VII.

In another embodiment, the determining comprises sequencing a human RMDN1 nucleic acid. In an embodiment, the above-mentioned CD8 T lymphocytes are *in vitro* expanded CD8 T lymphocytes.

In an embodiment, the above-mentioned method further comprises: (i) if said subject is
 5 the subject of (a), culturing CD8 T lymphocytes from a second subject comprising a C at a position corresponding to nucleotide 316 in the nucleic acid sequence of human RMDN1 and/or a RMDN1 polypeptide comprising an asparagine residue at a position corresponding to residue 52 in the protein sequence of human RMDN1 in the presence of cells expressing a MHC class I molecule of the HLA-A*0301 allele loaded with the peptide of sequence VII, wherein X²⁸ is K in
 10 said peptide, under conditions suitable for CD8 T lymphocyte expansion; or (ii) if said subject is the subject of (b) above, culturing CD8 T lymphocytes from a second subject comprising an A at a position corresponding to nucleotide 316 in the nucleic acid sequence of human RMDN1 and/or a RMDN1 polypeptide comprising a lysine residue at a position corresponding to residue 52 in the protein sequence of human RMDN1 in the presence of cells expressing a MHC class I
 15 molecule of the HLA-A*0301 allele loaded with the peptide of sequence VII, wherein X²⁸ is N in said peptide, under conditions suitable for CD8 T lymphocyte expansion.

In an embodiment, the above-mentioned subject is an allogeneic stem cell transplantation (ASCT) recipient.

In another aspect, the present invention provides a method of expanding CD8 T
 20 lymphocytes for adoptive T-cell immunotherapy, said method comprising: (a) determining whether a candidate donor expresses an RMDN1 nucleic acid comprising an A or C at a position corresponding to nucleotide 316 in the nucleic acid sequence of human RMDN1 (FIG. 7A, NCBI Reference Sequence: NM_016033.2) and/or an RMDN1 polypeptide having a lysine or an asparagine at a position corresponding to residue 52 in the RMDN1 protein sequence
 25 (FIG. 7B, NCBI Reference Sequence: NP_057117.2) and (b)(i) if said subject expresses an RMDN1 nucleic acid comprising an A at a position corresponding to nucleotide 316 in the nucleic acid sequence of human RMDN1 and/or an RMDN1 polypeptide comprising a lysine residue at a position corresponding to residue 52 in the protein sequence of human RMDN1, culturing CD8 T lymphocytes from said candidate donor in the presence of cells expressing a
 30 MHC class I molecule of the HLA-A*0301 allele loaded with the peptide of sequence VII, wherein X²⁸ is N in said peptide, under conditions suitable for CD8 T lymphocyte expansion; or (b)(ii) if said subject expresses an RMDN1 nucleic acid comprising a C at a position corresponding to nucleotide 316 in the nucleic acid sequence of human RMDN1 and/or a RMDN1 polypeptide comprising an asparagine residue at a position corresponding to residue
 35 52 in the protein sequence of human RMDN1, culturing CD8 T lymphocytes from said candidate donor in the presence of cells expressing a MHC class I molecule of the HLA-A*0301 allele

loaded with the peptide of sequence VII, wherein X^{28} is R in said peptide, under conditions suitable for CD8 T lymphocyte expansion.

In an embodiment, Z^1 is absent in the above-mentioned peptide (I)-(VII). In an embodiment, Z^2 is absent in the above-mentioned peptide (I)-(VII).

5 In another aspect, the present invention provides a MiHA identified by the above-mentioned method. In an embodiment, the MiHA is a peptide of sequences (I)-(VII) as defined herein.

In another aspect, the present invention provides a nucleic acid encoding the above-mentioned peptide (I)-(VII).

10 In another aspect, the present invention provides an isolated major histocompatibility complex (MHC) class I molecule loaded with the peptide (I)-(VII). In another aspect, the present invention provides an isolated cell expressing at its surface a MHC class I molecule loaded with the above-mentioned peptide (I)-(VII).

In an embodiment, the major histocompatibility complex (MHC) class I molecule is of the HLA-A*0301, HLA-B*0801 or HLA-B*4403 allele. In a further embodiment, the major histocompatibility complex (MHC) class I molecule is of the HLA-A*0301 allele. In another embodiment, the major histocompatibility complex (MHC) class I molecule is of the HLA-B*0801 allele. In another embodiment, the major histocompatibility complex (MHC) class I molecule is of the HLA-B*4403 allele.

20 Other objects, advantages and features of the present invention will become more apparent upon reading of the following non-restrictive description of specific embodiments thereof, given by way of example only with reference to the accompanying drawings.

BRIEF DESCRIPTION OF DRAWINGS

In the appended drawings:

25 **FIGs. 1A to C** show the nucleotide sequence of human centromer protein F, 350/400kDa (mitosin) (CENPF) cDNA (SEQ ID NO:1). The coding region is in italics;

FIG. 1D shows the amino acid sequence of human CENPF polypeptide (SEQ ID NO:2);

FIG. 2A shows the nucleotide sequence of human ZW10 interactor (ZWINT) cDNA (SEQ ID NO:3). The coding region is in italics;

FIG. 2B shows the amino acid sequence of human ZWINT polypeptide (SEQ ID NO:4);

FIG. 3A shows the nucleotide sequence of human mitochondrial carrier homolog 2 (MTCH2) cDNA (SEQ ID NO:5). The coding region is in italics;

35 **FIG. 3B** shows the amino acid sequence of human MTCH2 polypeptide (SEQ ID NO:6);

FIG. 4A and 4B show the nucleotide sequence of human ELF1 [E74-like factor 1 (ets domain transcription factor)] cDNA (SEQ ID NO:7). The coding region is in italics;

FIG. 4C shows the amino acid sequence of human ELF1 polypeptide (SEQ ID NO:8);

FIGs. 5A and 5B show the nucleotide sequence of human NQO1 [NAD(P)H dehydrogenase, quinone 1] cDNA (SEQ ID NO:9). The coding region is in italics;

FIG. 5C shows the amino acid sequence of human NQO1 polypeptide (SEQ ID NO:10);

FIG. 6A and 6B show the nucleotide sequence of human KIAA0226L cDNA (SEQ ID NO:11). The coding region is in italics;

FIG. 6C shows the amino acid sequence of human KIAA0226L polypeptide (SEQ ID NO:12);

FIG. 7A shows the nucleotide sequence of human RMDN1 cDNA (SEQ ID NO:13). The coding region is in italics; and

FIG. 7B shows the amino acid sequence of human RMDN1 polypeptide (SEQ ID NO:14).

DISCLOSURE OF INVENTION

Terms and symbols of genetics, molecular biology, biochemistry and nucleic acid used herein follow those of standard treatises and texts in the field, *e.g.* Kornberg and Baker, DNA Replication, Second Edition (W.H. Freeman, New York, 1992); Lehninger, Biochemistry, Second Edition (Worth Publishers, New York, 1975); Strachan and Read, Human Molecular Genetics, Second Edition (Wiley-Liss, New York, 1999); Eckstein, editor, Oligonucleotides and Analogs: A Practical Approach (Oxford University Press, New York, 1991); Gait, editor, Oligonucleotide Synthesis: A Practical Approach (IRL Press, Oxford, 1984); and the like. All terms are to be understood with their typical meanings established in the relevant art.

The articles "a" and "an" are used herein to refer to one or to more than one (i.e. to at least one) of the grammatical object of the article. By way of example, "an element" means one element or more than one element. Throughout this specification, unless the context requires otherwise, the words "comprise," "comprises" and "comprising" will be understood to imply the inclusion of a stated step or element or group of steps or elements but not the exclusion of any other step or element or group of steps or elements.

Described herein is a novel method for human MiHA discovery, novel MiHAs identified using this method, as well as uses of the novel MiHAs. One of the features of the method is the inclusion of personalized translated transcriptome and/or exome in the database used for peptide identification by mass spectroscopy (MS). Candidate MiHAs are identified by comparing the personalized transcriptome and/or exome to a reference genome and/or to the transcriptome and/or exome of an HLA-matched subject (*e.g.*, an HLA-identical sibling).

Accordingly, in a first aspect, the present invention provides a method of identifying a minor histocompatibility antigen (MiHA) candidate, the method comprising: (a) isolating and determining the sequence of MHC-associated peptides (MAPs) in a first cell sample from a first subject; (b) performing a whole transcriptome and/or exome sequencing on a second cell sample obtained from said first subject; (c) comparing the sequenced whole transcriptome and/or exome to a reference genome to identify single nucleotide variations (SNVs) between the transcriptome and/or exome of said first subject and the reference genome; (d) *in silico* translating the sequences containing the identified SNVs to identify peptide sequences comprising at least one non-synonymous mutation caused by said SNVs; (e) comparing the sequences of the MAPs isolated in (a) with the peptide sequences identified in (d); and (f) identifying a MiHA candidate based on said comparison.

In an embodiment, the MiHA candidate is a MAP whose sequence comprises at least one mutation relative to the corresponding sequence translated from the reference genome.

In another aspect, the present invention provides a method of identifying a minor histocompatibility antigen (MiHA) candidate, the method comprising: (a) isolating MHC-associated peptides (MAPs) in a first cell sample from a first and second subjects, wherein said first and second subjects are human leukocyte antigen (HLA)-matched; (b) performing a whole transcriptome and/or exome sequencing on a second cell sample obtained from said first and second subjects; (c) comparing the sequenced whole transcriptomes and/or exomes to identify single nucleotide variations (SNVs) between the transcriptomes and/or exomes of said first and second subjects; (d) *in silico* translating the sequences containing the identified SNVs to identify peptide sequences comprising at least one non-synonymous mutation caused by said SNVs; (e) comparing the sequences of the MAPs isolated in (a) with the peptide sequences identified in (d); and (f) identifying a MiHA candidate based on said comparison. In an embodiment, the MiHA candidate is a MAP present in the first cell sample from said first subject but absent from the first cell sample from said second subject.

The term "reference genome" as used herein refers to the human genome assemblies reported in the literature, and includes for example the Genome Reference Consortium Human Build 37 (GRCh37, Genome Reference Consortium; The International Human Genome Sequencing Consortium. *Nature*. 2004; 431:931–945), Hs_Celera_WGSA (Celera Genomics; Istrail S. *et al.*, *Proc Natl Acad Sci USA*. 2004 Feb 17;101(7):1916-21). Epub 2004 Feb 9), HuRef and HuRefPrime (J. Craig Venter Institute; Levy S, *et al.* *PLoS Biology*. 2007;5:2113–2144), YH1 and BGIAF (Beijing Genomics Institute; Li R, *et al.* *Genome Research*. 2010;20:265–272), as well as HsapALLPATHS1 (Broad Institute). In an embodiment, the reference genome is GRCh37.

In various embodiments, the above-noted first sample may be from any source that contains cells expressing MHC class I molecules, including a tissue or body fluid from the

subject, such as blood, serum, immune cells (e.g., lymphocytes), blood cells (e.g., PBMCs or a subset thereof), tissues, or a cell line derived from primary cells. In an embodiment, the first sample is a blood cell sample, for example a PBMC sample, or a cell line derived from blood cells such as PBMCs (e.g., an immortalized cell line). Methods for generating a cell line from primary cells, or for immortalizing primary cells, are known in the art and include, for example, immortalization of primary cells by recombinant expression of human telomerase reverse transcriptase (TERT) (Barsov EV, *Curr Protoc Immunol.* 2011 Nov;Chapter 7:Unit 7.21B), immortalization by recombinant expression of viral genes such as Simian virus 40 (SV40) T antigen, adenovirus E1A and E1B, human papillomavirus (HPV) E6 and E7 and Epstein-Barr Virus (EBV), as well as inactivation of tumor suppression genes such as p53 or Rb. Methods for immortalization of B lymphocytes by EBV are disclosed in Tosato G and Cohen JL. *Curr Protoc Immunol.* 2007 Feb;Chapter 7:Unit 7.22. Products/reagents for immortalizing mammalian cells are commercially available, for example from ATCC™. In an embodiment, the first sample is an immortalized cell line derived from primary cells obtained from the subject, in a further embodiment an immortalized B cell line, such as an EBV-transformed B lymphoblastoid cell line (B-LCL).

Methods for isolating MHC-associated peptides (MAPs) from a cell sample are well known in the art. The most commonly used technique is mild acid elution (MAE) of MHC-associated peptides from living cells, as described in Fortier *et al.* (*J. Exp. Med.* 205(3): 595-610, 2008). Another technique is immunoprecipitation or affinity purification of peptide-MHC class I complexes followed by peptide elution (see, e.g., Gebreselassie *et al.*, *Hum Immunol.* 2006 November; 67(11): 894-906). Two high-throughput strategies based on the latter approach have been implemented. The first is based on transfection of cell lines with expression vectors coding soluble secreted MHCs (lacking a functional transmembrane domain) and elution of peptides associated with secreted MHCs (Barnea *et al.*, *Eur J Immunol.* 2002 Jan;32(1):213-22; and Hickman HD *et al.*, *J Immunol.* 2004 Mar 1;172(5):2944-52). The second approach hinges on chemical or metabolic labeling to provide quantitative profiles of MHC-associated peptides (Weinzierl AO *et al.*, *Mol Cell Proteomics.* 2007 Jan;6(1):102-13. Epub 2006 Oct 29; Lemmel C *et al.*, *Nat Biotechnol.* 2004 Apr;22(4):450-4. Epub 2004 Mar 7; Milner E, *Mol Cell Proteomics.* 2006 Feb;5(2):357-65. Epub 2005 Nov 4).

Eluted MAPs may be subjected to any purification/enrichment steps, including size exclusion chromatography or ultrafiltration (using a filter with a cut-off of about 5000 Da, for example about 3000 Da), and/or ion exchange chromatography (e.g., cation exchange chromatography), prior to further analysis. The sequence of the eluted MAPs may be determined using any method known in the art for sequencing peptides/proteins, such as mass spectroscopy (as described below) and the Edman degradation reaction.

In various embodiments, the above-noted second sample may be from any source that contains genomic DNA, RNA, and/or proteins, for example a tissue or body fluid from the subject, such as blood, serum, immune cells (e.g., lymphocytes), blood cells (e.g., PBMCs), tissues, or a cell line derived from primary cells (as described above). In an embodiment, the second sample is an immortalized cell line derived from primary cells obtained from the subject, in a further embodiment an immortalized B cell line, such as an EBV-transformed B lymphoblastoid cell line (B-LCL). The cell sample may be subjected to commonly used isolation and/or purification techniques for enrichment in nucleic acids (genomic DNA, mRNA) and/or proteins.

In an embodiment, transcriptome libraries are generated/constructed from the RNA obtained from the sample. Transcriptome library construction may include one or more of the following steps: poly-A mRNA enrichment/purification; RNA fragmentation and priming for cDNA synthesis; reverse transcription (RT) (using random primers); second round of RT to generate a double-stranded cDNA, cDNA purification; end repair of fragmented cDNA, adenylation of the 3' ends, ligation of adaptors and enrichment of DNA fragments containing adapter molecules. Kits suitable for transcriptome library construction are commercially available, for example from Life Technologies (Ambion® RNA-Seq Library Construction Kit), Applied Biosystems (AB Library Builder™ Whole Transcriptome Core Kit), Qiagen (QuantiTect™ Whole Transcriptome Kit) and Sigma-Aldrich (TransPlex® Complete Whole Transcriptome Amplification Kit)

In an embodiment, genomic libraries are generated/constructed from the genomic DNA obtained from the sample. Genomic library construction may include one or more of the following steps: DNA shearing, DNA end repair, 3' ends adenylation, ligation of adaptors, purification of ligation products and amplification (e.g., PCR) to enrich DNA fragments that have adapter molecules. Kits suitable for genomic library construction are commercially available, for example from Illumina (TruSeq™ DNA Sample Preparation Kit (v2) (Cat. No. FC-930-1021), Life Technologies (SOLiD® Fragment Library Construction Kit and New England BioLabs (NEBNext® DNA Library Preparation).

In another embodiment, the genomic (DNA-Seq) libraries are subjected to an enrichment step to sequence only the coding portion (exome) of the human genome. Kits suitable for exome enrichment are commercially available, for example from Illumina (TruSeq™ exome enrichment kit, FC-930-1012), Life Technologies (TargetSeq™ Exome and Custom Enrichment System, A14060-A14063), FlexGen (Flexome whole exome enrichment kit v2), Roche NimbleGen (SeqCap EZ Human Exome Library v2.0) and Agilent Technologies (SureSelect All Exon kits)

Methods to perform whole transcriptome or exome sequencing (RNA-Seq) are known in the art (see, for example, Wang *et al.*, Nature Reviews Genetics 10, 57-63, January 2009; Genome Biology 2011, 12(9), Exome sequencing special issue). Various platforms for

performing whole transcriptome/exome sequencing exist, such as the Illumina Genome Analyzer platform, the Applied Biosystems (ABI) Solid™ Sequencing platform or Life Science's 454 Sequencing platform (Roche).

The identification of single nucleotide variations (SNVs) or single nucleotide polymorphisms (SNPs) between two or more sequences, for example between (i) the transcriptome and/or exome of a subject and a reference genome and/or (ii) the transcriptomes and/or exomes of two different subjects, may be performed using any sequence comparison/SNP identification methods or tools, including the SNP calling program Casava™ from Illumina, SNPdetector (Zhang et al., *PLoS Comput Biol.* 2005 October; 1(5): e53), the SNP & Variation Suite from Golden Helix, the Genome-wide human SNP arrays from Affymetrix, the SAMtools mpileup from MassGenomics, etc.

The *in silico* translation of nucleic acid sequences to protein sequences may be performed using any suitable softwares or tools, including the ExPASy Translate tool, Vector NTI™ (Life Technologies), pyGeno (Granados *et al.*, 2012), Virtual Ribosome (CBS, University of Denmark), etc. The *in silico* translation of transcriptomes and/or exomes permits the identification of peptide sequences comprising at least one non-synonymous mutation caused by the SNVs. In an embodiment, all possible amino acid (aa) sequence variants of 15 amino acids or less (in embodiments, 14, 13, 12, 11 amino acids or less) comprising at least one non-synonymous mutation are computed, listed and used in the comparing step (e). Thus, for each non-synonymous mutation caused by the SNVs, a window of 90 bp (84, 78, 72 or 66 bp) around each one of the polymorphic positions is computed to obtain the list of every possible amino acid (aa) sequence variant defined by these 90 bp (84, 78, 72 or 66 bp) (30, 28, 26, 24 or 22 aa) windows. In this way, a list of most possible aa sequences of at most 15 aa (in embodiments, 14, 13, 12, 11 aa) affected by non-synonymous polymorphisms may be obtained. For the identification of MHC class II-associated MAPs (which may be longer, up to about 30 amino acids), all possible amino acid (aa) sequence variants of for example 30 amino acids or less (in embodiments, 25, 20, 19, 18, 17, 16, 15, 14, 13, 12 amino acids or less) comprising at least one non-synonymous mutation are computed, listed and used in the comparing step (e). Thus, for each non-synonymous mutation caused by the SNVs, a window of 180 bp around each one of the polymorphic positions is computed to obtain the list of every possible amino acid (aa) sequence variant defined by these 180 bp (60 aa) windows. In this way, a list of most possible aa sequences of at most 30 aa affected by non-synonymous polymorphisms may be obtained.

In an embodiment, all possible amino acid (aa) sequence variants of 12 or 11 amino acids or less comprising at least one non-synonymous mutation are computed, listed and used in the comparing step (e). Thus, for each non-synonymous mutation caused by the SNVs, a window of 72 or 66 bp around each one of the polymorphic positions is computed to obtain the

list of every possible amino acid (aa) sequence variant defined by these 72 or 66 bp (24 or 22 aa) windows. In this way, a list of most possible aa sequences of at most 12 or 11 aa affected by non-synonymous polymorphisms was obtained.

In embodiment, the above-noted peptide sequences have a length of about 7 to about 15 amino acids (e.g., 7, 8, 9, 10, 11, 12, 13, 14 or 15), in a further embodiment of about 8 or 9 to about 11 or 12 amino acids (e.g., 8, 9, 10, 11 or 12).

In an embodiment, the comparison of the sequences of the MAPs isolated from the first sample with the peptide sequences derived from the transcriptome and/or exome identified above (i.e. comprising at least one non-synonymous mutation caused by the SNVs) comprises subjecting the isolated MAPs to mass spectrometry and comparing the MS spectra obtained with the peptide sequences derived from the transcriptome and/or exome. In an embodiment, the mass spectrometry is liquid chromatography–mass spectrometry (LC-MS), in a further embodiment LC-MS coupled to peptide mass fingerprinting (LC-MS/MS).

In an embodiment, the method further comprises determining the binding of the MiHA candidate identified to a MHC class I molecule. The binding may be a predicted binding affinity (IC_{50}) of peptides to the allelic products, which may be obtained using tools such as the NetMHCcons software version 1.0 (<http://www.cbs.dtu.dk/services/NetMHCcons/>) (Karosiene et al., 2011). An overview of the various available MHC class I peptide binding tools is provided in Peters B *et al.*, *PLoS Comput Biol* 2006, **2**(6):e65; Trost et al., *Immunome Res* 2007, **3**(1):5; Lin *et al.*, *BMC Immunology* 2008, **9**:8)

In an embodiment, peptides with a predicted IC_{50} below 50 nM are considered as strong binders and peptides with an IC_{50} between about 50 and about 500 nM are considered as weak binders.

The binding of the MiHA candidate identified to a MHC class I molecule may be determined using other known methods, for example the T2 Peptide Binding Assay. T2 cell lines are deficient in TAP but still express low amounts of MHC class I on the surface of the cells. The T2 binding assay is based upon the ability of peptides to stabilize the MHC class I complex on the surface of the T2 cell line. T2 cells are incubated with a specific peptide (e.g., a candidate MiHA), stabilized MHC class I complexes are detected using a pan-HLA class I antibody, an analysis is carried out (by flow cytometry, for example) and binding is assessed in relation to a non-binding negative control. The presence of stabilized peptide/MHC class I complexes at the surface is indicative that the peptide (e.g., candidate MiHA) binds to MHC class I molecules.

The binding of a peptide of interest (e.g., candidate MiHA) to MHC may also be assessed based on its ability to inhibit the binding of a radiolabeled probe peptide to MHC molecules. MHC molecules are solubilized with detergents and purified by affinity chromatography. They are then incubated for 2 days at room temperature with the inhibitor

peptide (e.g., candidate MiHA) and an excess of a radiolabeled probe peptide, in the presence of a cocktail of protease inhibitors. At the end of the incubation period, MHC-peptide complexes are separated from unbound radiolabeled peptide by size-exclusion gel-filtration chromatography, and the percent bound radioactivity is determined. The binding affinity of a particular peptide for an MHC molecule may be determined by co-incubation of various doses of unlabeled competitor peptide with the MHC molecules and labeled probe peptide. The concentration of unlabeled peptide required to inhibit the binding of the labeled peptide by 50% (IC₅₀) can be determined by plotting dose versus % inhibition (see, e.g., *Current Protocols in Immunology* (1998) 18.3.1-18.3.19, John Wiley & Sons, Inc.).

The binding of the MiHA candidate identified to a MHC class I molecule may be determined using an epitope discovery system, such as the ProImmune REVEAL & ProVE® epitope discovery system.

In another aspect, the present invention provides a peptide (e.g., an isolated or synthetic peptide) of 50 amino acids or less comprising the sequence (I) to (VII) described herein.

In another aspect, the present invention provides a peptide (e.g., an isolated or synthetic peptide) of 50 amino acids or less comprising the sequence (I):



wherein Z¹ is an amino terminal modifying group or is absent; X¹ is a sequence of 1 to 43 amino acids or is absent; X² is L or S; X³ is a sequence of 1 to 43 amino acids or is absent; and Z² is a carboxy terminal modifying group or is absent.

In another aspect, the present invention provides a peptide (e.g., an isolated or synthetic peptide) of 50 amino acids or less comprising the sequence (II):



wherein Z¹ is an amino terminal modifying group or is absent; X⁴ is a sequence of 1 to 43 amino acids or is absent; X⁵ is G or R; X⁶ is an amino acid or is absent; X⁷ is a sequence of 1 to 43 amino acids or is absent; and Z² is a carboxy terminal modifying group or is absent.

In another aspect, the present invention provides a peptide (e.g., an isolated or synthetic peptide) of 50 amino acids or less comprising the sequence (III):



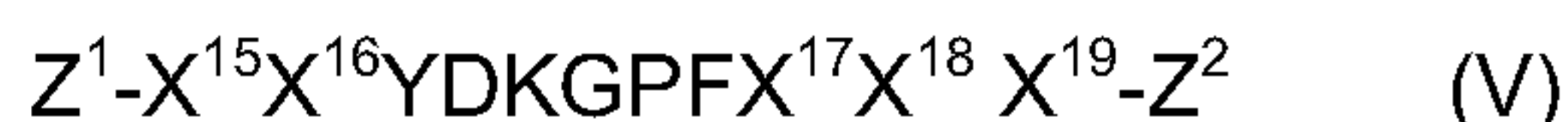
wherein Z¹ is an amino terminal modifying group or is absent; X⁸ is a sequence of 1 to 43 amino acids or is absent; X⁹ is P or A; X¹⁰ is a sequence of 1 to 43 amino acids or is absent; and Z² is a carboxy terminal modifying group or is absent.

In another aspect, the present invention provides a peptide (e.g., an isolated or synthetic peptide) of 50 amino acids or less comprising the sequence (IV):



wherein Z^1 is an amino terminal modifying group or is absent; X^{11} is a sequence of 1 to 43 amino acids or is absent; X^{12} is S or T; X^{13} is an amino acid or is absent; X^{14} is a sequence of 1 to 43 amino acids or is absent; and Z^2 is a carboxy terminal modifying group or is absent.

In another aspect, the present invention provides a peptide (e.g., an isolated or synthetic peptide) of 50 amino acids or less comprising the sequence (V):



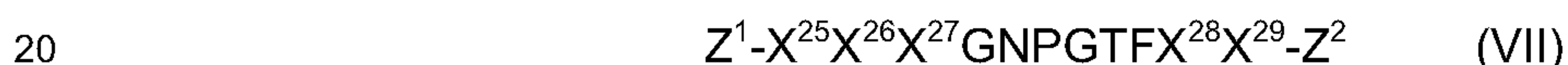
wherein Z^1 is an amino terminal modifying group or is absent; X^{15} is a sequence of 1 to 43 amino acids or is absent; X^{16} is an amino acid or is absent; X^{17} is R or W; X^{18} is an amino acid or is absent; X^{19} is a sequence of 1 to 43 amino acids or is absent; Z^2 is a carboxy terminal modifying group or is absent.

In another aspect, the present invention provides a peptide (e.g., an isolated or synthetic peptide) of 50 amino acids or less comprising the sequence (VI):



wherein Z^1 is an amino terminal modifying group or is absent; X^{20} is a sequence of 1 to 43 amino acids or is absent; X^{21} is an amino acid or is absent; X^{22} is an amino acid or is absent; X^{23} is G or R; X^{24} is a sequence of 1 to 43 amino acids or is absent; Z^2 is a carboxy terminal modifying group or is absent.

In another aspect, the present invention provides a peptide (e.g., an isolated or synthetic peptide) of 50 amino acids or less comprising the sequence (VII):



wherein Z^1 is an amino terminal modifying group or is absent; X^{25} is a sequence of 1 to 43 amino acids or is absent; X^{26} is an amino acid or is absent; X^{27} is an amino acid or is absent; X^{28} is K or N; X^{29} is a sequence of 1 to 43 amino acids or is absent; Z^2 is a carboxy terminal modifying group or is absent.

In another aspect, the present invention provides a peptide of 50 amino acids or less comprising any one of the sequences (I) to (VII) as defined herein.

In general, peptides presented in the context of HLA class I vary in length from about 7 to about 15 amino acid residues, and a longer peptide (e.g., of 50 amino acids or less) can be enzymatically processed to a peptide of such length. In embodiments, the peptide is 45, 40, 35, 30, 25, 20 or 15 amino acids or less. A peptide comprising the above-noted sequence/motif provided by the invention typically is at least 7 amino acids in length but preferably at least 8 or 9 amino acids. The upper length of a peptide provided by the invention is no more than 15 amino acids, but preferably no more than about 13, 12 or 11 amino acids in length. In embodiments, the above-mentioned peptide is about 8 to 12 amino acids long (e.g., 8, 9, 10, 11 or 12 amino acids long), small enough for a direct fit in an HLA class I molecule, but it may also be larger, between 12 to about 20, 25, 30, 35, 40, 45 or 50 amino acids and presented by HLA

molecules only after cellular uptake and intracellular processing by the proteasome and transport before presentation in the groove of an MHC molecule.

The term "amino acid" as used herein includes both L- and D-isomers of the naturally occurring amino acids as well as other amino acids (e.g., naturally-occurring amino acids, non-naturally-occurring amino acids, amino acids which are not encoded by nucleic acid sequences, etc.) used in peptide chemistry to prepare synthetic analogs of peptides. Examples of naturally occurring amino acids are glycine, alanine, valine, leucine, isoleucine, serine, threonine, etc.

Other amino acids include for example non-genetically encoded forms of amino acids, as well as a conservative substitution of an L-amino acid. Naturally-occurring non-genetically encoded amino acids include, for example, beta-alanine, 3-amino-propionic acid, 2,3-diaminopropionic acid, alpha-aminoisobutyric acid (Aib), 4-amino-butyric acid, *N*-methylglycine (sarcosine), hydroxyproline, ornithine (e.g., L-ornithine), citrulline, *t*-butylalanine, *t*-butylglycine, *N*-methyloleucine, phenylglycine, cyclohexylalanine, norleucine (Nle), norvaline, 2-naphthylalanine, pyridylalanine, 3-benzothienyl alanine, 4-chlorophenylalanine, 2-fluorophenylalanine, 3-fluorophenylalanine, 4-fluorophenylalanine, penicillamine, 1,2,3,4-tetrahydro-isoquinoline-3-carboxylic acid, beta-2-thienylalanine, methionine sulfoxide, L-homoarginine (Hoarg), *N*-acetyl lysine, 2-amino butyric acid, 2-amino butyric acid, 2,4,-diaminobutyric acid (D- or L-), *p*-aminophenylalanine, *N*-methylvaline, homocysteine, homoserine (HoSer), cysteic acid, epsilon-amino hexanoic acid, delta-amino valeric acid, or 2,3-diaminobutyric acid (D- or L-), etc. These amino acids are well known in the art of biochemistry/peptide chemistry.

In embodiments, the peptides of the present invention include peptides with altered sequences containing substitutions of functionally equivalent amino acid residues, relative to the above-mentioned sequences. For example, one or more amino acid residues within the sequence can be substituted by another amino acid of a similar polarity (having similar physico-chemical properties) which acts as a functional equivalent, resulting in a silent alteration. Substitution for an amino acid within the sequence may be selected from other members of the class to which the amino acid belongs. For example, positively charged (basic) amino acids include arginine, lysine and histidine (as well as homoarginine and ornithine). Nonpolar (hydrophobic) amino acids include leucine, isoleucine, alanine, phenylalanine, valine, proline, tryptophan and methionine. Uncharged polar amino acids include serine, threonine, cysteine, tyrosine, asparagine and glutamine. Negatively charged (acidic) amino acids include glutamic acid and aspartic acid. The amino acid glycine may be included in either the nonpolar amino acid family or the uncharged (neutral) polar amino acid family. Substitutions made within a family of amino acids are generally understood to be conservative substitutions.

The above-mentioned peptide may comprise all L-amino acids, all D-amino acids or a mixture of L- and D-amino acids. In an embodiment, the above-mentioned peptide comprises all L-amino acids.

The peptide may also be N- and/or C-terminally capped or modified to prevent degradation, increase stability or uptake. In an embodiment, the amino terminal residue (*i.e.*, the free amino group at the N-terminal end) of the peptide is modified (*e.g.*, for protection against degradation), for example by covalent attachment of a moiety/chemical group (Z^1). Z^1 may be a straight chained or branched alkyl group of one to eight carbons, or an acyl group ($R-CO-$), wherein R is a hydrophobic moiety (*e.g.*, acetyl, propionyl, butanyl, iso-propionyl, or iso-butanyl), or an aroyl group ($Ar-CO-$), wherein Ar is an aryl group. In an embodiment, the acyl group is a C_1-C_{16} or C_3-C_{16} acyl group (linear or branched, saturated or unsaturated), in a further embodiment, a saturated C_1-C_6 acyl group (linear or branched) or an unsaturated C_3-C_6 acyl group (linear or branched), for example an acetyl group (CH_3-CO- , Ac). In an embodiment, Z^1 is absent.

The carboxy terminal residue (*i.e.*, the free carboxy group at the C-terminal end of the peptide) of the peptide may be modified (*e.g.*, for protection against degradation), for example by amidation (replacement of the OH group by a NH_2 group), thus in such a case Z^2 is a NH_2 group. In an embodiment, Z^2 may be an hydroxamate group, a nitrile group, an amide (primary, secondary or tertiary) group, an aliphatic amine of one to ten carbons such as methyl amine, iso-butylamine, iso-valerylamine or cyclohexylamine, an aromatic or arylalkyl amine such as aniline, naphthylamine, benzylamine, cinnamylamine, or phenylethylamine, an alcohol or CH_2OH . In an embodiment, Z^2 is absent.

In an embodiment, the peptide comprises the sequence ELQEKFLSL (SEQ ID NO:15), in a further embodiment the peptide is ELQEKFLSL (SEQ ID NO:15). In another embodiment, the peptide comprises the sequence ELQEKFSSL (SEQ ID NO:16), in a further embodiment the peptide is ELQEKFSSL (SEQ ID NO:16).

In an embodiment, the peptide comprises the sequence QELDGVFQKL (SEQ ID NO:17). In a further embodiment, the peptide is QELDGVFQKL (SEQ ID NO:17). In another embodiment, the peptide comprises the sequence QELDRVFQKL (SEQ ID NO:18). In a further embodiment, the peptide is QELDRVFQKL (SEQ ID NO:18).

In an embodiment, the peptide comprises the sequence SLFFRKVPF (SEQ ID NO:19). In a further embodiment, the peptide is SLFFRKVPF (SEQ ID NO:19). In another embodiment, the peptide comprises the sequence SLFFRKVAF (SEQ ID NO:20). In a further embodiment, the peptide is SLFFRKVAF (SEQ ID NO:20).

In an embodiment, the above-mentioned peptide comprises the sequence SVLKPGNSK (SEQ ID NO:21). In a further embodiment, the peptide is SVLKPGNSK (SEQ ID NO:21). In another embodiment, the above-mentioned peptide comprises the sequence

TVLKPGNSK (SEQ ID NO:22). In a further embodiment, the peptide is TVLKPGNSK (SEQ ID NO:22).

In an embodiment, the above-mentioned peptide comprises the sequence AMYDKGPFRSK (SEQ ID NO:23). In an embodiment, the above-mentioned peptide is
 5 AMYDKGPFRSK (SEQ ID NO:23). In another embodiment, the above-mentioned peptide comprises the sequence AMYDKGPFWK (SEQ ID NO:24). In an embodiment, the above-mentioned peptide is AMYDKGPFWK (SEQ ID NO:24).

In an embodiment, the above-mentioned peptide comprises the sequence RVSLPTSPG (SEQ ID NO:25). In an embodiment, the above-mentioned peptide is
 10 RVSLPTSPG (SEQ ID NO:25). In another embodiment, the above-mentioned peptide comprises the sequence RVSLPTSPR (SEQ ID NO:26). In an embodiment, the above-mentioned peptide is RVSLPTSPR (SEQ ID NO:26).

In an embodiment, the above-mentioned peptide comprises the sequence VMGNPGTFK (SEQ ID NO:27). In an embodiment, the above-mentioned peptide is
 15 VMGNPGTFK (SEQ ID NO:27). In another embodiment, the above-mentioned peptide comprises the sequence VMGNPGTFN (SEQ ID NO:28). In an embodiment, the above-mentioned peptide is VMGNPGTFN (SEQ ID NO:28).

The peptides of the invention may be produced by expression in a host cell comprising a nucleic acid encoding the peptides (recombinant expression) or by chemical synthesis (e.g.,
 20 solid-phase peptide synthesis). Peptides can be readily synthesized by manual and/or automated solid phase procedures well known in the art. Suitable syntheses can be performed for example by utilizing "T-boc" or "Fmoc" procedures. Techniques and procedures for solid phase synthesis are described in for example Solid Phase Peptide Synthesis: A Practical Approach, by E. Atherton and R. C. Sheppard, published by IRL, Oxford University Press, 1989.
 25 Alternatively, the peptides may be prepared by way of segment condensation, as described, for example, in Liu *et al.*, *Tetrahedron Lett.* 37: 933-936, 1996; Baca *et al.*, *J. Am. Chem. Soc.* 117: 1881-1887, 1995; Tam *et al.*, *Int. J. Peptide Protein Res.* 45: 209-216, 1995; Schnolzer and Kent, *Science* 256: 221-225, 1992; Liu and Tam, *J. Am. Chem. Soc.* 116: 4149-4153, 1994; Liu and Tam, *Proc. Natl. Acad. Sci. USA* 91: 6584-6588, 1994; and Yamashiro and Li, *Int. J. Peptide Protein Res.* 31: 322-334, 1988). Other methods useful for synthesizing the peptides
 30 are described in Nakagawa *et al.*, *J. Am. Chem. Soc.* 107: 7087-7092, 1985.

Peptides comprising naturally occurring amino acids encoded by the genetic code may also be prepared using recombinant DNA technology using standard methods.

Accordingly, in another aspect, the invention further provides a nucleic acid (isolated)
 35 encoding the above-mentioned peptides of sequences I-VII. In an embodiment, the nucleic acid does not encode the full-length CENPF, ZWINT, MTCH2, ELF1, NQO1, KIAA0226L or RMDN1 polypeptide. In an embodiment, the nucleic acid has a length of 150 nucleotides or less, in

further embodiments 135, 120, 105, 90, 75, 60, 45, 42 or 39 nucleotides or less. In other embodiments, the nucleic acid comprises from about 21 nucleotides to about 45 nucleotides, from about 24 to about 36 nucleotides, for example 24, 27, 30, 33 or 36 nucleotides.

"Isolated", as used herein, refers to a peptide or nucleic molecule separated from other components that are present in the natural source of the macromolecule (other nucleic acids, proteins, lipids, sugars, etc.). "Synthetic", as used herein, refers to a peptide or nucleic molecule that is not isolated from its natural sources, e.g., which is produced through recombinant technology or using chemical synthesis.

In an embodiment, the above-mentioned peptide is substantially pure. A compound is "substantially pure" when it is separated from the components that naturally accompany it. Typically, a compound is substantially pure when it is at least 60%, more generally 75%, 80% or 85%, preferably over 90% and more preferably over 95%, by weight, of the total material in a sample. Thus, for example, a polypeptide that is chemically synthesized or produced by recombinant technology will generally be substantially free from its naturally associated components. A nucleic acid molecule is substantially pure when it is not immediately contiguous with (i.e., covalently linked to) the coding sequences with which it is normally contiguous in the naturally occurring genome of the organism from which the nucleic acid is derived. A substantially pure compound can be obtained, for example, by extraction from a natural source; by expression of a recombinant nucleic acid molecule encoding a peptide compound; or by chemical synthesis. Purity can be measured using any appropriate method such as column chromatography, gel electrophoresis, HPLC, etc.

The nucleic acid may be in a vector, such as a cloning vector or an expression vector, that may be transfected into a host cell. Alternatively, the nucleic acid may be incorporated into the genome of the host cell. In either event, the host cell expresses the nucleic acid, and in turn the encoded peptide. The invention also provides a vector or plasmid comprising the above-mentioned nucleic acid. The vector or plasmid contains the necessary elements for the transcription and translation of the inserted coding sequence, and may contain other components such as resistance genes, cloning sites, etc. Methods that are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding peptides or polypeptides and appropriate transcriptional and translational control/regulatory elements operably linked thereto. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination. Such techniques are described in Sambrook. *et al.* (1989) *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Press, Plainview, N.Y., and Ausubel, F. M. *et al.* (1989) *Current Protocols in Molecular Biology*, John Wiley & Sons, New York, N.Y.

"Operably linked" refers to a juxtaposition of components, particularly nucleotide sequences, such that the normal function of the components can be performed. Thus, a coding

sequence that is operably linked to regulatory sequences refers to a configuration of nucleotide sequences wherein the coding sequences can be expressed under the regulatory control, that is, transcriptional and/or translational control, of the regulatory sequences. "Regulatory/control region" or "regulatory/control sequence", as used herein, refers to the non-coding nucleotide sequences that are involved in the regulation of the expression of a coding nucleic acid. Thus the term regulatory region includes promoter sequences, regulatory protein binding sites, upstream activator sequences, and the like.

In another aspect, the present invention provides an MHC class I molecule loaded with the above-mentioned peptide. In an embodiment, the MHC molecule is a HLA-B8 molecule, in a further embodiment a HLA-B*0801 molecule. In an embodiment, the peptide is non-covalently bound to the MHC class I molecule (i.e., the peptide is loaded into the peptide binding groove/pocket but is not covalently attached to the MHC class I molecule). In another embodiment, the peptide is covalently attached/bound to the MHC class I molecule. In such a construct, the peptide and the MHC class I molecule are produced as a fusion protein, typically with a short (e.g., 5 to 20 residues, preferably about 10) flexible linker or spacer (e.g., a polyglycine linker). In another aspect, the invention provides a nucleic acid encoding a peptide - MHC class I fusion protein. In an embodiment, the MHC class I molecule - peptide complex is multimerized. Accordingly, in another aspect, the present invention provides a multimer of MHC class I molecule loaded (covalently or not) with the above-mentioned peptide. A great number of strategies have been developed for the production of MHC multimers, including MHC dimers, tetramers, pentamers, octamers, etc. (reviewed in Bakker and Schumacher, *Current Opinion in Immunology* 2005, 17:428-433). MHC multimers are useful, for example, for the detection and purification of antigen-specific T cells.

In yet another aspect, the present invention provides a cell (e.g., a host cell), in an embodiment an isolated cell, comprising the above-mentioned nucleic acid or vector. In another aspect, the present invention provides a cell expressing at its cell surface an MHC molecule (e.g., a HLA-B8 molecule, such as a HLA-B*0801 molecule, and/or an HLA-A3 molecule, such as a HLA-A*0301 molecule and/or an HLA-B44 allele such as HLA-B*4403) loaded with the above-mentioned peptide. In an embodiment, the host cell is a primary cell, a cell line or an immortalized cell. In another embodiment, the cell is an antigen-presenting cell (APC).

Nucleic acids and vectors can be introduced into cells via conventional transformation or transfection techniques. The terms "transformation" and "transfection" refer to techniques for introducing foreign nucleic acid into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, electroporation, microinjection and viral-mediated transfection. Suitable methods for transforming or transfecting host cells can for example be found in Sambrook *et al.* (*supra*), and other laboratory manuals. Methods for

introducing nucleic acids into mammalian cells *in vivo* are also known, and may be used to deliver the vector DNA of the invention to a subject for gene therapy.

In another embodiment, the present invention provides T cell receptor (TCR) molecules capable of interacting with the above-mentioned MHC molecule/peptide (MiHA) complex, and nucleic acid molecules encoding such TCR molecules. A TCR according to the present invention will preferably be capable of specifically interacting with an MiHA of the present invention loaded on an MHC molecule, preferably at the surface of a living cell *in vitro* or *in vivo*. T cell receptors and in particular nucleic acids encoding TCR' s according to the invention may for instance be applied to transfer a TCR from one T cell to another T cell and generate new T cell clones capable of specifically recognizing the MiHA. By this TCR cloning method, T cell clones may be provided that essentially are of the genetic make-up of an allogeneic donor, for instance a donor of lymphocytes. The method to provide T cell clones capable of recognizing an MiHA according to the invention may be generated for and can be specifically targeted to tumor cells expressing the MiHA in a graft recipient, preferably an ASCT and/or donor lymphocyte infusion (DLI) recipient subject. Hence the invention provides CD8 T lymphocytes encoding and expressing a T cell receptor capable of interacting with the above-mentioned peptide/MHC molecule complex. Said T lymphocyte may be a recombinant or a naturally selected T lymphocyte. CD8 T lymphocytes of the invention may also be used for or in the methods and pharmaceutical compositions (see below). This specification thus provides at least two methods for producing CD8 T lymphocytes of the invention, comprising the step of bringing undifferentiated lymphocytes into contact with a peptide/MHC molecule complex (typically expressed at the surface of cells, such as APCs) under conditions conducive of triggering an immune response, which may be done *in vitro* or *in vivo* for instance in a patient receiving a graft. Alternatively, it may be carried out *in vitro* by cloning a gene encoding the TCR specific for interacting with a peptide/MHC molecule complex, which may be obtained from a cell obtained from the previous method or from a subject exhibiting an immune response against peptide/MHC molecule complex, into a host cell and/or a host lymphocyte obtained from a graft recipient or graft donor, and optionally differentiate to cytotoxic T lymphocytes (CTLs).

The potential impact of MiHA-based cancer immunotherapy is significant. For hematologic cancers (e.g., leukemia), the use of anti-MiHA T cells may replace conventional AHCT because it may provide superior anti-leukemic activity without causing GVHD. As a corollary it may benefit to many patients with hematologic malignancy who cannot be treated by conventional AHCT because their risk/reward (GVHD/GVT) ratio is too high. Finally, since studies in mice have shown that MiHA-based immunotherapy may be effective for treatment of solid tumors, MiHA-based cancer immunotherapy may be used to MiHA-targeted therapy of non-hematologic cancers, such as solid cancers.

High-avidity T cell responses capable of eradicating tumors can be generated in an allogeneic setting. In hematological malignancies, allogeneic HLA-matched hematopoietic stem cell transplantation (ASCT) provides a platform for allogeneic immunotherapy due to the induction of T cell-mediated graft-versus-tumor (GVT) immune responses. Immunotherapy in an allogeneic setting enables induction of effective T cell responses due to the fact that T cells of donor origin are not selected for low reactivity against self-antigens of the recipient. Therefore, high-affinity T cells against tumor- or recipient-specific antigens can be found in the T cell inoculum administered to the patient during or after ASCT. The main targets of the tumor-reactive T cell responses are polymorphic proteins for which donor and recipient are disparate, namely MiHAs. The MiHA peptide sequences identified herein may be used for the production of synthetic peptides to be used i) for *in vitro* priming and expansion of MiHA-specific T cells to be injected into transplant (AHCT) recipients and/or ii) as vaccines to boost the graft-vs.-tumor effect (GVTE) in recipients of MiHA-specific T cells, subsequent to the transplantation.

In another aspect, the present invention provides the use of the above-mentioned peptide of sequences (I) to (VII) in the immunotherapy of cancer.

Accordingly, in another aspect, the present invention provides a method of treating cancer, said method comprising administering to a subject in need thereof an effective amount of CD8 T lymphocytes recognizing a MHC class I molecule loaded with the above-mentioned peptide.

In another aspect, the present invention provides the use of CD8 T lymphocytes recognizing a MHC class I molecule loaded with the above-mentioned peptide for treating cancer in a subject. In another aspect, the present invention provides the use of CD8 T lymphocytes recognizing a MHC class I molecule loaded with the above-mentioned peptide for the preparation/manufacture of a medicament for treating cancer in a subject. In an embodiment, the subject is a transplant (e.g., AHCT) recipient.

In an embodiment, the above-mentioned method or use further comprises determining whether said subject expresses a CENPF nucleic acid comprising a T or a C at a position corresponding to nucleotide 4409 in the nucleic acid sequence of human CENPF (FIGs. 1A to 1C, NCBI Reference Sequence: NM_016343.3), and/or a CENPF polypeptide comprising a leucine or serine residue at a position corresponding to residue 1412 in the protein sequence of human CENPF (FIG. 1D, NCBI Reference Sequence: NP_057427.3), wherein (a) if said subject expresses a CENPF nucleic acid comprising a T at a position corresponding to nucleotide 4409 in the nucleic acid sequence of human CENPF and/or a CENPF polypeptide comprising a leucine residue at a position corresponding to residue 1412 in the protein sequence of human CENPF, X^2 is L in the peptide; (b) if said subject expresses a CENPF nucleic acid comprising a C at a position corresponding to nucleotide 4409 in the nucleic acid sequence of human CENPF

and/or a CENPF polypeptide comprising a serine residue at a position corresponding to residue 1412 in the protein sequence of human CENPF, X^2 is S in the peptide.

In an embodiment, the above-mentioned method or use further comprises determining whether said subject expresses a ZWINT nucleic acid comprising an A or a G at a position
5 corresponding to nucleotide 596 in the nucleic acid sequence of human ZWINT (FIG. 2A, NCBI Reference Sequence: NM_007057.3), and/or a ZWINT polypeptide comprising a arginine or glycine residue at a position corresponding to residue 187 in the protein sequence of human ZWINT (FIG. 2B, NCBI Reference Sequence: NP_008988.2), wherein (a) if said subject expresses a ZWINT nucleic acid comprising an A at a position corresponding to nucleotide 596
10 in the nucleic acid sequence of human ZWINT and/or a ZWINT polypeptide comprising an arginine residue at a position corresponding to residue 187 in the protein sequence of human ZWINT, X^5 is R in said peptide; (b) if said subject expresses a ZWINT nucleic acid comprising a G at a position corresponding to nucleotide 596 in the nucleic acid sequence of human ZWINT and/or a ZWINT polypeptide comprising a glycine residue at a position corresponding to residue
15 187 in the protein sequence of human ZWINT, X^5 is G in said peptide.

In an embodiment, the above-mentioned method further comprises determining whether said subject expresses a MTCH2 nucleic acid comprising a C or a G at a position corresponding to nucleotide 1057 in the nucleic acid sequence of human MTCH2 (FIG. 3A, NCBI Reference Sequence: NM_014342.3), and/or a MTCH2 polypeptide comprising a proline
20 or alanine residue at a position corresponding to residue 290 in the protein sequence of human MTCH2 (FIG. 3B, NCBI Reference Sequence: NP_055157.1), wherein (a) if said subject expresses a MTCH2 nucleic acid comprising a C at a position corresponding to nucleotide 1057 in the nucleic acid sequence of human MTCH2 and/or a MTCH2 polypeptide comprising a proline residue at a position corresponding to residue 290 in the protein sequence of human
25 MTCH2, X^9 is P in said peptide; (b) if said subject expresses a MTCH2 nucleic acid comprising a G at a position corresponding to nucleotide 1057 in the nucleic acid sequence of human MTCH2 and/or a MTCH2 polypeptide comprising an alanine residue at a position corresponding to residue 290 in the protein sequence of human MTCH2, X^5 is A in said peptide.

In an embodiment, the above-mentioned method further comprises determining
30 whether said subject expresses an ELF1 nucleic acid comprising an A or T at a position corresponding to nucleotide 1400 in the nucleic acid sequence of human ELF1 (FIGs. 4A and 4B, NCBI Reference Sequence: NM_172373.3) and/or an ELF1 polypeptide having a threonine or a serine at a position corresponding to residue 343 in the ELF1 protein sequence (FIG. 4C, NCBI Reference Sequence: NP_758961.1), wherein (a) if said subject expresses an ELF1
35 nucleic acid comprising an A at a position corresponding to nucleotide 1400 in the nucleic acid sequence of human ELF1 and/or an ELF1 polypeptide comprising a threonine residue at a position corresponding to residue 343 in the protein sequence of human ELF1, X^{12} is T in said

peptide; (b) if said subject expresses an ELF1 nucleic acid comprising a T at a position corresponding to nucleotide 1400 in the nucleic acid sequence of human ELF1 and/or an ELF1 polypeptide comprising a serine residue at a position corresponding to residue 343 in the protein sequence of human ELF1, X^{12} is S in said peptide.

5 In an embodiment, the above-mentioned method further comprises determining whether said subject expresses an NQO1 nucleic acid comprising an C or T at a position corresponding to nucleotide 615 in the nucleic acid sequence of human NQO1 (FIG. 5A, NCBI Reference Sequence: NM_000903.2) and/or an NQO1 polypeptide having an arginine or a tryptophan at a position corresponding to residue 139 in the NQO1 protein sequence (FIG. 5B,
10 NCBI Reference Sequence: NP_000894.1), wherein (a) if said subject expresses a NQO1 nucleic acid comprising an C at a position corresponding to nucleotide 615 in the nucleic acid sequence of human NQO1 and/or a NQO1 polypeptide comprising an arginine residue at a position corresponding to residue 139 in the protein sequence of human NQO1, X^{17} is R in said peptide; (b) if said subject expresses an NQO1 nucleic acid comprising a T at a position
15 corresponding to nucleotide 615 in the nucleic acid sequence of human NQO1 and/or a NQO1 polypeptide comprising a tryptophan residue at a position corresponding to residue 139 in the protein sequence of human NQO1, X^{17} is W in said peptide.

In an embodiment, the above-mentioned method further comprises determining whether said subject expresses a KIAA0226L nucleic acid comprising a G or A at a position
20 corresponding to nucleotide 1059 in the nucleic acid sequence of human KIAA0226L (FIGs. 6A and 6B, NCBI Reference Sequence: NM_025113.2) and/or an KIAA0226L polypeptide having a glycine or an arginine at a position corresponding to residue 152 in the KIAA0226L protein sequence (FIG. 6C, NCBI Reference Sequence: NP_079389.2), wherein (a) if said subject expresses a KIAA0226L nucleic acid comprising a G at a position corresponding to nucleotide
25 1059 in the nucleic acid sequence of human KIAA0226L and/or a KIAA0226L polypeptide comprising a glycine residue at a position corresponding to residue 152 in the protein sequence of human KIAA0226L, X^{23} is G in said peptide; (b) if said subject expresses a KIAA0226L nucleic acid comprising an A at a position corresponding to nucleotide 1059 in the nucleic acid sequence of human a KIAA0226L and/or a KIAA0226L polypeptide comprising an arginine
30 residue at a position corresponding to residue 152 in the protein sequence of human KIAA0226L, X^{23} is R in said peptide.

In an embodiment, the above-mentioned method further comprises determining whether said subject expresses an RMDN1 nucleic acid comprising an A or C at a position corresponding to nucleotide 316 in the nucleic acid sequence of human RMDN1 (FIG. 7A, NCBI
35 Reference Sequence: NM_016033.2) and/or an RMDN1 polypeptide having a lysine or an asparagine at a position corresponding to residue 52 in the RMDN1 protein sequence (FIG. 7B, NCBI Reference Sequence: NP_057117.2), wherein (a) if said subject expresses an RMDN1

nucleic acid comprising an A at a position corresponding to nucleotide 316 in the nucleic acid sequence of human RMDN1 and/or an RMDN1 polypeptide comprising a lysine residue at a position corresponding to residue 52 in the protein sequence of human RMDN1, X²⁸ is K in said peptide; (b) if said subject expresses an RMDN1 nucleic acid comprising a C at a position
 5 corresponding to nucleotide 316 in the nucleic acid sequence of human RMDN1 and/or a RMDN1 polypeptide comprising an asparagine residue at a position corresponding to residue 52 in the protein sequence of human RMDN1, X²⁸ is N in said peptide.

The above-noted polymorphism (nucleotide variation) in the nucleic acid and/or protein of interest (e.g., CENPF, ZWINT, MTCH2, ELF1, NQO1, KIAA0226L or RMDN1) may be
 10 detected by a number of methods which are known in the art. Examples of suitable methods for detecting alterations at the nucleic acid level include sequencing of the nucleic acid sequence of the nucleic acid of interest (e.g., CENPF, ZWINT, MTCH2, ELF1, NQO1, KIAA0226L or RMDN1); hybridization of a nucleic acid probe capable of specifically hybridizing to a nucleic acid of interest (e.g., CENPF, ZWINT, MTCH2, ELF1, NQO1, KIAA0226L or RMDN1)
 15 comprising the polymorphism (the first allele) and not to (or to a lesser extent to) a corresponding nucleic acid that do not comprise the polymorphism (the second allele) (under comparable hybridization conditions, such as stringent hybridization conditions), or vice-versa; restriction fragment length polymorphism analysis (RFLP); Amplified fragment length polymorphism PCR (AFLP-PCR); amplification of a nucleic acid fragment using a primer specific
 20 for one of the allele, wherein the primer produces an amplified product if the allele is present and does not produce the same amplified product when the other allele is used as a template for amplification (e.g., allele-specific PCR). Other methods include *in situ* hybridization analyses and single-stranded conformational polymorphism analyses. Further, nucleic acids of interest (e.g., CENPF, ZWINT, MTCH2, ELF1, NQO1, KIAA0226L or RMDN1 nucleic acids) may be
 25 amplified using known methods (e.g., polymerase chain reaction [PCR]) prior to or in conjunction with the detection methods noted herein. The design of various primers for such amplification is known in the art. The nucleic acid (mRNA) may also be reverse transcribed into cDNA prior to analysis.

Examples of suitable methods for detecting alterations/polymorphisms at the
 30 polypeptide level include sequencing of the polypeptide of interest (e.g., CENPF, ZWINT, MTCH2, ELF1, NQO1, KIAA0226L or RMDN1); digestion of the polypeptide followed by mass spectrometry or HPLC analysis of the peptide fragments, wherein the alteration/polymorphism of the polypeptide of interest (e.g., CENPF, ZWINT, MTCH2, ELF1, NQO1, KIAA0226L or RMDN1) results in an altered mass spectrometry or HPLC spectrum as compared to the native
 35 polypeptide of interest (e.g., CENPF, ZWINT, MTCH2, ELF1, NQO1, KIAA0226L or RMDN1); and immunodetection using an immunological reagent (e.g., an antibody, a ligand) which exhibits altered immunoreactivity with a polypeptide comprising the alteration (first allele)

relative to a corresponding native polypeptide not comprising the alteration (second allele), for example by targeting an epitope comprising the amino acid change. Immunodetection can measure the amount of binding between a polypeptide molecule and an anti-protein antibody by the use of enzymatic, chromodynamic, radioactive, magnetic, or luminescent labels which are attached to either the anti-protein antibody or a secondary antibody which binds the anti-protein antibody. In addition, other high affinity ligands may be used. Immunoassays which can be used include e.g. ELISAs, Western blots, and other techniques known to those of ordinary skill in the art (see Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1999 and Edwards R, *Immunodiagnostics: A Practical Approach*, Oxford University Press, Oxford; England, 1999).

All these detection techniques may also be employed in the format of microarrays, protein-arrays, antibody microarrays, tissue microarrays, electronic biochip or protein-chip based technologies (see Schena M., *Microarray Biochip Technology*, Eaton Publishing, Natick, Mass., 2000).

Further, nucleic acids of interest (e.g., CENPF, ZWINT, MTCH2, ELF1, NQO1, KIAA0226L or RMDN1 nucleic acids) may be amplified using known methods (e.g., polymerase chain reaction [PCR]) prior to or in conjunction with the detection methods noted herein. The design of various primers for such amplification is known in the art.

In an embodiment, the above determining comprises sequencing in a biological sample from a subject a region of a nucleic acid corresponding to the region encompassing (i) nucleotide 4409 of a human CENPF nucleic acid (SEQ ID NO:1), (ii) nucleotide 596 in the nucleic acid of a human ZWINT nucleic acid (SEQ ID NO:3), (iii) nucleotide 1057 in the nucleic acid of a human MTCH2 nucleic acid (SEQ ID NO:5), (iv) nucleotide 1400 in the nucleic acid of a human ELF1 nucleic acid (SEQ ID NO:7), (v) nucleotide 615 in the nucleic acid of a human NQO1 nucleic acid (SEQ ID NO:9), (vi) nucleotide 1059 in the nucleic acid of a human KIAA0226L nucleic acid (SEQ ID NO:11) and/or (vii) nucleotide 316 in the nucleic acid of a human RMDN1 nucleic acid (SEQ ID NO:13).

In an embodiment, the above-mentioned CD8 T lymphocytes are *in vitro* expanded CD8 T lymphocytes. Expanded CD8 T lymphocytes may be obtained by culturing primary CD8 T lymphocytes (from a donor) under conditions permitting the proliferation and/or differentiation of the CD8 T lymphocytes. Such conditions typically include contacting the CD8 T lymphocytes with cells, such as APCs, expressing at their surface peptide/MHC complexes, in the presence of growth factors and/or cytokines such as IL-2, IL-7 and/or IL-15 (see, e.g., Montes *et al.*, *Clin Exp Immunol.* 2005 Nov;142(2):292-302). Such expanded CD8 T lymphocytes are then administered to the recipient, for example through intravenous infusion.

In an embodiment, the subject is an allogeneic stem cell transplantation (ASCT) or donor lymphocyte infusion (DLI) recipient.

In another aspect, the present invention provides a method of expanding CD8 T lymphocytes for adoptive T-cell immunotherapy, said method comprising (a) determining whether a candidate donor expresses a CENPF nucleic acid comprising a T or a C at a position corresponding to nucleotide 4409 in the nucleic acid sequence of human CENPF (FIGs. 1A to 1D, NCBI Reference Sequence: NM_016343.3), and/or a CENPF polypeptide comprising a leucine or serine residue at a position corresponding to residue 1412 in the protein sequence of human CENPF (FIG. 1E, NCBI Reference Sequence: NP_057427.3) and (b)(i) if said candidate donor expresses a CENPF nucleic acid comprising a T at a position corresponding to nucleotide 4409 in the nucleic acid sequence of human CENPF and/or a CENPF polypeptide comprising a leucine residue at a position corresponding to residue 1412 in the protein sequence of human CENPF, culturing CD8 T lymphocytes from said candidate donor in the presence of cells expressing a MHC class I molecule of the HLA-B*0801 allele loaded with the above-mentioned peptide, wherein X^2 is S in said peptide, under conditions suitable for CD8 T lymphocyte expansion; or (b)(ii) if said candidate donor expresses a CENPF nucleic acid comprising a C at a position corresponding to nucleotide 4409 in the nucleic acid sequence of human CENPF and/or a CENPF polypeptide comprising a serine residue at a position corresponding to residue 1412 in the protein sequence of human CENPF, culturing CD8 T lymphocytes from said candidate donor in the presence of cells expressing a MHC class I molecule of the HLA-B*0801 allele loaded with the above-mentioned peptide, wherein X^2 is L in said peptide, under conditions suitable for CD8 T lymphocyte expansion.

In another aspect, the present invention provides a method of expanding CD8 T lymphocytes for adoptive T-cell immunotherapy, said method comprising: (a) determining whether a candidate donor expresses a ZWINT nucleic acid comprising an A or a G at a position corresponding to nucleotide 596 in the nucleic acid sequence of human ZWINT (FIG. 2A, NCBI Reference Sequence: NM_007057.3), and/or a ZWINT polypeptide comprising an arginine or glycine residue at a position corresponding to residue 187 in the protein sequence of human ZWINT (FIG. 2B, NCBI Reference Sequence: NP_008988.2) and (b)(i) if said candidate donor expresses a ZWINT nucleic acid comprising an A at a position corresponding to nucleotide 596 in the nucleic acid sequence of human ZWINT and/or a ZWINT polypeptide comprising an arginine residue at a position corresponding to residue 187 in the protein sequence of human ZWINT, culturing CD8 T lymphocytes from said candidate donor in the presence of cells expressing a MHC class I molecule of the HLA-B*4403 allele loaded with the above-mentioned peptide (II), wherein X^5 is G in said peptide, under conditions suitable for CD8 T lymphocyte expansion; or (b)(ii) if said candidate donor expresses a ZWINT nucleic acid comprising a G at a position corresponding to nucleotide 596 in the nucleic acid sequence of human ZWINT and/or a ZWINT polypeptide comprising a glycine residue at a position corresponding to residue 187 in the protein sequence of human ZWINT, culturing CD8 T

lymphocytes from said candidate donor in the presence of cells expressing a MHC class I molecule of the HLA-B*4403 allele loaded with the above-mentioned peptide (II), wherein X⁵ is R in said peptide, under conditions suitable for CD8 T lymphocyte expansion.

In another aspect, the present invention provides a method of expanding CD8 T lymphocytes for adoptive T-cell immunotherapy, said method comprising: (a) determining whether a candidate donor expresses a MTCH2 nucleic acid comprising a C or a G at a position corresponding to nucleotide 1057 in the nucleic acid sequence of human MTCH2 (FIG. 3A, NCBI Reference Sequence: NM_014342.3), and/or a MTCH2 polypeptide comprising a proline or alanine residue at a position corresponding to residue 290 in the protein sequence of human MTCH2 (FIG. 3B, NCBI Reference Sequence: NP_055157.1) and (b)(i) if said candidate donor expresses a MTCH2 nucleic acid comprising a C at a position corresponding to nucleotide 1057 in the nucleic acid sequence of human MTCH2 and/or a MTCH2 polypeptide comprising a proline residue at a position corresponding to residue 290 in the protein sequence of human MTCH2, culturing CD8 T lymphocytes from said candidate donor in the presence of cells expressing a MHC class I molecule of the HLA-B*0801 allele loaded with the peptide of any one of claims 64 to 75, wherein X⁹ is A in said peptide, under conditions suitable for CD8 T lymphocyte expansion; or (b)(ii) if said candidate donor expresses a MTCH2 nucleic acid comprising a G at a position corresponding to nucleotide 1057 in the nucleic acid sequence of human MTCH2 and/or a MTCH2 polypeptide comprising an alanine residue at a position corresponding to residue 290 in the protein sequence of human MTCH2, culturing CD8 T lymphocytes from said candidate donor in the presence of cells expressing a MHC class I molecule of the HLA-B*0801 allele loaded with the peptide of any one of claims 64 to 75, wherein X⁹ is P in said peptide, under conditions suitable for CD8 T lymphocyte expansion.

In another aspect, the present invention provides a method of expanding CD8 T lymphocytes for adoptive T-cell immunotherapy, said method comprising: (a) determining whether a candidate donor expresses an ELF1 nucleic acid comprising an A or T at a position corresponding to nucleotide 1400 in the nucleic acid sequence of human ELF1 (FIGs. 4A and 4B, NCBI Reference Sequence: NM_172373.3) and/or an ELF1 polypeptide having a threonine or a serine at a position corresponding to residue 343 in the ELF1 protein sequence (FIG. 4C, NCBI Reference Sequence: NP_758961.1) and (b)(i) if said candidate donor expresses an ELF1 nucleic acid comprising an A at a position corresponding to nucleotide 1400 in the nucleic acid sequence of human ELF1 and/or an ELF1 polypeptide having a threonine at a position corresponding to residue 343 in the ELF1 protein sequence, culturing CD8 T lymphocytes from said candidate donor in the presence of cells expressing a MHC class I molecule of the HLA-A*0301 allele loaded with the above-mentioned peptide (IV), wherein X¹² is S in said peptide, under conditions suitable for CD8 T lymphocyte expansion; or (b)(ii) if said candidate donor expresses an ELF1 nucleic acid comprising a T at a position corresponding to nucleotide 1400

in the nucleic acid sequence of human ELF1 and/or an ELF1 polypeptide having a serine at a position corresponding to residue 343 in the ELF1 protein sequence, culturing CD8 T lymphocytes from said candidate donor in the presence of cells expressing a MHC class I molecule of the HLA-A*0301 allele loaded with the above-mentioned peptide (IV), wherein X¹² is T in said peptide, under conditions suitable for CD8 T lymphocyte expansion.

In another aspect, the present invention provides a method of expanding CD8 T lymphocytes for adoptive T-cell immunotherapy, said method comprising: (a) determining whether said subject expresses an NQO1 nucleic acid comprising a C or T at a position corresponding to nucleotide 615 in the nucleic acid sequence of human NQO1 (FIG. 5A, NCBI Reference Sequence: NM_000903.2) and/or an NQO1 polypeptide having an arginine or a tryptophan at a position corresponding to residue 139 in the NQO1 protein sequence (FIG. 5B, NCBI Reference Sequence: NP_000894.1) and (b)(i) if said candidate donor expresses a NQO1 nucleic acid comprising a C at a position corresponding to nucleotide 615 in the nucleic acid sequence of human NQO1 and/or a NQO1 polypeptide comprising an arginine residue at a position corresponding to residue 139 in the protein sequence of human NQO1, culturing CD8 T lymphocytes from said candidate donor in the presence of cells expressing a MHC class I molecule of the HLA-A*0301 allele loaded with the above-mentioned peptide (V), wherein X¹⁷ is W in said peptide, under conditions suitable for CD8 T lymphocyte expansion; or (b)(ii) if said candidate donor expresses a NQO1 nucleic acid comprising a T at a position corresponding to nucleotide 615 in the nucleic acid sequence of human NQO1 and/or a NQO1 polypeptide comprising a tryptophan residue at a position corresponding to residue 139 in the protein sequence of human NQO1, culturing CD8 T lymphocytes from said candidate donor in the presence of cells expressing a MHC class I molecule of the HLA-A*0301 allele loaded with the above-mentioned peptide (V), wherein X¹⁷ is R in said peptide, under conditions suitable for CD8 T lymphocyte expansion.

In another aspect, the present invention provides a method of expanding CD8 T lymphocytes for adoptive T-cell immunotherapy, said method comprising: (a) determining whether said subject expresses a KIAA0226L nucleic acid comprising a G or A at a position corresponding to nucleotide 1059 in the nucleic acid sequence of human KIAA0226L (FIGs. 6A and 6B, NCBI Reference Sequence: NM_025113.2) and/or a KIAA0226L polypeptide having a glycine or an arginine at a position corresponding to residue 152 in the KIAA0226L protein sequence (FIG. 6C, NCBI Reference Sequence: NP_079389.2) and (b)(i) if said candidate donor expresses a KIAA0226L nucleic acid comprising a G at a position corresponding to nucleotide 1059 in the nucleic acid sequence of human KIAA0226L and/or a KIAA0226L polypeptide comprising a glycine residue at a position corresponding to residue 152 in the protein sequence of human KIAA0226L, culturing CD8 T lymphocytes from said candidate donor in the presence of cells expressing a MHC class I molecule of the HLA-A*0301 allele

loaded with the above-mentioned peptide (VI), wherein X^{23} is R in said peptide, under conditions suitable for CD8 T lymphocyte expansion; or (b)(ii) if said candidate donor expresses a KIAA0226L nucleic acid comprising an A at a position corresponding to nucleotide 1059 in the nucleic acid sequence of human KIAA0226L and/or a KIAA0226L polypeptide comprising an arginine residue at a position corresponding to residue 152 in the protein sequence of human KIAA0226L, culturing CD8 T lymphocytes from said candidate donor in the presence of cells expressing a MHC class I molecule of the HLA-A*0301 allele loaded with the above-mentioned peptide (VI), wherein X^{23} is G in said peptide, under conditions suitable for CD8 T lymphocyte expansion.

In another aspect, the present invention provides a method of expanding CD8 T lymphocytes for adoptive T-cell immunotherapy, said method comprising: (a) determining whether said subject expresses an RMDN1 nucleic acid comprising an A or C at a position corresponding to nucleotide 316 in the nucleic acid sequence of human RMDN1 (FIG. 7A, NCBI Reference Sequence: NM_016033.2) and/or an RMDN1 polypeptide having a lysine or an asparagine at a position corresponding to residue 52 in the RMDN1 protein sequence (FIG. 7B, NCBI Reference Sequence: NP_057117.2) and (b)(i) if said candidate donor expresses a RMDN1 nucleic acid comprising a A at a position corresponding to nucleotide 316 in the nucleic acid sequence of human RMDN1 and/or a RMDN1 polypeptide comprising a lysine residue at a position corresponding to residue 52 in the protein sequence of human RMDN1, culturing CD8 T lymphocytes from said candidate donor in the presence of cells expressing a MHC class I molecule of the HLA-A*0301 allele loaded with the above-mentioned peptide (VII), wherein X^{28} is N in said peptide, under conditions suitable for CD8 T lymphocyte expansion; or (b)(ii) if said candidate donor expresses a RMDN1 nucleic acid comprising a C at a position corresponding to nucleotide 316 in the nucleic acid sequence of human RMDN1 and/or a RMDN1 polypeptide comprising an arginine residue at a position corresponding to residue 52 in the protein sequence of human RMDN1, culturing CD8 T lymphocytes from said candidate donor in the presence of cells expressing a MHC class I molecule of the HLA-A*0301 allele loaded with the above-mentioned peptide (VII), wherein X^{28} is K in said peptide, under conditions suitable for CD8 T lymphocyte expansion.

In an embodiment, the above-mentioned cancer comprises tumor cells expressing CENPF, ZWINT, MTCH2, ELF1, NQO1, KIAA0226L and/or RMDN1.

In an embodiment, the above-mentioned cancer comprises tumor cells expressing CENPF, and is a hematopoietic cancer, such as leukemia, lymphoma, or myeloma, or a solid tumor such as head and neck squamous cell carcinomas, breast cancer, non-Hodgkin's lymphoma or gastrointestinal cancer.

In another embodiment, the above-mentioned cancer comprises tumor cells expressing ZWINT, and is a breast, prostate or bladder cancer.

In another embodiment, the above-mentioned cancer comprises tumor cells expressing MTCH2, and is a solid tumor/cancer, in a further embodiment lung, thyroid, liver, esophagus, colon or breast cancer, or osteosarcoma.

In another embodiment, the above-mentioned cancer comprises tumor cells expressing
5 ELF1, and is leukemia, lung cancer (e.g., non-small cell lung cancer), breast cancer or ovarian cancer.

In another embodiment, the above-mentioned cancer comprises tumor cells expressing NQO1, and is lung cancer (e.g., non-small cell lung cancer), skin cancer, breast cancer, liver cancer (e.g., intrahepatic cholangiocarcinoma) digestive tract cancer such as colorectal cancer,
10 or pancreatic cancer.

In another embodiment, the above-mentioned cancer comprises tumor cells expressing KIAA0226L, and is a lymphoma (e.g., Burkitt's lymphoma), skin cancer, breast cancer, liver cancer (e.g., intrahepatic cholangiocarcinoma) digestive tract cancer such as colorectal cancer, or pancreatic cancer.
15

Although the present invention has been described hereinabove by way of specific embodiments thereof, it can be modified, without departing from the spirit and nature of the subject invention as defined in the appended claims. In the claims, the word "comprising" is used as an open-ended term, substantially equivalent to the phrase "including, but not limited to". The singular forms "a", "an" and "the" include corresponding plural references unless the
20 context clearly dictates otherwise.

MODE(S) FOR CARRYING OUT THE INVENTION

The present invention is illustrated in further details by the following non-limiting examples.

25 Example 1: Materials and Methods

Cell culture. Peripheral blood mononuclear cells (PBMCs) were isolated from blood samples of 2 HLA-identical siblings. Epstein-Barr virus (EBV)-transformed B lymphoblastoid cell lines (B-LCLs) were derived from PBMCs with Ficoll-Paque™ Plus (Amersham) as described (Tosato and Cohen, 2007). Established B-LCLs were maintained in RPMI 1640 medium
30 supplemented with 10% fetal bovine serum, 25 mM of HEPES, 2 mM L-glutamine and penicillin-streptomycin (all from Invitrogen).

HLA typing. High-resolution HLA genotyping was performed at the Maisonneuve-Rosemont Hospital. The HLA genotype of our subjects was HLA-A*0301/2902, HLA-B*0801/*4403, HLA-C*0701/*1601 and HLA-DRB1*0301/*0701.

35 *RNA extraction.* Total RNA was isolated from 5 million B-LCLs using RNeasy™ mini kit including DNase I treatment (Qiagen) according to the manufacturer's instructions. Total RNA

was quantified using the NanoDrop™ 2000 (Thermo Scientific) and RNA quality was assessed with the 2100 Bioanalyzer™ (Agilent Technologies).

Preparation of transcriptome libraries. Transcriptome libraries were generated from 1 µg of total RNA using the TruSeq™ RNA Sample Prep Kit (v2) (RS-930–1021, Illumina) following the manufacturer's protocol. Briefly, poly-A mRNA was purified using poly-T oligo-attached magnetic beads using two rounds of purification. During the second elution of the poly-A RNA, the RNA was fragmented and primed for cDNA synthesis. Reverse transcription (RT) of the first strand was done using random primers and SuperScript™ II (InvitroGene). A second round of RT was also done to generate a double-stranded cDNA, which was then purified using Agencourt AMPure™ XP PCR purification system (Beckman Coulter). End repair of fragmented cDNA, adenylation of the 3' ends and ligation of adaptors were done following the manufacturer's protocol. Enrichment of DNA fragments containing adapter molecules on both ends was done using 15 cycles of PCR amplification and the Illumina™ PCR mix and primers cocktail.

DNA extraction. Genomic DNA was extracted from 5 million B-LCLs using the PureLink™ Genomic DNA Mini Kit (Invitrogen) according to the manufacturer's instructions. DNA was quantified and quality-assessed using the NanoDrop™ 2000 (Thermo Scientific).

Preparation of genomic DNA libraries and exome enrichment. Genomic libraries were constructed from 1µg of genomic DNA using the TruSeq™ DNA Sample Preparation Kit (v2) (FC-930–1021, Illumina) following the manufacturer's protocol. This included the following steps: DNA shearing using a Covaris™ S2 instrument, DNA end repair, 3' ends adenylation, ligation of adaptors, purification of ligation products and PCR amplification to enrich DNA fragments that have adapter molecules.

DNA-Seq libraries were subjected to an enrichment step to sequence only the coding portion (exome) of the human genome. 500 ng of DNA-Seq libraries were used for hybrid selection-based exome enrichment with the TruSeq™ exome enrichment kit (FC-930-1012, Illumina) according to the manufacturer's instructions.

Whole transcriptome sequencing (RNA-Seq) and exome sequencing. Paired-end (2 x 100 bp) sequencing was performed using the Illumina HiSeq2000™ machine running TruSeq™ v3 chemistry. Cluster density was targeted at around 600-800k clusters/mm². Two RNA-Seq or four exomes libraries were sequenced per lane (8 lanes per slide). Detail of the Illumina sequencing technologies can be found at http://www.illumina.com/applications/detail/sequencing/dna_sequencing.ilmn. Briefly, DNA or RNA libraries are incorporated into a fluidic flow cell design with 8 individual lanes. The flow cell surface is populated with capture oligonucleotide anchors, which hybridize the appropriately modified DNA segments of a sequencing library. By a process called "bridge amplification," captured DNA templates are amplified in the flow cell by "arching" over and hybridizing to an

adjacent anchor oligonucleotide primer. The sequencing reaction is performed by hybridizing a primer complementary to the adapter sequence and then cyclically adding DNA polymerase and a mixture of 4 differently colored fluorescent reversible dye terminators to the captured DNA in the flow cell. By using this approach, unmodified DNA fragments and unincorporated nucleotides are washed away, while captured DNA fragments are extended one nucleotide at a time. After each nucleotide-coupling cycle takes place, the flow cell is scanned, and digital images are acquired to record the locations of fluorescently labeled nucleotide incorporations. Following imaging, the fluorescent dye and the terminal 3' blocker are chemically removed from the DNA before the next nucleotide coupling cycle.

Read mapping. Sequence data were mapped to the human reference genome (hg19) using the Illumina Casava™ 1.8.1 and the Eland™ v2 mapping softwares. First, the *.bcl files were converted into compressed FASTQ files, following by demultiplexing of separate multiplexed sequence runs by index. Then, single reads were aligned to the human reference genome using the multiseed and gapped alignment method. Multiseed alignment works by aligning the first seed of 32 bases and consecutive seeds separately. Gapped alignment extends each candidate alignment to the full length of the read and allows for gaps up to 10 bases. The following criteria were applied: i) a read has to have at least one seed that matches with at most 2 mismatches, and for that seed no gaps are allowed and ii) for the whole read any number of gaps were allowed, as long as they correct at least five mismatches downstream. For each candidate alignment a probability score was calculated. This score is based on the sequencing base quality values and the positions of the mismatches. The alignment score of a read, which is expressed on the Phred scale, was computed from the probability scores of the candidate alignments. The best alignment for a given read corresponded to the candidate alignment with the highest probability score and it was picked if the alignment score exceeded a threshold. Reads that mapped at 2 or more locations (multireads) were not included in further analyses. An additional alignment was done against splice junctions and contaminants (mitochondrial and ribosomal RNA).

Identification of single nucleotide variations. The first step of the process consists of retrieving the list of all single nucleotide variations (SNV) observed between the reference genome (GRCh37.p2, NCBI) and the sequenced transcriptome of each of our subjects. This was done by using the SNP calling program Casava™ v1.8.2 from Illumina (http://support.illumina.com/sequencing/sequencing_software/casava.ilmn). Casava calculates and retrieves statistics about every observed SNV including its position, the reference base, the raw counts for each base, the most probable genotype (max_gt) and the probability of the most probable genotype (Qmax_gt). Among the SNVs called by Casava™, only those that had a high confidence (Qmax_gt value>20) were considered. SNVs with Qmax_gt value below this

threshold were assigned with the reference base instead. This strategy was used to identify SNVs at the transcript level between each of our subjects and the reference genome.

In silico translated transcriptome. The sequences containing the identified SNVs of each individual were further processed. For each sequence, all transcripts reported in Ensembl
5 (<http://useast.ensembl.org/info/data/ftp/index.html>, Flicek et al., Ensembl 2012, *Nucleic Acids Research* 2012 40 Database issue:D84-D90) were retrieved and *in silico* translated each of them into proteins using our in-house software pyGeno (Granados et al., 2012). The *in silico* translated transcriptomes included cases in which more than one non-synonymous allele was found for a given position. Supposing that a polymorphic position could
10 affect 11mers or smaller peptides upstream or downstream, a window of 66 bp around each one of these polymorphic positions was considered and computed every possible amino acid (aa) sequence variant defined by these 66 bp (or 22 aa) windows. In this way, a list of most possible aa sequences of at most 11 aa affected by non-synonymous polymorphisms was obtained. The number of aa sequences affected by one non-synonymous polymorphism was
15 limited to 10240 to limit the size of the file. All translated sequences were compiled into a single FASTA file that was used as a database for the identification of MHC class I-associated peptides (see "MS/MS sequencing and peptide clustering" section).

Mass spectrometry and peptide sequencing. Three biological replicates of 4×10^8 exponentially growing B-LCLs were prepared from each subject. MHC class I-associated
20 peptides were released by mild acid treatment, pretreated by desalting with an HLB cartridge, filtered with a 3000 Da cut-off column and separated by cation exchange chromatography (SCX) using an off-line 1100 series binary LC system (Agilent Technologies) as previously described (Fortier et al., 2008). Peptides were loaded at 8 μ L/min on a homemade strong cation exchange (SCX) column (0.3 mm internal diameter $\times$ 50 mm length) packed with SCX bulk
25 material (Polysulfoethyl ATM, PolyLC). Peptides were separated into five fractions using a linear gradient of 0–25% B in 25 min (solvent A: 5 mM ammonium formate, 15% acetonitrile, ACN, pH 3.0; solvent B: 2 M ammonium formate, 15% ACN, pH 3.0) and brought to dryness using a Speedvac.

MHC class I-associated peptide fractions were resuspended in 2% aqueous ACN
30 (0.2% formic acid) and analyzed by LC-MS/MS using an EksigentTM LC system coupled to a LTQ-OrbitrapTM mass spectrometer (Thermo Electron) (Fortier et al., 2008; de Verteuil et al., 2010; Caron et al., 2011). Peptides were separated in a custom C₁₈ reversed phase column (150 μ m i.d. $\times$ 100 mm, Jupiter ProteoTM 4 μ m, Phenomenex) at a flow rate of 600 nL/min using a linear gradient of 3-60% aqueous ACN (0.2% formic acid) in 69 mins. Full mass spectra were
35 acquired with the OrbitrapTM analyzer operated at a resolving power of 60 000 (at m/z 400) and collision-activated dissociation tandem mass spectra were acquired in data-dependent mode with the linear ion trap analyzer. Mass calibration used an internal lock mass (protonated

(Si(CH₃)₂O)₆; *m/z* 445.120029) and mass accuracy of peptide measurements was within 5 ppm.

MS/MS sequencing and peptide clustering. Mass spectra were analyzed using Xcalibur™ software and peak lists were generated using Mascot™ distiller version 2.1.1 (Matrix Science, <http://www.matrixscience.com>). Database searches were performed against a non-redundant human Uniprot database (containing 110,361 sequences, released on 28.07.2011) (version 101 of the UniProt Gene Ontology Annotation available from the European Bioinformatics Institute (EBI), an academic research institute located on the Wellcome Trust Genome Campus in Hinxton near Cambridge (UK), part of the European Molecular Biology Laboratory; <http://www.ebi.ac.uk/Information/>; Magrane *et al.*, Database Vol. 2011, Article ID bar009; The UniProt Consortium, *Nucleic Acids Res.* 2011 January; 39(Database issue): D214-D219) and databases specific for each individual (see “*in silico* translated transcriptome” section) using Mascot (version 2.3, Matrix Science). A Mascot search against a concatenated target/decoy database consisting of combined forward and reverse versions of the Ensembl human reference genome database and of each subject-specific database. Non-redundant peptide sequences with a cutoff score threshold above 15 were selected. The tolerances for precursor and fragment mass values were set to 0.02 and 0.05 Dalton, respectively. Searches were performed without enzyme specificity and a variable modification for oxidation (Met) and deamidation (Asn, Gln). Raw data files were converted to peptide maps comprising *m/z* values, charge state, retention time and intensity for all detected ions above a threshold of 8000 counts using in-house software (Proteoprofile) (Fortier *et al.*, 2008; de Verteuil *et al.*, 2010; Caron *et al.*, 2011). Peptide maps were aligned together and peptide ions of each map (including non-sequenced ions) were aggregated to their corresponding Mascot identification, creating a peptide abundance profile. When multiple Mascot identifications were associated to the same ions, only the one with the highest score was kept. The intensity counts were then summed for identical peptide sequences resulting in a non-redundant abundance profile of identified peptide sequences.

Selection of MiHA candidates and identification of proteins source of MiHAs. Peptides were filtered by their length and those peptides with the canonical MHC I-associated peptide length (typically 8-11 mers) were kept. Peptides were considered to be undetected/not expressed if they were absent in 3 replicates per subject, and detected/expressed if they were identified in at least 2 replicates per subject. The predicted binding affinity (IC₅₀) of peptides to the allelic products was obtained using NetMHCcons version 1.0 (<http://www.cbs.dtu.dk/services/NetMHCcons/>) (Karosiene *et al.*, 2011) and was used to classify MHC I peptides. Peptides with an IC₅₀ below 50 nM were considered as strong binders and peptides with an IC₅₀ between 50 and 500 nM were considered as weak binders.

MiHA peptides were selected according to the following criteria:

i) Presence of a non-synonymous SNV between the 2 subjects in the peptide-coding region leading to the exclusive surface expression of the corresponding peptide(s) in one of the 2 subjects. These constitute MiHA differences between the subjects.

5 ii) Presence of a reported non-synonymous SNP in the peptide-coding region of the subjects leading to surface expression of the corresponding peptide(s). These constitute MiHA differences between the subjects and other individuals harboring the alternate allele for the reported SNP.

The RNA (cDNA) and DNA sequences encoding MiHAs candidates were manually inspected using the Integrative Genomics Viewer v2.0 (The Broad Institute). The UCSC Repeat
10 Masker track was included to discard candidates that corresponded to repetitive regions. A dbSNP (build 135) track was also used to identify those SNVs in the MiHA-encoding region that corresponded to reported SNPs. MiHA candidates were further inspected for mass accuracy and MS/MS spectra were validated manually using the Xcalibur™ software (Thermo Xcalibur 2.2 SP1.48 version).

15

Example 2: Novel high-throughput method for discovery of human MiHAs

Since MiHAs are i) peptides ii) whose presence depends on genetic polymorphisms, it was reasoned that high-throughput MiHA discovery would involve a combination of MS and personalized whole transcriptome and/or exome sequencing. Genomic data alone are
20 insufficient for MiHA discovery because even if a genetic polymorphism is identified in a gene, i) only 0.1% of expressed peptides are presented as MAPs at the cell surface and ii) the effect of polymorphisms in trans cannot be predicted (5'6). In MS analyses, peptides are identified through database search with softwares such as Mascot. These databases contain the translated genome of reference and do not include genetic polymorphisms. Since MiHAs are
25 the result of genetic polymorphisms, they cannot be discovered by standard MS alone.

A novel method for human MiHA discovery was developed. One of the key elements of the method is the inclusion of personalized translated transcriptome in the database used for peptide identification by MS.

PBMCs were isolated from blood samples of 2 HLA-identical siblings who are HLA-
30 A*0301/2902, HLA-B*0801/*4403, HLA-C*0701/*1601. Epstein-Barr virus (EBV)-transformed B lymphoblastoid cell lines (B-LCLs) were derived from PBMCs with Ficoll-Paque™ Plus. Three biological replicates of 4×10^8 exponentially growing B-LCLs were prepared from each subject. MAPs were released by mild acid treatment and separated by cation exchange chromatography. MAP fractions were analyzed by LC-MS/MS.

35 On each subject, a whole-exome and transcriptome sequencing was performed, and variants were identified. Based on the list of variants for each subject and human transcript annotations, a complete repertoire of subject-specific protein sequences was produced by

translating annotated transcripts. While whole-exome sequencing provided a comprehensive identification of polymorphisms in coding regions, RNA-seq provided a most useful complementary dataset since i) it can highlight differences in transcript levels between two subjects (due for instance to polymorphisms in promoter regions), and ii) it covers unannotated genes or pseudogenes not covered by the exome capture protocol. The latter genes can be a source of MAPs and may have a significant contribution to the MiHA landscape. The translated exome-transcriptome were then used as a personalized Mascot protein sequence database, helping to retrieve MAPs overlapping variant locations and most importantly documenting differences in the MAP repertoire of HLA-identical siblings: a MAP present in only one of two HLA-identical siblings is an MiHA.

Example 3: Novel MiHAs identified

Allelic MiHAs comprising the amino acid sequence ELQEKFL_LSL vs. ELQEKF_SSSL have been identified. These MiHAs are presented at the cell surface by the MHC class I allele HLA-B*0801. These peptides are not listed in the Immune Epitope Database (Vita *et al.*, 2010) which includes a repertoire of all well characterized peptides presented by MHC molecules (HLA in humans). These MiHAs derive from a single nucleotide polymorphism in the centromer protein F, 350/400kDa (mitosin) (CENPF) gene. This single nucleotide polymorphism is listed as rs3795517 in the dbSNP database (Sherry *et al.*, 2001) (<http://www.ncbi.nlm.nih.gov/projects/SNP/>). Two alleles are found at this locus; one codes for a CENPF protein containing the ELQEKFL_LSL sequence, and the other codes for a CENPF protein with the ELQEKF_SSSL sequence. The single nucleotide polymorphism corresponds to a T to C substitution at a position corresponding to nucleotide 4409 in the nucleic acid sequence of human CENPF (FIGs. 1A to 1D, NCBI Reference Sequence: NM_016343.3), leading to a leucine to serine substitution at a position corresponding to residue 1412 in the CENPF protein sequence (FIG. 1E, NCBI Reference Sequence: NP_057427.3). CENPF is a transient kinetochore protein that plays multiple roles in cell division. CENPF expression increases in the G2 phase but it is rapidly proteolysed at the end of mitosis. Overexpression of CENPF has been found in various hematopoietic cancers and solid tumors including head and neck squamous cell carcinomas (de la Guardia *et al.*, 2001), breast cancer (O'Brien *et al.*, 2007), non-Hodgkin's lymphoma (Bencimon *et al.*, 2005) and gastrointestinal cancer (Chen *et al.*, 2011). Also, according to the EMBL-EBI Gene Expression Atlas (Kapushesky *et al.*, *Nucl. Acids Res.* (2012) 40 (D1): D1077-D1081), overexpression of CENPF was detected in sarcomas (leiomyosarcomas), glioblastomas, lung adenocarcinomas, colorectal cancers/colon carcinomas, prostate cancer, bladder cancers, lymphomas (peripheral T-cell lymphomas) and leukemias (acute lymphocytic leukemias, T acute lymphoblastic leukemias). Accordingly, in an

embodiment, the peptide of the sequence (I) described herein may be used in the immunotherapy of one or more of these cancers.

The amino acid sequence of another pair of MiHAs identified herein is QELDGVFQKL vs. QELDRVVFQKL. These MiHAs are presented at the cell surface by HLA-B*4403 molecules.

5 These peptides are not listed in the Immune Epitope Database (Vita et al., 2010). These MiHAs derive from a single nucleotide polymorphism in the ZWINT (ZW10 interactor) gene. This single nucleotide polymorphism is listed as rs2241666 in the dbSNP database (Sherry et al., 2001). Two alleles are found at this locus; one codes for a ZWINT protein containing the QELDGVFQKL sequence, and the other codes for a ZWINT protein with the
10 QELDRVVFQKL.sequence. The single nucleotide polymorphism corresponds to an A to G substitution at a position corresponding to nucleotide 596 in the nucleic acid sequence of human ZWINT (FIG. 2A, NCBI Reference Sequence: NM_007057.3), leading to an arginine to glycine substitution at a position corresponding to residue 187 in the ZWINT protein sequence (FIG. 2B, NCBI Reference Sequence: NP_008988.2). ZWINT is a known component of the
15 kinetochore complex that is required for the mitotic spindle checkpoint. Overexpression of ZWINT has been observed in several types of cancer (notably breast, prostate and bladder) and correlates with an increased proliferation of cancer cells (Endo et al., 2012; Ho et al., 2012; Urbanucci et al., 2012). Also, according to the EMBL-EBI Gene Expression Atlas (Kapushesky et al., *Nucl. Acids Res.* (2012) 40 (D1): D1077-D1081), overexpression of ZWINT was detected
20 in lung cancer (adenocarcinoma), head and neck squamous cell carcinoma, colorectal/colon cancer, renal carcinomas as well as lymphomas (mucosa-associated lymphoid tissue lymphomas, peripheral T-cell lymphomas). Accordingly, in an embodiment, the peptide of the sequence (II) described herein may be used in the immunotherapy of one or more of these cancers.

25 The amino acid sequence of a third pair of MiHAs identified herein is SLFFRKVPF vs. SLFFRKVAF. These MiHAs are presented at the cell surface by HLA-B*0801 molecules. These peptides are not listed in the Immune Epitope Database (Vita et al., 2010). These MiHAs derive from a single nucleotide polymorphism in the MTCH2 (mitochondrial carrier homologue 2) gene. This single nucleotide polymorphism is listed as rs1064608 in the dbSNP database (Sherry et al., 2001). Two alleles are found at this locus; one codes for a MTCH2 protein containing the SLFFRKVAF sequence, and the other codes for a MTCH2 protein with the SLFFRKVPF
30 sequence. The single nucleotide polymorphism corresponds to a C to G substitution at a position corresponding to nucleotide 1057 in the nucleic acid sequence of human MTCH2 (FIG. 3A, NCBI Reference Sequence: NM_014342.3), leading to a proline to alanine substitution at a
35 position corresponding to residue 290 in the MTCH2 protein sequence (FIG. 3B, NCBI Reference Sequence: NP_055157.1). MTCH2 interact with proapoptotic truncated BID and thereby regulate apoptosis, and is upregulated/involved in several types of cancer, notably solid

tumors (lung, thyroid, liver, esophagus, colon, breast) and osteosarcoma (Yu *et al.*, 2008, Grinberg *et al.*, 2005; Katz *et al.*, 2012). Also, according to the EMBL-EBI Gene Expression Atlas (Kapushesky *et al.*, *Nucl. Acids Res.* (2012) 40 (D1): D1077-D1081), overexpression of MTCH2 was detected in lymphomas, leukemias and myelomas. Accordingly, in an embodiment,
 5 the peptide of the sequence (III) described herein may be used in the immunotherapy of one or more of these cancers.

The amino acid sequence of a fourth pair of MiHAs identified herein is SVLKPGNSK vs. TVLKPGNSK. These MiHAs are presented at the cell surface by HLA-A*0301 molecules. These peptides are not listed in the Immune Epitope Database (Vita et al., 2010). These MiHAs
 10 derive from a single nucleotide polymorphism in the ELF1 [E74-like factor 1 (ets domain transcription factor)] gene. This single nucleotide polymorphism is listed as rs1056820 in the dbSNP database (Sherry et al., 2001). Two alleles are found at this locus; one codes for an ELF1 protein containing the SVLKPGNSK sequence, and the other codes for an ELF1 protein with the TVLKPGNSK sequence. The single nucleotide polymorphism corresponds to an A to T
 15 substitution at a position corresponding to nucleotide 1400 in the nucleic acid sequence of human ELF1 (FIGs. 4A and 4B, NCBI Reference Sequence: NM_172373.3), leading to a threonine to serine substitution at a position corresponding to residue 343 in the ELF1 protein sequence (FIG. 4C, NCBI Reference Sequence: NP_758961.1). ELF1 is a transcriptional factor that is involved in the transcriptional activation of oncogenic pathways in cancer cells, such as
 20 leukemias, non-small cell lung cancer, breast and ovarian cancer (Andrews et al., 2008; Xiang et al., 2010; Yang et al., 2010). Also, according to the EMBL-EBI Gene Expression Atlas (Kapushesky *et al.*, *Nucl. Acids Res.* (2012) 40 (D1): D1077-D1081), overexpression of ELF1 was detected in sarcomas (e.g., osteosarcomas), glioblastomas, pancreatic cancer and leukemias (acute myeloid leukemia), lymphomas (Burkitt's lymphomas). Accordingly, in an
 25 embodiment, the peptide of the sequence (IV) described herein may be used in the immunotherapy of one or more of these cancers.

The amino acid sequence of another pair of MiHAs is AMYDKGPFRSK vs. AMYDKGPFWSK. These MiHAs are presented at the cell surface by HLA-A*0301. These peptides are not listed in the Immune Epitope Database (Vita et al., 2010). These MiHAs derive
 30 from a single nucleotide polymorphism in the NQO1 [NAD(P)H dehydrogenase, quinone 1] gene. This single nucleotide polymorphism is listed as rs1131341 in the dbSNP database (Sherry et al., 2001). Two alleles are found at this locus; one codes for an NQO1 protein containing the AMYDKGPFRSK sequence, and the other codes for an NQO1 protein with the AMYDKGPFWSK sequence. The single nucleotide polymorphism corresponds to a C to T
 35 substitution at a position corresponding to nucleotide 615 in the nucleic acid sequence of human NQO1 (FIG. 5A, NCBI Reference Sequence: NM_000903.2), leading to an arginine to tryptophan substitution at a position corresponding to residue 139 in the NQO1 protein

sequence (FIG. 5B, NCBI Reference Sequence: NP_000894.1). NQO1 is a cytosolic enzyme that catalyzes the reduction of various quinones using flavin adenine dinucleotide (FAD) as a cofactor. This protein's enzymatic activity prevents the one electron reduction of quinones that results in the production of radical species. Mutations in this NQO1 have been associated with susceptibility to various forms of cancer and altered expression of this NQO1 protein has been seen in many tumors, including non-small cell lung cancer, skin cancer (e.g., melanomas), breast cancer, liver cancer (e.g., intrahepatic cholangiocarcinoma) and digestive tract cancer such as colorectal cancer (Kolesar et al., 2011; Wakai et al., 2011; Jamieson et al., 2011; Ding et al., 2012; Yang et al., 2012; Patrick and Jaiswal, 2012). An NQO1 Substrate was recently shown to possess potent antitumor activity against a wide spectrum of cancer cells, such as pancreatic and lung cancer cells (Huang *et al.*, 2012). Also, according to the EMBL-EBI Gene Expression Atlas (Kapushesky *et al.*, *Nucl. Acids Res.* (2012) 40 (D1): D1077-D1081), overexpression of NQO1 was detected in lymphomas (Hodgkin's lymphomas, anaplastic large cell lymphoma) and brain cancers (glioblastomas, subependymal giant cell astrocytomas). Accordingly, in an embodiment, the peptide of the sequence (V) described herein may be used in the immunotherapy of one or more of these cancers.

The amino acid sequence of another pair of MiHAs is RVSLPTSPG vs. RVSLPTSPR. These MiHAs are presented at the cell surface by HLA-A*0301. These peptides are not listed in the Immune Epitope Database (Vita et al., 2010) which includes a repertoire of all well characterized peptides presented by MHC molecules (HLA in humans).

These MiHAs derive from a single nucleotide polymorphism in the KIAA0226-like gene (also known as C13orf18). This single nucleotide polymorphism is listed as rs1408184 in the dbSNP database (Sherry et al., 2001). Two alleles are found at this locus; one codes for an KIAA0226L protein containing the RVSLPTSPG sequence, and the other codes for an KIAA0226L protein with the RVSLPTSPR sequence. The single nucleotide polymorphism corresponds to a G to A substitution at a position corresponding to nucleotide 1059 in the nucleic acid sequence of human KIAA0226L (FIG. 6A, NCBI Reference Sequence: NM_025113.2), leading to a glycine to arginine substitution at a position corresponding to residue 152 in the KIAA0226L protein sequence (FIG. 6B, NCBI Reference Sequence: NP_079389.2). The function of KIAA0226L is largely unknown, and according to the BioGPS database, KIAA0226L is upregulated in B lymphocytes and Burkitt's lymphoma cells. Also, according to the EMBL-EBI Gene Expression Atlas (Kapushesky *et al.*, *Nucl. Acids Res.* (2012) 40 (D1): D1077-D1081), overexpression of KIAA0226L was detected in colon cancer (carcinoma), glioblastoma, and anaplastic large cell lymphoma (Chowdary D *et al.*, *J Mol Diagn.* 2006 Feb;8(1):31-9; Ancona N et al., *BMC Bioinformatics.* 2006 Aug 19;7:387; Freije WA *et al.*, *Cancer Res.* 2004 Sep 15;64(18):6503-10; Sun *et al.*, *Cancer Cell.* 2006 Apr;9(4):287-300; Piccaluga *et al.*, *J Clin Invest.* 2007 Mar;117(3):823-34. Epub 2007 Feb 15). Accordingly, in an

embodiment, the peptide of the sequence (VI) described herein may be used in the immunotherapy of one or more of these cancers.

The amino acid sequence of another pair of MiHAs is VMGNPGTFK vs. VMGNPGTFN. These MiHAs are presented at the cell surface by HLA-A*0301. These peptides are not listed in the Immune Epitope Database (Vita *et al.*, 2010) which includes a repertoire of all well characterized peptides presented by MHC molecules (HLA in humans).

These MiHAs derive from a single nucleotide polymorphism in the RMDN1 (regulator of microtubule dynamics 1) gene (also known as FAM82B). This single nucleotide polymorphism is listed as rs6980476 in the dbSNP database (Sherry *et al.*, 2001). Two alleles are found at this locus; one codes for an RMDN1 protein containing the VMGNPGTFK sequence, and the other codes for an RMDN1 protein with the VMGNPGTFN sequence. The single nucleotide polymorphism corresponds to an A to C substitution at a position corresponding to nucleotide 316 in the nucleic acid sequence of human RMDN1 (FIG. 7A, NCBI Reference Sequence: NM_016033.2), leading to a lysine to asparagine substitution at a position corresponding to residue 52 in the RMDN1 protein sequence (FIG. 7B, NCBI Reference Sequence: NP_057117.2). RMDN1 is a microtubule-associated protein that plays a role in chromosome segregation in *Caenorhabditis elegans* (Oishi *et al.*, *J Cell Biol.* **179**, 1149-1162, 2007). According to the BioGPS database, RMDN1 is upregulated in B lymphoblasts and Burkitt's lymphoma cells. Also, according to the EMBL-EBI Gene Expression Atlas (Kapushesky *et al.*, *Nucl. Acids Res.* (2012) 40 (D1): D1077-D1081), overexpression of RMDN1 was detected in leiomyosarcoma (Perot *et al.*, *Cancer Res.* 2009 Mar 15;69(6):2269-78; Chibon *et al.*, *Nat Med.* 2010 Jul;16(7):781-7), breast cancer/carcinoma such as invasive ductal carcinoma (Chen *et al.*, *Breast Cancer Res Treat.* 2010 Jan;119(2):335-46; Cheng *et al.*, *Cancer Res.* 2008 Mar 15;68(6):1786-96), and brain cancers such as glioblastomas and astrocytomas (subependymal giant cell astrocytomas (Sun *et al.*, *Cancer Cell.* 2006 Apr;9(4):287-300; Tyburczy *et al.*, *Am J Pathol.* 2010 Apr;176(4):1878-90). Accordingly, in an embodiment, the peptide of the sequence (VII) described herein may be used in the immunotherapy of one or more of these cancers.

Although the present invention has been described hereinabove by way of specific embodiments thereof, it can be modified, without departing from the spirit and nature of the subject invention as defined in the appended claims. In the claims, the word "comprising" is used as an open-ended term, substantially equivalent to the phrase "including, but not limited to". The singular forms "a", "an" and "the" include corresponding plural references unless the context clearly dictates otherwise.

REFERENCES

1. Sykes,M., K.Woods, and D.H.Sachs. 2008. Transplantation Immunology. In Fundamental Immunology. W.E.Paul, editor. Lippincott Williams & Wilkins, Philadelphia. 1426-1488.
- 5 2. Loveland,B. and E.Simpson. 1986. The non-MHC transplantation antigens - neither weak nor minor. Immunol.Today 7:223-229.
3. Perreault,C., F.Decary, S.Brochu, M.Gyger, R.Belanger, and D.Roy. 1990. Minor histocompatibility antigens. Blood 76:1269-1280.
4. Perreault,C. 2010. The origin and role of MHC class I-associated self-peptides. Prog.Mol
10 Biol.Transl.Sci. 92:41-60.
5. de Verteuil,D., D.P.Granados, P.Thibault, and C.Perreault. 2012. Origin and plasticity of MHC I-associated self peptides. Autoimmun.Rev. Epub November 2011.
6. Yewdell,J.W., E.Reits, and J.Neefjes. 2003. Making sense of mass destruction: quantitating MHC class I antigen presentation. Nature Rev.Immunol. 3:952-961.
- 15 7. Neefjes,J., M.L.M.Jongsma, P.Paul, and O.Bakke. 2011. Towards a systems understanding of MHC class I and MHC class II antigen presentation. Nat.Rev.Immunol. 11:823-836.
8. Yewdell,J.W. 2011. DRiPs solidify: progress in understanding endogenous MHC class I antigen processing. Trends Immunol. 32:548-558.
- 20 9. Roopenian,D., E.Y.Choi, and A.Brown. 2002. The immunogenomics of minor histocompatibility antigens. Immunol.Rev. 190:86-94.
10. Spierings,E., M.Hendriks, L.Absi, A.Canossi, S.Chhaya, J.Crowley, H.Dolstra, J.F.Eliaou, T.Ellis, J.Enczmann, M.E.Fasano, T.Gervais, C.Gorodezky, B.Kircher, D.Laurin, M.S.Leffell, P.Loiseau, M.Malkki, M.Markiewicz, M.Martinetti, E.Maruya, N.Mehra, F.Oguz,
25 M.Oudshoorn, N.Pereira, R.Rani, R.Sergeant, J.Thomson, T.H.Tran, H.Turpeinen, K.L.Yang, R.Zunec, M.Carrington, P.de Knijff, and E.Goulmy. 2007. Phenotype frequencies of autosomal minor histocompatibility antigens display significant differences among populations. PLoS.Genet. 3:e103.
11. Perreault,C. and S.Brochu. 2002. Adoptive cancer immunotherapy: discovering the best
30 targets. J.Mol.Med. 80:212-218.

12. Mullally,A. and J.Ritz. 2007. Beyond HLA: the significance of genomic variation for allogeneic hematopoietic stem cell transplantation. *Blood* 109:1355-1362.
13. Feng,X., K.M.Hui, H.M.Younes, and A.G.Brickner. 2008. Targeting minor histocompatibility antigens in graft versus tumor or graft versus leukemia responses. *Trends Immunol.* 29:624-632.
14. Brickner,A.G. 2006. Mechanisms of minor histocompatibility antigen immunogenicity: the role of infinitesimal versus structurally profound polymorphisms. *Immunol.Res.* 36:33-41.
15. Bleakley,M. and S.R.Riddell. 2011. Exploiting T cells specific for human minor histocompatibility antigens for therapy of leukemia. *Immunol.Cell Biol.* 89:396-407.
- 10 16. Kawase,T., Y.Akatsuka, H.Torikai, S.Morishima, A.Oka, A.Tsujimura, M.Miyazaki, K.Tsujimura, K.Miyamura, S.Ogawa, H.Inoko, Y.Morishima, Y.Kodera, K.Kuzushima, and T.Takahashi. 2007. Alternative splicing due to an intronic SNP in HMSD generates a novel minor histocompatibility antigen. *Blood* 110:1055-1063.
- 15 17. Rosenberg,S.A., J.C.Yang, and N.P.Restifo. 2004. Cancer immunotherapy: moving beyond current vaccines. *Nat Med.* 10:909-915.
18. Bleakley,M. and S.R.Riddell. 2004. Molecules and mechanisms of the graft-versus-leukaemia effect. *Nat.Rev.Cancer* 4:371-380.
- 20 19. Rosenberg,S.A., J.C.Yang, R.M.Sherry, U.S.Kammula, M.S.Hughes, G.Q.Phan, D.E.Citrin, N.P.Restifo, P.F.Robbins, J.R.Wunderlich, K.E.Morton, C.M.Laurencot, S.M.Steinberg, D.E.White, and M.E.Dudley. 2011. Durable complete responses in heavily pretreated patients with metastatic melanoma using T cell transfer immunotherapy. *Clin.Cancer Res.* 17:4550-4557.
- 20 20. Vincent,K., D.C.Roy, and C.Perreault. 2011. Next-generation leukemia immunotherapy. *Blood* 118:2951-2959.
- 25 21. Rezvani,K. and A.J.Barrett. 2008. Characterizing and optimizing immune responses to leukaemia antigens after allogeneic stem cell transplantation. *Best.Pract.Res.Clin.Haematol.* 21:437-453.
22. Barrett,A.J. 2008. Understanding and harnessing the graft-versus-leukaemia effect. *Br.J.Haematol.* 142:877-888.

23. Horowitz,M.M., R.P.Gale, P.M.Sondel, J.M.Goldman, J.Kersey, H.J.Kolb, A.A.Rimm, O.Ringden, C.Rozman, B.Speck, and . 1990. Graft-versus-leukemia reactions after bone marrow transplantation. *Blood* 75:555-562.
24. Thomas,E.D. 2005. Foreword. In *Graft-vs.-host disease*. J.L.M.Ferrara, K.R.Coke, and H.J.Deeg, editors. Marcel Dekker, New York. iii-iv.
25. Molldrem,J.J. and W.D.Shlomchik. 2005. Graft-vs.-leukemia effects. In *Graft-vs.-host disease*. J.L.M.Ferrara, K.R.Cooke, and H.J.Deeg, editors. Marcel Dekker, New York. 155-194.
26. O'Reilly,R.J., T.Dao, G.Koehne, D.Scheinberg, and E.Doubrovina. 2010. Adoptive transfer of unselected or leukemia-reactive T-cells in the treatment of relapse following allogeneic hematopoietic cell transplantation. *Semin.Immunol.* 22:162-172.
27. Kolb,H.J., A.Schattenberg, J.M.Goldman, B.Hertenstein, N.Jacobsen, W.Arcese, P.Ljungman, A.Ferrant, L.Verdonck, and D.Niederwieser. 1995. Graft-versus-leukemia effect of donor lymphocyte transfusions in marrow grafted patients. *European Group for Blood and Marrow Transplantation Working Party Chronic Leukemia. Cancer Cell* 86:2041-2050.
28. Kolb,H.-J. 2008. Graft-versus-leukemia effects of transplantation and donor lymphocytes. *Blood* 112:4371-4383.
29. Childs,R.W. and J.Barrett. 2004. Nonmyeloablative allogeneic immunotherapy for solid tumors. *Annu.Rev.Med.* 55:459-475.
30. Tykodi,S.S., E.H.Warren, J.A.Thompson, S.R.Riddell, R.W.Childs, B.E.Otterud, M.F.Leppert, R.Storb, and B.M.Sandmaier. 2004. Allogeneic hematopoietic cell transplantation for metastatic renal cell carcinoma after nonmyeloablative conditioning: toxicity, clinical response, and immunological response to minor histocompatibility antigens. *Clin.Cancer Res.* 10:7799-7811.
31. Bishop,M.R., D.H.Fowler, D.Marchigiani, K.Castro, C.Kasten-Sportes, S.M.Steinberg, J.C.Gea-Banacloche, R.Dean, C.K.Chow, C.Carter, E.J.Read, S.Leitman, and R.Gress. 2004. Allogeneic lymphocytes induce tumor regression of advanced metastatic breast cancer. *J.Clin.Oncol.* 22:3886-3892.
32. Takahashi,Y., N.Harashima, S.Kajigaya, H.Yokoyama, E.Cherkasova, J.P.McCoy, K.I.Hanada, O.Mena, R.Kurlander, T.Abdul, R.Srinivasan, A.Lundqvist, E.Malinzak, N.Geller, M.I.Lerman, and R.W.Childs. 2008. Regression of human kidney cancer following allogeneic

stem cell transplantation is associated with recognition of an HERV-E antigen by T cells. *J.Clin.Invest.* 118:1099-1109.

33. Meunier,M.C., J.S.Delisle, J.Bergeron, V.Rineau, C.Baron, and C.Perreault. 2005. T cells targeted against a single minor histocompatibility antigen can cure solid tumors. *Nat.Med.* 11:1222-1229.
34. Shlomchik,W.D. 2007. Graft-versus-host disease. *Nat.Rev.Immunol.* 7:340-352.
35. Ferrara,J.L., J.E.Levine, P.Reddy, and E.Holler. 2009. Graft-versus-host disease. *Lancet* 373:1550-1561.
36. Socie,G. and B.R.Blazar. 2009. Acute graft-versus-host disease; from the bench to the bedside. *Blood* 114:4327-4336.
37. Greinix,H.T., C.Loddenkemper, S.Z.Pavletic, E.Holler, G.Socie, A.Lawitschka, J.Halter, and D.Wolff. 2011. Diagnosis and staging of chronic graft-versus-host disease in the clinical practice. *Biol.Blood Marrow Transplant.* 17:167-175.
38. Vogelsang,G.B., L.Lee, and D.M.Bensen-Kennedy. 2003. Pathogenesis and treatment of graft-versus-host disease after bone marrow transplant. *Annu.Rev.Med.* 54:29-52.
39. Inaba,M., K.Kurasawa, M.Mamura, K.Kumano, Y.Saito, and I.Iwamoto. 1999. Primed T cells are more resistant to Fas-mediated activation-induced cell death than naive T cells. *J Immunol* 163:1315-1320.
40. Yang,J., M.O.Brook, M.Carvalho-Gaspar, J.Zhang, H.E.Ramon, M.H.Sayegh, K.J.Wood, L.A.Turka, and N.D.Jones. 2007. Allograft rejection mediated by memory T cells is resistant to regulation. *Proc.Natl.Acad.Sci.U.S.A* 104:19954-19959.
41. Massague,J. 2008. TGFb in Cancer. *Cell* 134:215-230.
42. Hanahan,D. and R.A.Weinberg. 2011. Hallmarks of cancer: the next generation. *Cell* 144:646-674.
43. Fontaine,P., G.Roy-Proulx, L.Knafo, C.Baron, D.C.Roy, and C.Perreault. 2001. Adoptive transfer of T lymphocytes targeted to a single immunodominant minor histocompatibility antigen eradicates leukemia cells without causing graft-versus-host disease. *Nat.Med.* 7:789-794.
44. Meunier,M.C., C.Baron, and C.Perreault. 2009. Two host factors regulate persistence of H7a-specific T cells injected in tumor bearing mice. *PLoS One* 4:e4116.

45. Fortier, M.H., E. Caron, M.P. Hardy, G. Voisin, S. Lemieux, C. Perreault, and P. Thibault. 2008. The MHC class I peptide repertoire is molded by the transcriptome. *J. Exp. Med.* 205:595-610.
46. de Verteuil, D., T.L. Muratore-Schroeder, D.P. Granados, M.H. Fortier, M.P. Hardy, A. Bramoullé, E. Caron, K. Vincent, S. Mader, S. Lemieux, P. Thibault, and C. Perreault. 2010. Deletion of immunoproteasome subunits imprints on the transcriptome and has a broad impact on peptides presented by major histocompatibility complex I molecules. *Mol Cell Proteomics* 9:2034-2047.
47. Caron, E., K. Vincent, M.H. Fortier, J.P. Laverdure, A. Bramoullé, M.P. Hardy, G. Voisin, P. Roux, S. Lemieux, P. Thibault, and C. Perreault. 2011. The MHC I immunopeptidome conveys to the cell surface an integrative view of cellular regulation. *Mol. Syst. Biol.* 7:533.
48. den Haan, J.M., N.E. Sherman, E. Blokland, E. Huczko, F. Koning, J.W. Drijfhout, J. Skipper, J. Shabanowitz, D.F. Hunt, V.H. Engelhard, and E. Goulmy. 1995. Identification of a graft versus host disease-associated human minor histocompatibility antigen. *Science* 268:1476-1480.
49. den Haan, J.M., L.M. Meadows, W. Wang, J. Pool, E. Blokland, T.L. Bishop, C. Reinhardus, J. Shabanowitz, R. Offringa, D.F. Hunt, V.H. Engelhard, and E. Goulmy. 1998. The minor histocompatibility antigen HA-1: a diallelic gene with a single amino acid polymorphism. *Science* 279:1054-1057.
50. Brickner, A.G., E.H. Warren, J.A. Caldwell, Y. Akatsuka, T.N. Golovina, A.L. Zarling, J. Shabanowitz, L.C. Eisenlohr, D.F. Hunt, V.H. Engelhard, and S.R. Riddell. 2001. The immunogenicity of a new human minor histocompatibility antigen results from differential antigen processing. *J. Exp. Med.* 193:195-205.
51. Spierings, E., A.G. Brickner, J.A. Caldwell, S. Zegveld, N. Tatsis, E. Blokland, J. Pool, R.A. Pierce, S. Mollah, J. Shabanowitz, L.C. Eisenlohr, V.P. van, F. Ossendorp, D.F. Hunt, E. Goulmy, and V.H. Engelhard. 2003. The minor histocompatibility antigen HA-3 arises from differential proteasome-mediated cleavage of the lymphoid blast crisis (Lbc) oncoprotein. *Blood* 102:621-629.
52. Brickner, A.G., A.M. Evans, J.K. Mito, S.M. Xuereb, X. Feng, T. Nishida, L. Fairfull, R.E. Ferrell, K.A. Foon, D.F. Hunt, J. Shabanowitz, V.H. Engelhard, S.R. Riddell, and E.H. Warren. 2006. The PANE1 gene encodes a novel human minor histocompatibility antigen that is selectively expressed in B-lymphoid cells and B-CLL. *Blood* 107:3779-3786.

53. van Bergen, C.A.M., M.G.D.Kester, I.Jedema, M.H.M.Heemskerk, S.A.P.van Luxemburg-Heijs, F.M.Kloosterboer, W.A.E.Marijt, A.H.de Ru, M.R.Schaafsma, R.Willemze, P.A.van Veelen, and J.H.F.Falkenburg. 2007. Multiple myeloma-reactive T cells recognize an activation-induced minor histocompatibility antigen encoded by the ATP-dependent interferon-responsive (ADIR) gene. *Blood* 109:4089-4096.
54. Slager, E.H., M.W.Honders, E.D.Van Der Meijden, S.A.Van Luxemburg-Heijs, F.M.Kloosterboer, M.G.Kester, I.Jedema, W.A.Marijt, M.R.Schaafsma, R.Willemze, and J.H.Falkenburg. 2006. Identification of the angiogenic endothelial-cell growth factor-1/thymidine phosphorylase as a potential target for immunotherapy of cancer. *Blood* 107:4954-4960.
55. Dolstra, H., H.Fredrix, F.Maas, P.G.Coulie, F.Brasseur, E.Mensink, G.J.Adema, T.M.de Witte, C.G.Figdor, and E.van de Wiel-van Kemenade. 1999. A human minor histocompatibility antigen specific for B cell acute lymphoblastic leukemia. *J.Exp.Med.* 189:301-308.
56. Murata, M., E.H.Warren, and S.R.Riddell. 2003. A human minor histocompatibility antigen resulting from differential expression due to a gene deletion. *J.Exp.Med.* 197:1279-1289.
57. Warren, E.H., N.J.Vigneron, M.A.Gavin, P.G.Coulie, V.Stroobant, A.Dalet, S.S.Tykodi, S.M.Xuereb, J.K.Mito, S.R.Riddell, and B.J.Van Den Eynde. 2006. An antigen produced by splicing of noncontiguous peptides in the reverse order. *Science* 313:1444-1447.
58. Griffioen, M., E.D.Van Der Meijden, E.H.Slager, M.W.Honders, C.E.Rutten, S.A.Van Luxemburg-Heijs, P.A.von dem Borne, J.J.van Rood, R.Willemze, and J.H.Falkenburg. 2008. Identification of phosphatidylinositol 4-kinase type II beta as HLA class II-restricted target in graft versus leukemia reactivity. *Proc.Natl.Acad.Sci.U.S.A* 105:3837-3842.
59. Stumpf, A.N., E.D.Van Der Meijden, C.A.Van Bergen, R.Willemze, J.H.Falkenburg, and M.Griffioen. 2009. Identification of 4 new HLA-DR-restricted minor histocompatibility antigens as hematopoietic targets in antitumor immunity. *Blood* 114:3684-3692.
60. Akatsuka, Y., T.Nishida, E.Kondo, M.Miyazaki, H.Taji, H.Iida, K.Tsujimura, M.Yazaki, T.Naoe, Y.Morishima, Y.Kodera, K.Kuzushima, and T.Takahashi. 2003. Identification of a polymorphic gene, BCL2A1, encoding two novel hematopoietic lineage-specific minor histocompatibility antigens. *J.Exp.Med.* 197:1489-1500.
61. Rijke, B.D., A.Horssen-Zoetbrood, J.M.Beekman, B.Otterud, F.Maas, R.Woestenenk, M.Kester, M.Leppert, A.V.Schattenberg, T.de Witte, van de Wiel-van Kemenade, and H.Dolstra.

2005. A frameshift polymorphism in P2X5 elicits an allogeneic cytotoxic T lymphocyte response associated with remission of chronic myeloid leukemia. *J.Clin.Invest* 115:3506-3516.

62. Kawase,T., Y.Nannya, H.Torikai, G.Yamamoto, M.Onizuka, S.Morishima, K.Tsujimura, K.Miyamura, Y.Kodera, Y.Morishima, T.Takahashi, K.Kuzushima, S.Ogawa, and Y.Akatsuka.

5 2008. Identification of human minor histocompatibility antigens based on genetic association with highly parallel genotyping of pooled DNA. *Blood* 111:3286-3294.

63. Kamei,M., Y.Nannya, H.Torikai, T.Kawase, K.Taura, Y.Inamoto, T.Takahashi, M.Yazaki, S.Morishima, K.Tsujimura, K.Miyamura, T.Ito, H.Togari, S.R.Riddell, Y.Kodera, Y.Morisima, T.Takahashi, K.Kuzushima, S.Ogawa, and Y.Akatsuka. 2009. HapMap scanning of novel

10 human minor histocompatibility antigens. *Blood* 113:5041-5048.

64. Van Bergen,C.A., C.E.Rutten, E.D.Van Der Meijden, S.A.Van Luxemburg-Heijs, E.G.Lurvink, J.J.Houwing-Duistermaat, M.G.Kester, A.Mulder, R.Willemze, J.H.Falkenburg, and M.Griffioen. 2010. High-throughput characterization of 10 new minor histocompatibility antigens by whole genome association scanning. *Cancer Res.* 70:9073-9083.

15 65. Spaapen,R.M., H.M.Lokhorst, K.van den Oudenalder, B.E.Otterud, H.Dolstra, M.F.Leppert, M.C.Minnema, A.C.Bloem, and T.Mutis. 2008. Toward targeting B cell cancers with CD4+ CTLs: identification of a CD19-encoded minor histocompatibility antigen using a novel genome-wide analysis. *J Exp.Med* 205:2863-2872.

66. Spaapen,R.M., R.A.de Kort, K.van den Oudenalder, E.M.van, A.C.Bloem, H.M.Lokhorst, 20 and T.Mutis. 2009. Rapid identification of clinical relevant minor histocompatibility antigens via genome-wide zygosity-genotype correlation analysis. *Clin.Cancer Res.* 15:7137-7143.

67. Warren,E.H., N.Fujii, Y.Akatsuka, C.N.Chaney, J.K.Mito, K.R.Loeb, T.A.Gooley, M.L.Brown, K.K.Koo, K.V.Rosinski, S.Ogawa, A.Matsubara, F.R.Appelbaum, and S.R.Riddell. 2010. Therapy of relapsed leukemia after allogeneic hematopoietic cell transplant with T cells 25 specific for minor histocompatibility antigens. *Blood* 115:3869-3878.

68. Mason,D. 1998. A very high level of crossreactivity is an essential feature of the T-cell receptor. *Immunol.Today* 19:395-404.

69. Kessler,J.H. and C.J.Melief. 2007. Identification of T-cell epitopes for cancer immunotherapy. *Leukemia* 21:1859-1874.

70. Popovic,J., L.P.Li, P.M.Kloetzel, M.Leisegang, W.Uckert, and T.Blankenstein. 2011. The only proposed T-cell epitope derived from the TEL-AML1 translocation is not naturally processed. *Blood* 118:946-954.
71. Schreiber,H., J.D.Rowley, and D.A.Rowley. 2011. Targeting mutations predictably.
5 *Blood* 118:830-831.
72. Granados,D.P., W.Yahyaoui, C.M.Laumont, T.Daouda, T.L.Muratore-Schroeder, C.Cote, J.P.Laverdure, S.Lemieux, P.Thibault, and C.Perreault. 2012. MHC I-associated peptides preferentially derive from transcripts bearing miRNA recognition elements. *Blood Epub* March 21, 2012.
- 10 73. Karosiene,E., Lundegaard,C., Lund,O., and Nielsen,M. (2011). NetMHCcons: a consensus method for the major histocompatibility complex class I predictions. *Immunogenetics*.
74. Tosato,G. and Cohen,J.I. (2007). Generation of Epstein-Barr Virus (EBV)-immortalized B cell lines. *Curr. Protoc. Immunol.* Chapter 7, Unit.
- 15 75. Bencimon,C., Salles,G., Moreira,A., Guyomard,S., Coiffier,B., Bienvenu,J., and Fabien,N. (2005). Prevalence of anticentromere F protein autoantibodies in 347 patients with non-Hodgkin's lymphoma. *Ann. N. Y. Acad. Sci.* 1050, 319-326.
76. Chen,W.B., Cheng,X.B., Ding,W., Wang,Y.J., Chen,D., Wang,J.H., and Fei,R.S. (2011). Centromere protein F and survivin are associated with high risk and a poor prognosis in
20 colorectal gastrointestinal stromal tumours. *J Clin. Pathol.* 64, 751-755.
77. de la Guardia,C., Casiano,C.A., Trinidad-Pinedo,J., and Baez,A. (2001). CENP-F gene amplification and overexpression in head and neck squamous cell carcinomas. *Head Neck* 23, 104-112.
78. O'Brien,S.L., Fagan,A., Fox,E.J., Millikan,R.C., Culhane,A.C., Brennan,D.J.,
25 McCann,A.H., Hegarty,S., Moyna,S., Duffy,M.J., Higgins,D.G., Jirstrom,K., Landberg,G., and Gallagher,W.M. (2007). CENP-F expression is associated with poor prognosis and chromosomal instability in patients with primary breast cancer. *Int. J Cancer* 120, 1434-1443.
79. Sherry,S.T., Ward,M.H., Kholodov,M., Baker,J., Phan,L., Smigielski,E.M., and Sirotkin,K. (2001). dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* 29, 308-311.

80. Vita,R., Zarebski,L., Greenbaum,J.A., Emami,H., Hoof,I., Salimi,N., Damle,R., Sette,A., and Peters,B. (2010). The immune epitope database 2.0. *Nucleic Acids Res.* 38, D854-D862.
81. Endo,H., Ikeda,K., Urano,T., Horie-Inoue,K., and Inoue,S. (2012). Terf/TRIM17 stimulates degradation of kinetochore protein ZWINT and regulates cell proliferation. *J. Biochem.* 151, 139-144.
82. Ho,J.R., Chapeaublanc,E., Kirkwood,L., Nicolle,R., Benhamou,S., Lebret,T., Allory,Y., Southgate,J., Radvanyi,F., and Goud,B. (2012). Deregulation of rab and rab effector genes in bladder cancer. *PLoS. ONE.* 7, e39469.
83. Urbanucci,A., Sahu,B., Seppala,J., Larjo,A., Latonen,L.M., Waltering,K.K., Tammela,T.L., Vessella,R.L., Lahdesmaki,H., Janne,O.A., and Visakorpi,T. (2012). Overexpression of androgen receptor enhances the binding of the receptor to the chromatin in prostate cancer. *Oncogene* 31, 2153-2163.
84. Grinberg,M., Schwarz,M., Zaltsman,Y., Eini,T., Niv,H., Pietrokovski,S., and Gross,A. (2005). Mitochondrial carrier homolog 2 is a target of tBID in cells signaled to die by tumor necrosis factor alpha. *Mol. Cell Biol.* 25, 4579-4590.
85. Katz,C., Zaltsman-Amir,Y., Mostizky,Y., Kollet,N., Gross,A., and Friedler,A. (2012). Molecular basis of the interaction between proapoptotic truncated BID (tBID) protein and mitochondrial carrier homologue 2 (MTCH2) protein: key players in mitochondrial death pathway. *J. Biol. Chem.* 287, 15016-15023.
86. Yu,K., Ganesan,K., Tan,L.K., Laban,M., Wu,J., Zhao,X.D., Li,H., Leung,C.H., Zhu,Y., Wei,C.L., Hooi,S.C., Miller,L., and Tan,P. (2008). A precisely regulated gene expression cassette potently modulates metastasis and survival in multiple solid cancers. *PLoS. Genet.* 4, e1000129.
87. Andrews,P.G., Kennedy,M.W., Popadiuk,C.M., and Kao,K.R. (2008). Oncogenic activation of the human Pygopus2 promoter by E74-like factor-1. *Mol. Cancer Res.* 6, 259-266.
88. Xiang,P., Lo,C., Argiropoulos,B., Lai,C.B., Rouhi,A., Imren,S., Jiang,X., Mager,D., and Humphries,R.K. (2010). Identification of E74-like factor 1 (ELF1) as a transcriptional regulator of the Hox cofactor MEIS1. *Exp. Hematol.* 38, 798-8, 808.
89. Yang,D.X., Li,N.E., Ma,Y., Han,Y.C., and Shi,Y. (2010). Expression of Elf-1 and survivin in non-small cell lung cancer and their relationship to intratumoral microvessel density. *Chin J. Cancer* 29, 396-402.

90. Ding,R., Lin,S., and Chen,D. (2012). Association of NQO1 rs1800566 polymorphism and the risk of colorectal cancer: a meta-analysis. *Int. J Colorectal Dis.* 27, 885-892.
91. Jamieson,D., Cresti,N., Bray,J., Sludden,J., Griffin,M.J., Hawsawi,N.M., Famie,E., Mould,E.V., Verrill,M.W., May,F.E., and Boddy,A.V. (2011). Two minor NQO1 and NQO2 alleles
5 predict poor response of breast cancer patients to adjuvant doxorubicin and cyclophosphamide therapy. *Pharmacogenet. Genomics* 21, 808-819.
92. Kolesar,J.M., Dahlberg,S.E., Marsh,S., McLeod,H.L., Johnson,D.H., Keller,S.M., and Schiller,J.H. (2011). The NQO1*2/*2 polymorphism is associated with poor overall survival in patients following resection of stages II and IIIa non-small cell lung cancer. *Oncol. Rep.* 25,
10 1765-1772.
93. Patrick, B.A. and Jaiswal, A.K. (2012). Stress-induced NQO1 controls stability of C/EBPalpha against 20S proteasomal degradation to regulate p63 expression with implications in protection against chemical-induced skin cancer. *Oncogene.* 2012 Jan 16. doi: 10.1038/onc.2011.600. [Epub ahead of print].
- 15 94. Wakai,T., Shirai,Y., Sakata,J., Matsuda,Y., Korita,P.V., Takamura,M., Ajioka,Y., and Hatakeyama,K. (2011). Prognostic significance of NQO1 expression in intrahepatic cholangiocarcinoma. *Int. J Clin. Exp Pathol.* 4, 363-370.
95. Yang,F.Y., Guan,Q.K., Cui,Y.H., Zhao,Z.Q., Rao,W., and Xi,Z. (2012). NAD(P)H quinone oxidoreductase 1 (NQO1) genetic C609T polymorphism is associated with the risk of digestive
20 tract cancer: a meta-analysis based on 21 case-control studies. *Eur. J Cancer Prev.* 2012 Sep;21(5):432-41.
96. Huang X, Dong Y, Bey EA, Kilgore JA, Bair JS, Li LS, Patel M, Parkinson EI, Wang Y, Williams NS, Gao J, Hergenrother PJ, Boothman DA. *Cancer Res.* 2012 Jun 15;72(12):3038-47. Epub 2012 Apr 24.
- 25 97. Oishi,K., Okano,H., and Sawa,H. (2007). RMD-1, a novel microtubule-associated protein, functions in chromosome segregation in *Caenorhabditis elegans*. *J Cell Biol.* 179, 1149-1162.

WHAT IS CLAIMED IS:

1. A method of identifying a minor histocompatibility antigen (MiHA) candidate, the method comprising:

(a) isolating and sequencing MHC-associated peptides (MAPs) in a first cell sample
5 from a first subject;

(b) performing a whole transcriptome and/or exome sequencing on a second cell sample obtained from said first subject;

(c) comparing the sequenced whole transcriptome and/or exome to a reference genome to identify single nucleotide variations (SNVs) between the transcriptome and/or
10 exome of said first subject and the reference genome;

(d) *in silico* translating the sequences containing the identified SNVs to identify peptide sequences comprising at least one non-synonymous mutation caused by said SNVs;

(e) comparing the sequences of the MAPs isolated in (a) with the peptide sequences
15 identified in (d); and

(f) identifying a MiHA candidate based on said comparison.

2. A method of identifying a minor histocompatibility antigen (MiHA) candidate, the method comprising:

(a) isolating and sequencing MHC-associated peptides (MAPs) in a first cell sample
20 from a first and second subjects, wherein said first and second subjects are human leukocyte antigen (HLA)-matched;

(b) performing a whole transcriptome and/or exome sequencing on a second cell sample obtained from said first and second subjects;

(c) comparing the sequenced whole transcriptomes and/or exomes to identify single nucleotide variations (SNVs) between the transcriptomes and/or exomes of said first and
25 second subjects;

(d) *in silico* translating the sequences containing the identified SNVs to identify peptide sequences comprising at least one non-synonymous mutation caused by said SNVs;

(e) comparing the sequences of the MAPs isolated in (a) with the peptide sequences
30 identified in (d); and

(f) identifying a MiHA candidate based on said comparison.

3. The method of claim 1, wherein said MiHA candidate is a MAP whose sequence comprises at least one mutation relative to the corresponding sequence translated from the
35 reference genome.

4. The method of claim 2, wherein said MiHA candidate is a MAP present in the first cell sample from said first subject but absent from the first cell sample from said second subject.
5. The method of claim 1 or 3, wherein said reference genome is the Genome Reference Consortium Human Build 37 (GRCh37).
- 5 6. The method of any one of claims 1 to 5, wherein said first and/or second cell sample is a peripheral blood cell sample.
7. The method of claim 6, wherein said peripheral blood cell sample is an immortalized peripheral blood cell sample.
8. The method of claim 7, wherein said immortalized peripheral blood cell sample is an
10 Epstein-Barr virus (EBV)-transformed B lymphoblastoid cell line.
9. The method of any one of claims 1 to 8, wherein said isolating MAPs comprises (i) releasing said MAPs from said cell sample by mild acid treatment; and (ii) subjecting the released MAPs to chromatography.
10. The method of claim 9, wherein said method further comprises filtering the released
15 peptides with a size exclusion column prior to said chromatography.
11. The method of claim 10, wherein said size exclusion column has a cut-off of about 3000 Da.
12. The method of any one of claim 9 to 11, wherein said chromatography is cation exchange chromatography.
- 20 13. The method of any one of claims 1 to 12, wherein said peptide sequences of (d) have a length of 12 amino acids or less.
14. The method of claim 13, wherein said peptide sequences of (d) have a length of 8 to 11 amino acids.
15. The method of any one of claims 1 to 14, wherein said comparing comprises subjecting
25 the MAPs isolated in (a) to mass spectrometry and comparing the MS spectra obtained with the peptide sequences identified in (d).
16. The method of any one of claims 1 to 15, further comprising determining the binding of the MiHA candidate identified in (f) to a major histocompatibility complex (MHC) class I molecule.

17. A peptide of 50 amino acids or less comprising the sequence



wherein

Z^1 is an amino terminal modifying group or is absent;

5 X^1 is a sequence of 1 to 43 amino acids or is absent;

X^2 is L or S;

X^3 is a sequence of 1 to 43 amino acids or is absent; and

Z^2 is a carboxy terminal modifying group or is absent.

18. The peptide of claim 17, wherein said peptide has a length of 8 to 12 amino acids.

10 19. The peptide of claim 17 or 18, wherein X^1 is an acidic amino acid.

20. The peptide of claim 19, wherein X^1 is glutamic acid (E).

21. The peptide of any one of claims 17 to 20, wherein X^3 is an amino acid.

22. The peptide of claim 21, wherein X^3 is a hydrophobic amino acid.

23. The peptide of claim 22, wherein X^3 is leucine (L).

15 24. The peptide of claim any one of claims 17 to 23, wherein Z^1 is absent.

25. The peptide of any one of claims 17 to 24, wherein Z^2 is absent.

26. The peptide of any one of claims 17 to 25, wherein X^2 is L.

27. The peptide of any one of claims 17 to 25, wherein X^2 is S.

28. The peptide of claim 26, wherein said peptide is ELQEKFLSL (SEQ ID NO:15).

20 29. The peptide of claim 27, wherein said peptide is ELQEKFSSL (SEQ ID NO:16).

30. A nucleic acid encoding the peptide of any one of claims 17 to 29.

31. An isolated major histocompatibility complex (MHC) class I molecule of the HLA-B*0801 allele loaded with the peptide of any one of claims 17 to 29.

25 32. An isolated cell expressing at its surface a MHC class I molecule of the HLA-B*0801 allele loaded with the peptide of any one of claims 17 to 29.

33. A method of treating cancer, said method comprising administering to a subject in need thereof an effective amount of CD8 T lymphocytes recognizing a MHC class I molecule of the HLA-B*0801 allele loaded with the peptide of any one of claims 17 to 29.

34. The method of claim 33, said method further comprising determining whether said
 5 subject expresses a CENPF nucleic acid comprising a T or a C at a position corresponding to nucleotide 4409 in the nucleic acid sequence of human CENPF (FIGs. 1A to 1D, NCBI Reference Sequence: NM_016343.3), and/or a CENPF polypeptide comprising a leucine or serine residue at a position corresponding to residue 1412 in the protein sequence of human CENPF (FIG. 1E, NCBI Reference Sequence: NP_057427.3), wherein

10 (a) if said subject expresses a CENPF nucleic acid comprising a T at a position corresponding to nucleotide 4409 in the nucleic acid sequence of human CENPF and/or a CENPF polypeptide comprising a leucine residue at a position corresponding to residue 1412 in the protein sequence of human CENPF, X^2 is L in said peptide;

(b) if said subject expresses a CENPF nucleic acid comprising a C at a position
 15 corresponding to nucleotide 4409 in the nucleic acid sequence of human CENPF and/or a CENPF polypeptide comprising a serine residue at a position corresponding to residue 1412 in the protein sequence of human CENPF, X^2 is S in said peptide.

35. The method of claim 34, wherein said determining comprises sequencing a CENPF nucleic acid.

20 36. The method of claim 34, wherein said CD8 T lymphocytes are *in vitro* expanded CD8 T lymphocytes.

37. The method of any one of claims 34 to 36, wherein said method further comprises:

(i) if said subject is the subject of (a), culturing CD8 T lymphocytes from a second
 25 subject comprising a C at a position corresponding to nucleotide 4409 in the nucleic acid sequence of human CENPF and/or a CENPF polypeptide comprising a serine residue at a position corresponding to residue 1412 in the protein sequence of human CENPF in the presence of cells expressing a MHC class I molecule of the HLA-B*0801 allele loaded with the peptide of any one of claims 17 to 29, wherein X^2 is L in said peptide, under conditions suitable for CD8 T lymphocyte expansion; or

30 (ii) if said subject is the subject of (b), culturing CD8 T lymphocytes from a second subject comprising a T at a position corresponding to nucleotide 4409 in the nucleic acid sequence of human CENPF and/or a CENPF polypeptide comprising a leucine residue at a position corresponding to residue 1412 in the protein sequence of human CENPF in the presence of cells expressing a MHC class I molecule of the HLA-B*0801 allele loaded with the

peptide of any one of claims 17 to 29, wherein X^2 is S in said peptide, under conditions suitable for CD8 T lymphocyte expansion.

38. The method of any one of claims 33 to 37, wherein said subject is an allogeneic stem cell transplantation (ASCT) recipient.

5 39. A method of expanding CD8 T lymphocytes for adoptive T-cell immunotherapy, said method comprising:

(a) determining whether a candidate donor expresses a CENPF nucleic acid comprising a T or a C at a position corresponding to nucleotide 4409 in the nucleic acid sequence of human CENPF (FIGs. 1A to 1D, NCBI Reference Sequence: NM_016343.3),
10 and/or a CENPF polypeptide comprising a leucine or serine residue at a position corresponding to residue 1412 in the protein sequence of human CENPF (FIG. 1E, NCBI Reference Sequence: NP_057427.3) and

(b)(i) if said candidate donor expresses a CENPF nucleic acid comprising a T at a position corresponding to nucleotide 4409 in the nucleic acid sequence of human CENPF
15 and/or a CENPF polypeptide comprising a leucine residue at a position corresponding to residue 1412 in the protein sequence of human CENPF, culturing CD8 T lymphocytes from said candidate donor in the presence of cells expressing a MHC class I molecule of the HLA-B*0801 allele loaded with the peptide of any one of claims 17 to 29, wherein X^2 is S in said peptide, under conditions suitable for CD8 T lymphocyte expansion; or

20 (b)(ii) if said candidate donor expresses a CENPF nucleic acid comprising a C at a position corresponding to nucleotide 4409 in the nucleic acid sequence of human CENPF and/or a CENPF polypeptide comprising a serine residue at a position corresponding to residue 1412 in the protein sequence of human CENPF, culturing CD8 T lymphocytes from said candidate donor in the presence of cells expressing a MHC class I molecule of the HLA-B*0801
25 allele loaded with the peptide of any one of claims 17 to 29, wherein X^2 is L in said peptide, under conditions suitable for CD8 T lymphocyte expansion.

40. A peptide of 50 amino acids or less comprising the sequence



wherein

30 Z^1 is an amino terminal modifying group or is absent;

X^4 is a sequence of 1 to 43 amino acids or is absent;

X^5 is G or R;

X^6 is an amino acid or is absent;

X^7 is a sequence of 1 to 43 amino acids or is absent; and

35 Z^2 is a carboxy terminal modifying group or is absent.

41. The peptide of claim 40, wherein said peptide has a length of 8 to 12 amino acids.
42. The peptide of claim 40 or 41, wherein X^4 is glutamine (Q).
43. The peptide of any one of claims 40 to 42, wherein X^6 is an amino acid.
44. The peptide of claim 43, wherein X^6 is a basic amino acid.
- 5 45. The peptide of claim 44, wherein X^6 is lysine (K).
46. The peptide of any one of claims 40 to 45, wherein X^7 is an amino acid.
47. The peptide of claim 46, wherein X^7 is leucine (L).
48. The peptide of claim any one of claims 40 to 47, wherein Z^1 is absent.
49. The peptide of any one of claims 40 to 48, wherein Z^2 is absent.
- 10 50. The peptide of any one of claims 40 to 49, wherein X^5 is G.
51. The peptide of any one of claims 40 to 49, wherein X^5 is R.
52. The peptide of claim 50, wherein said peptide is QELDGVFQKL (SEQ ID NO:17).
53. The peptide of claim 51, wherein said peptide is QELDRV FQKL (SEQ ID NO:18).
54. A nucleic acid encoding the peptide of any one of claims 40 to 53.
- 15 55. An isolated major histocompatibility complex (MHC) class I molecule of the HLA-B*4403 allele loaded with the peptide of any one of claims 40 to 53.
56. An isolated cell expressing at its surface a MHC class I molecule of the HLA-B*4403 allele loaded with the peptide of any one of claims 40 to 53.
57. A method of treating cancer, said method comprising administering to a subject in need
20 thereof an effective amount of CD8 T lymphocytes recognizing a MHC class I molecule of the HLA-B*4403 allele loaded with the peptide of any one of claims 40 to 53.
58. The method of claim 57, said method further comprising determining whether said
subject expresses a ZWINT nucleic acid comprising an A or a G at a position corresponding to
nucleotide 596 in the nucleic acid sequence of human ZWINT (FIG. 2A, NCBI Reference
25 Sequence: NM_007057.3), and/or a ZWINT polypeptide comprising a arginine or glycine

residue at a position corresponding to residue 187 in the protein sequence of human ZWINT (FIG. 2B, NCBI Reference Sequence: NP_008988.2), wherein

(a) if said subject expresses a ZWINT nucleic acid comprising an A at a position corresponding to nucleotide 596 in the nucleic acid sequence of human ZWINT and/or a ZWINT polypeptide comprising an arginine residue at a position corresponding to residue 187 in the protein sequence of human ZWINT, X⁵ is R in said peptide;

(b) if said subject expresses a ZWINT nucleic acid comprising a G at a position corresponding to nucleotide 596 in the nucleic acid sequence of human ZWINT and/or a ZWINT polypeptide comprising a glycine residue at a position corresponding to residue 187 in the protein sequence of human ZWINT, X⁵ is G in said peptide.

59. The method of claim 58, wherein said determining comprises sequencing a human ZWINT nucleic acid.

60. The method of any one of claims 57 to 59, wherein said CD8 T lymphocytes are *in vitro* expanded CD8 T lymphocytes.

61. The method of any one of claims 58 to 60, wherein said method further comprises:

(i) if said subject is the subject of (a), culturing CD8 T lymphocytes from a second subject comprising a G at a position corresponding to nucleotide 596 in the nucleic acid sequence of human ZWINT and/or a ZWINT polypeptide comprising a glycine residue at a position corresponding to residue 187 in the protein sequence of human ZWINT in the presence of cells expressing a MHC class I molecule of the HLA-B*4403 allele loaded with the peptide of any one of claims 40 to 53, wherein X⁵ is R in said peptide, under conditions suitable for CD8 T lymphocyte expansion; or

(ii) if said subject is the subject of (b), culturing CD8 T lymphocytes from a second subject comprising an A at a position corresponding to nucleotide 596 in the nucleic acid sequence of human ZWINT and/or a ZWINT polypeptide comprising an arginine residue at a position corresponding to residue 187 in the protein sequence of human ZWINT in the presence of cells expressing a MHC class I molecule of the HLA-B*4403 allele loaded with the peptide of any one of claims 40 to 53, wherein X⁵ is G in said peptide, under conditions suitable for CD8 T lymphocyte expansion.

62. The method of any one of claims 57 to 61, wherein said subject is an allogeneic stem cell transplantation (ASCT) recipient.

63. A method of expanding CD8 T lymphocytes for adoptive T-cell immunotherapy, said method comprising:

(a) determining whether a candidate donor expresses a ZWINT nucleic acid comprising an A or a G at a position corresponding to nucleotide 596 in the nucleic acid sequence of human ZWINT (FIG. 2A, NCBI Reference Sequence: NM_007057.3), and/or a ZWINT polypeptide comprising a arginine or glycine residue at a position corresponding to residue 187 in the protein sequence of human ZWINT (FIG. 2B, NCBI Reference Sequence: NP_008988.2) and

(b)(i) if said candidate donor expresses a ZWINT nucleic acid comprising an A at a position corresponding to nucleotide 596 in the nucleic acid sequence of human ZWINT and/or a ZWINT polypeptide comprising an arginine residue at a position corresponding to residue 187 in the protein sequence of human ZWINT, culturing CD8 T lymphocytes from said candidate donor in the presence of cells expressing a MHC class I molecule of the HLA-B*4403 allele loaded with the peptide of any one of claims 40 to 53, wherein X⁵ is G in said peptide, under conditions suitable for CD8 T lymphocyte expansion; or

(b)(ii) if said candidate donor expresses a ZWINT nucleic acid comprising a G at a position corresponding to nucleotide 596 in the nucleic acid sequence of human ZWINT and/or a ZWINT polypeptide comprising a glycine residue at a position corresponding to residue 187 in the protein sequence of human ZWINT, culturing CD8 T lymphocytes from said candidate donor in the presence of cells expressing a MHC class I molecule of the HLA-B*4403 allele loaded with the peptide of any one of claims 40 to 53, wherein X⁵ is R in said peptide, under conditions suitable for CD8 T lymphocyte expansion.

64. A peptide of 50 amino acids or less comprising the sequence



wherein

Z¹ is an amino terminal modifying group or is absent;

X⁸ is a sequence of 1 to 43 amino acids or is absent;

X⁹ is P or A;

X¹⁰ is a sequence of 1 to 43 amino acids or is absent; and

Z² is a carboxy terminal modifying group or is absent.

65. The peptide of claim 64, wherein said peptide has a length of 8 to 12 amino acids.

66. The peptide of claim 64 or 65, wherein X⁸ is serine (S).

67. The peptide of any one of claims 64 to 66, wherein X¹⁰ is an amino acid.

68. The peptide of claim 67, wherein X¹⁰ is an aromatic amino acid.

69. The peptide of claim 68, wherein X¹⁰ is phenylalanine (F).

70. The peptide of claim any one of claims 64 to 69, wherein Z^1 is absent.
71. The peptide of any one of claims 64 to 70, wherein Z^2 is absent.
72. The peptide of any one of claims 64 to 71, wherein X^9 is P.
73. The peptide of any one of claims 64 to 71, wherein X^9 is A.
- 5 74. The peptide of claim 50, wherein said peptide is SLFFRKVPF (SEQ ID NO:19).
75. The peptide of claim 51, wherein said peptide is SLFFRKVAF (SEQ ID NO:20).
76. A nucleic acid encoding the peptide of any one of claims 64 to 75.
77. An isolated major histocompatibility complex (MHC) class I molecule of the HLA-B*0801 allele loaded with the peptide of any one of claims 64 to 75.
- 10 78. An isolated cell expressing at its surface a MHC class I molecule of the HLA-B*0801 allele loaded with the peptide of any one of claims 64 to 75.
79. A method of treating cancer, said method comprising administering to a subject in need thereof an effective amount of CD8 T lymphocytes recognizing a MHC class I molecule of the HLA-B*0801 allele loaded with the peptide of any one of claims 64 to 75.
- 15 80. The method of claim 79, said method further comprising determining whether said subject expresses a MTCH2 nucleic acid comprising a C or a G at a position corresponding to nucleotide 1057 in the nucleic acid sequence of human MTCH2 (FIG. 3A, NCBI Reference Sequence: NM_014342.3), and/or a MTCH2 polypeptide comprising a proline or alanine residue at a position corresponding to residue 290 in the protein sequence of human MTCH2 (FIG. 3B,
- 20 NCBI Reference Sequence: NP_055157.1), wherein
- (a) if said subject expresses a MTCH2 nucleic acid comprising a C at a position corresponding to nucleotide 1057 in the nucleic acid sequence of human MTCH2 and/or a MTCH2 polypeptide comprising a proline residue at a position corresponding to residue 290 in the protein sequence of human MTCH2, X^9 is P in said peptide;
- 25 (b) if said subject expresses a MTCH2 nucleic acid comprising a G at a position corresponding to nucleotide 1057 in the nucleic acid sequence of human MTCH2 and/or a MTCH2 polypeptide comprising an alanine residue at a position corresponding to residue 290 in the protein sequence of human MTCH2, X^5 is A in said peptide.
81. The method of claim 80, wherein said determining comprises sequencing a human
- 30 MTCH2 nucleic acid.

82. The method of claim 79, wherein said CD8 T lymphocytes are *in vitro* expanded CD8 T lymphocytes.

83. The method of any one of claims 80 to 82, wherein said method further comprises:

(i) if said subject is the subject of (a), culturing CD8 T lymphocytes from a second
5 subject comprising a G at a position corresponding to nucleotide 1057 in the nucleic acid sequence of human MTCH2 and/or a MTCH2 polypeptide comprising an alanine residue at a position corresponding to residue 290 in the protein sequence of human MTCH2 in the presence of cells expressing a MHC class I molecule of the HLA-B*0801 allele loaded with the peptide of any one of claims 64 to 75, wherein X⁹ is P in said peptide, under conditions suitable
10 for CD8 T lymphocyte expansion; or

(ii) if said subject is the subject of (b), culturing CD8 T lymphocytes from a second
subject comprising a C at a position corresponding to nucleotide 1057 in the nucleic acid sequence of human MTCH2 and/or a MTCH2 polypeptide comprising a proline residue at a position corresponding to residue 290 in the protein sequence of human MTCH2 in the
15 presence of cells expressing a MHC class I molecule of the HLA-B*0801 allele loaded with the peptide of any one of claims 64 to 75, wherein X⁹ is A in said peptide, under conditions suitable for CD8 T lymphocyte expansion.

84. The method of any one of claims 79 to 83, wherein said subject is an allogeneic stem cell transplantation (ASCT) recipient.

20 85. A method of expanding CD8 T lymphocytes for adoptive T-cell immunotherapy, said method comprising:

(a) determining whether a candidate donor expresses a MTCH2 nucleic acid comprising a C or a G at a position corresponding to nucleotide 1057 in the nucleic acid sequence of human MTCH2 (FIG. 3A, NCBI Reference Sequence: NM_014342.3), and/or a
25 MTCH2 polypeptide comprising a proline or alanine residue at a position corresponding to residue 290 in the protein sequence of human MTCH2 (FIG. 3B, NCBI Reference Sequence: NP_055157.1) and

(b)(i) if said candidate donor expresses a MTCH2 nucleic acid comprising a C at a position corresponding to nucleotide 1057 in the nucleic acid sequence of human MTCH2
30 and/or a MTCH2 polypeptide comprising a proline residue at a position corresponding to residue 290 in the protein sequence of human MTCH2, culturing CD8 T lymphocytes from said candidate donor in the presence of cells expressing a MHC class I molecule of the HLA-B*0801 allele loaded with the peptide of any one of claims 64 to 75, wherein X⁹ is A in said peptide, under conditions suitable for CD8 T lymphocyte expansion; or

(b)(ii) if said candidate donor expresses a MTCH2 nucleic acid comprising a G at a position corresponding to nucleotide 1057 in the nucleic acid sequence of human MTCH2 and/or a MTCH2 polypeptide comprising an alanine residue at a position corresponding to residue 290 in the protein sequence of human MTCH2, culturing CD8 T lymphocytes from said candidate donor in the presence of cells expressing a MHC class I molecule of the HLA-B*0801 allele loaded with the peptide of any one of claims 64 to 75, wherein X⁹ is P in said peptide, under conditions suitable for CD8 T lymphocyte expansion.

86. A peptide of 50 amino acids or less comprising the sequence



wherein

Z¹ is an amino terminal modifying group or is absent;

X¹¹ is a sequence of 1 to 43 amino acids or is absent;

X¹² is S or T;

X¹³ is an amino acid or is absent;

X¹⁴ is a sequence of 1 to 43 amino acids or is absent; and

Z² is a carboxy terminal modifying group or is absent.

87. The peptide of claim 86, wherein said peptide has a length of 8 to 12 amino acids.

88. The peptide of claim 86 or 87, wherein X¹¹ is absent.

89. The peptide of any one of claims 86 to 88, wherein X¹³ is an amino acid.

90. The peptide of claim 89, wherein X¹³ is serine.

91. The peptide of any one of claims 86 to 90, wherein X¹⁴ is an amino acid.

92. The peptide of claim 91, wherein X¹⁴ is a basic amino acid.

93. The peptide of claim 92, wherein X¹⁴ is lysine (K).

94. The peptide of any one of claims 86 to 93, wherein Z¹ is absent.

95. The peptide of any one of claims 86 to 94, wherein Z² is absent.

96. The peptide of any one of claims 86 to 95, wherein X¹² is S.

97. The peptide of any one of claims 86 to 95, wherein X¹² is T.

98. The peptide of claim 96, wherein said peptide is SVLKPGNSK (SEQ ID NO:21).

99. The peptide of claim 97, wherein said peptide is TVLKPGNSK (SEQ ID NO:22).

100. A nucleic acid encoding the peptide of any one of claims 86 to 99.

101. An isolated major histocompatibility complex (MHC) class I molecule of the HLA-A*0301 allele loaded with the peptide of any one of claims 86 to 99.

5 102. An isolated cell expressing at its surface a MHC class I molecule of the HLA-A*0301 allele loaded with the peptide of any one of claims 86 to 99.

103. A method of treating cancer, said method comprising administering to a subject in need thereof an effective amount of CD8 T lymphocytes recognizing a MHC class I molecule of the HLA-A*0301 allele loaded with the peptide of any one of claims 86 to 99.

10 104. The method of claim 103, said method further comprising determining whether said subject expresses an ELF1 nucleic acid comprising an A or T at a position corresponding to nucleotide 1400 in the nucleic acid sequence of human ELF1 (FIGs. 4A and 4B, NCBI Reference Sequence: NM_172373.3) and/or an ELF1 polypeptide having a threonine or a serine at a position corresponding to residue 343 in the ELF1 protein sequence (FIG. 4C, NCBI
15 Reference Sequence: NP_758961.1), wherein

(a) if said subject expresses an ELF1 nucleic acid comprising an A at a position corresponding to nucleotide 1400 in the nucleic acid sequence of human ELF1 and/or an ELF1 polypeptide comprising a threonine residue at a position corresponding to residue 343 in the protein sequence of human ELF1, X¹² is T in said peptide;

20 (b) if said subject expresses an ELF1 nucleic acid comprising a T at a position corresponding to nucleotide 1400 in the nucleic acid sequence of human ELF1 and/or an ELF1 polypeptide comprising a serine residue at a position corresponding to residue 343 in the protein sequence of human ELF1, X¹² is S in said peptide.

105. The method of claim 104, wherein said determining comprises sequencing a human
25 ELF1 nucleic acid.

106. The method of claim 103, wherein said CD8 T lymphocytes are *in vitro* expanded CD8 T lymphocytes.

107. The method of any one of claims 104 to 106, wherein said method further comprises:

(i) if said subject is the subject of (a), culturing CD8 T lymphocytes from a second
30 subject comprising a T at a position corresponding to nucleotide 1400 in the nucleic acid sequence of human ELF1 and/or an ELF1 polypeptide comprising a serine residue at a position corresponding to residue 343 in the protein sequence of human ELF1 in the presence of cells

expressing a MHC class I molecule of the HLA-A*0301 allele loaded with the peptide of any one of claims 86 to 99, wherein X^{12} is T in said peptide, under conditions suitable for CD8 T lymphocyte expansion; or

(ii) if said subject is the subject of (b), culturing CD8 T lymphocytes from a second subject comprising an A at a position corresponding to nucleotide 1400 in the nucleic acid sequence of human ELF1 and/or an ELF1 polypeptide comprising a threonine residue at a position corresponding to residue 343 in the protein sequence of human ELF1 in the presence of cells expressing a MHC class I molecule of the HLA-A*0301 allele loaded with the peptide of any one of claims 86 to 99, wherein X^{12} is S in said peptide, under conditions suitable for CD8 T lymphocyte expansion.

108. The method of any one of claims 103 to 107, wherein said subject is an allogeneic stem cell transplantation (ASCT) recipient.

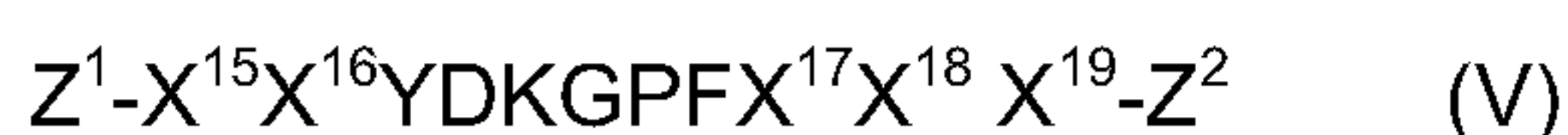
109. A method of expanding CD8 T lymphocytes for adoptive T-cell immunotherapy, said method comprising:

(a) determining whether a candidate donor expresses an ELF1 nucleic acid comprising an A or T at a position corresponding to nucleotide 1400 in the nucleic acid sequence of human ELF1 (FIGs. 4A and 4B, NCBI Reference Sequence: NM_172373.3) and/or an ELF1 polypeptide having a threonine or a serine at a position corresponding to residue 343 in the ELF1 protein sequence (FIG. 4C, NCBI Reference Sequence: NP_758961.1) and

(b)(i) if said candidate donor expresses an ELF1 nucleic acid comprising an A at a position corresponding to nucleotide 1400 in the nucleic acid sequence of human ELF1 and/or an ELF1 polypeptide having a threonine at a position corresponding to residue 343 in the ELF1 protein sequence, culturing CD8 T lymphocytes from said candidate donor in the presence of cells expressing a MHC class I molecule of the HLA-A*0301 allele loaded with the peptide of any one of claims 86 to 99, wherein X^{12} is S in said peptide, under conditions suitable for CD8 T lymphocyte expansion; or

(b)(ii) if said candidate donor expresses an ELF1 nucleic acid comprising a T at a position corresponding to nucleotide 1400 in the nucleic acid sequence of human ELF1 and/or an ELF1 polypeptide having a serine at a position corresponding to residue 343 in the ELF1 protein sequence, culturing CD8 T lymphocytes from said candidate donor in the presence of cells expressing a MHC class I molecule of the HLA-A*0301 allele loaded with the peptide of any one of claims 86 to 99, wherein X^{12} is T in said peptide, under conditions suitable for CD8 T lymphocyte expansion.

110. A peptide of 50 amino acids or less comprising the sequence (V)



wherein

Z^1 is an amino terminal modifying group or is absent;

X^{15} is a sequence of 1 to 43 amino acids or is absent;

X^{16} is an amino acid or is absent;

5 X^{17} is R or W;

X^{18} is an amino acid or is absent;

X^{19} is a sequence of 1 to 43 amino acids or is absent;

Z^2 is a carboxy terminal modifying group or is absent.

111. The peptide of claim 110, wherein said peptide has a length of 8 to 12 amino acids.

10 112. The peptide of claim 110 or 111, wherein X^{16} is an amino acid.

113. The peptide of claim 112, wherein X^{16} is a methionine (M).

114. The peptide of any one of claims 110 to 113, wherein X^{15} is an amino acid.

115. The peptide of claim 114, wherein X^{15} is an alanine.

116. The peptide of any one of claims 110 to 115, wherein X^{18} is an amino acid.

15 117. The peptide of claim 116, wherein X^{18} is serine (S).

118. The peptide of any one of claims 110 to 117, wherein X^{19} is an amino acid

119. The peptide of claim 118, wherein X^{19} is a basic amino acid

120. The peptide of claim 119, wherein X^{19} is lysine (K).

121. The peptide of any one of claims 110 to 120, wherein Z^1 is absent.

20 122. The peptide of any one of claims 110 to 121, wherein Z^2 is absent.

123. The peptide of any one of claims 110 to 122, wherein X^{17} is R.

124. The peptide of any one of claims 86 to 95, wherein X^{17} is W.

125. The peptide of claim 123, wherein said peptide is AMYDKGPFRSK (SEQ ID NO:23).

126. The peptide of claim 124, wherein said peptide is AMYDKGPFWSK (SEQ ID NO:24).

25 127. A nucleic acid encoding the peptide of any one of claims 110 to 126.

128. An isolated major histocompatibility complex (MHC) class I molecule of the HLA-A*0301 allele loaded with the peptide of any one of claims 110 to 126.

129. An isolated cell expressing at its surface a MHC class I molecule of the HLA-A*0301 allele loaded with the peptide of any one of claims 110 to 126.

5 130. A method of treating cancer, said method comprising administering to a subject in need thereof an effective amount of CD8 T lymphocytes recognizing a MHC class I molecule of the HLA-A*0301 allele loaded with the peptide of any one of claims 110 to 126.

131. The method of claim 130, said method further comprising determining whether said subject expresses an NQO1 nucleic acid comprising an C or T at a position corresponding to
10 nucleotide 615 in the nucleic acid sequence of human NQO1 (FIG. 5A, NCBI Reference Sequence: NM_000903.2) and/or an NQO1 polypeptide having an arginine or a tryptophan at a position corresponding to residue 139 in the NQO1 protein sequence (FIG. 5B, NCBI Reference Sequence: NP_000894.1), wherein

(a) if said subject expresses a NQO1 nucleic acid comprising an C at a position
15 corresponding to nucleotide 615 in the nucleic acid sequence of human NQO1 and/or a NQO1 polypeptide comprising an arginine residue at a position corresponding to residue 139 in the protein sequence of human NQO1, X¹⁷ is R in said peptide;

(b) if said subject expresses an NQO1 nucleic acid comprising a T at a position
20 corresponding to nucleotide 615 in the nucleic acid sequence of human NQO1 and/or a NQO1 polypeptide comprising a tryptophan residue at a position corresponding to residue 139 in the protein sequence of human NQO1, X¹⁷ is W in said peptide.

132. The method of claim 131, wherein said determining comprises sequencing a human NQO1 nucleic acid.

133. The method of claim 131 or 132, wherein said CD8 T lymphocytes are *in vitro* expanded
25 CD8 T lymphocytes.

134. The method of any one of claims 131 to 133, wherein said method further comprises:

(i) if said subject is the subject of (a), culturing CD8 T lymphocytes from a second subject comprising a T at a position corresponding to nucleotide 615 in the nucleic acid sequence of human NQO1 and/or a NQO1 polypeptide comprising a tryptophan residue at a
30 position corresponding to residue 139 in the protein sequence of human NQO1 in the presence of cells expressing a MHC class I molecule of the HLA-A*0301 allele loaded with the peptide of any one of claims 110 to 126, wherein X¹⁷ is R in said peptide, under conditions suitable for CD8 T lymphocyte expansion; or

(ii) if said subject is the subject of (b), culturing CD8 T lymphocytes from a second subject comprising a C at a position corresponding to nucleotide 615 in the nucleic acid sequence of human NQO1 and/or a NQO1 polypeptide comprising an arginine residue at a position corresponding to residue 139 in the protein sequence of human NQO1 in the presence
 5 of cells expressing a MHC class I molecule of the HLA-A*0301 allele loaded with the peptide of any one of claims 110 to 126, wherein X^{17} is W in said peptide, under conditions suitable for CD8 T lymphocyte expansion.

135. The method of any one of claims 130 to 134, wherein said subject is an allogeneic stem cell transplantation (ASCT) recipient.

10 136. A method of expanding CD8 T lymphocytes for adoptive T-cell immunotherapy, said method comprising:

(a) determining whether said subject expresses an NQO1 nucleic acid comprising an C or T at a position corresponding to nucleotide 615 in the nucleic acid sequence of human NQO1 (FIG. 5A, NCBI Reference Sequence: NM_000903.2) and/or an NQO1 polypeptide
 15 having an arginine or a tryptophan at a position corresponding to residue 139 in the NQO1 protein sequence (FIG. 5B, NCBI Reference Sequence: NP_000894.1) and

(b)(i) if said candidate donor expresses a NQO1 nucleic acid comprising an C at a position corresponding to nucleotide 615 in the nucleic acid sequence of human NQO1 and/or a NQO1 polypeptide comprising an arginine residue at a position corresponding to residue 139 in
 20 the protein sequence of human NQO1, culturing CD8 T lymphocytes from said candidate donor in the presence of cells expressing a MHC class I molecule of the HLA-A*0301 allele loaded with the peptide of any one of claims 110 to 126, wherein X^{17} is W in said peptide, under conditions suitable for CD8 T lymphocyte expansion; or

(b)(ii) if said candidate donor expresses a NQO1 nucleic acid comprising a T at a
 25 position corresponding to nucleotide 615 in the nucleic acid sequence of human NQO1 and/or a NQO1 polypeptide comprising a tryptophan residue at a position corresponding to residue 139 in the protein sequence of human NQO1, culturing CD8 T lymphocytes from said candidate donor in the presence of cells expressing a MHC class I molecule of the HLA-A*0301 allele loaded with the peptide of any one of claims 110 to 126, wherein X^{17} is R in said peptide, under
 30 conditions suitable for CD8 T lymphocyte expansion.

137. A peptide of 50 amino acids or less comprising the sequence (VI)



wherein

Z^1 is an amino terminal modifying group or is absent;

35 X^{20} is a sequence of 1 to 43 amino acids or is absent;

X^{21} is an amino acid or is absent;

X^{22} is an amino acid or is absent;

X^{23} is G or R;

X^{24} is a sequence of 1 to 43 amino acids or is absent;

5 Z^2 is a carboxy terminal modifying group or is absent.

138. The peptide of claim 137, wherein said peptide has a length of 8 to 12 amino acids.

139. The peptide of claim 137 or 138, wherein X^{22} is an amino acid.

140. The peptide of claim 139, wherein X^{22} is valine (V).

141. The peptide of any one of claims 137 to 140, wherein X^{21} is an amino acid.

10 142. The peptide of claim 141, wherein X^{21} is arginine (R).

143. The peptide of any one of claims 137 to 142, wherein Z^1 is absent.

144. The peptide of any one of claims 137 to 143, wherein Z^2 is absent.

145. The peptide of any one of claims 137 to 144, wherein X^{23} is G.

146. The peptide of any one of claims 137 to 144, wherein X^{23} is R.

15 147. The peptide of claim 145, wherein said peptide is RVSLPTSPG (SEQ ID NO:25).

148. The peptide of claim 146, wherein said peptide is RVSLPTSPR (SEQ ID NO:26).

149. A nucleic acid encoding the peptide of any one of claims 137 to 148.

150. An isolated major histocompatibility complex (MHC) class I molecule of the HLA-A*0301 allele loaded with the peptide of any one of claims 137 to 148.

20 151. An isolated cell expressing at its surface a MHC class I molecule of the HLA-A*0301 allele loaded with the peptide of any one of claims 137 to 148.

152. A method of treating cancer, said method comprising administering to a subject in need thereof an effective amount of CD8 T lymphocytes recognizing a MHC class I molecule of the HLA-A*0301 allele loaded with the peptide of any one of claims 137 to 148.

25 153. The method of claim 152, said method further comprising determining whether said subject expresses a KIAA0226L nucleic acid comprising a G or A at a position corresponding to nucleotide 1059 in the nucleic acid sequence of human KIAA0226L (FIGs. 6A and 6B, NCBI

Reference Sequence: NM_025113.2) and/or an KIAA0226L polypeptide having a glycine or an arginine at a position corresponding to residue 152 in the KIAA0226L protein sequence (FIG. 6C, NCBI Reference Sequence: NP_079389.2), wherein

(a) if said subject expresses a KIAA0226L nucleic acid comprising a G at a position
5 corresponding to nucleotide 1059 in the nucleic acid sequence of human KIAA0226L and/or a KIAA0226L polypeptide comprising a glycine residue at a position corresponding to residue 152 in the protein sequence of human KIAA0226L, X²³ is G in said peptide;

(b) if said subject expresses a KIAA0226L nucleic acid comprising an A at a position
10 corresponding to nucleotide 1059 in the nucleic acid sequence of human a KIAA0226L and/or a KIAA0226L polypeptide comprising an arginine residue at a position corresponding to residue 152 in the protein sequence of human KIAA0226L, X²³ is R in said peptide.

154. The method of claim 153, wherein said determining comprises sequencing a human KIAA0226L nucleic acid.

155. The method of claim 153 or 154, wherein said CD8 T lymphocytes are *in vitro* expanded
15 CD8 T lymphocytes.

156. The method of any one of claims 153 to 155, wherein said method further comprises:

(i) if said subject is the subject of (a), culturing CD8 T lymphocytes from a second
subject comprising an A at a position corresponding to nucleotide to nucleotide 1059 in the
nucleic acid sequence of human KIAA0226L and/or a KIAA0226L polypeptide comprising an
20 arginine residue at a position corresponding to residue 152 in the protein sequence of human
KIAA0226L in the presence of cells expressing a MHC class I molecule of the HLA-A*0301
allele loaded with the peptide of any one of claims 137 to 148, wherein X²³ is G in said peptide,
under conditions suitable for CD8 T lymphocyte expansion; or

(ii) if said subject is the subject of (b), culturing CD8 T lymphocytes from a second
25 subject comprising a G at a position corresponding to nucleotide 1059 in the nucleic acid
sequence of human KIAA0226L and/or a KIAA0226L polypeptide comprising a glycine residue
at a position corresponding to residue 152 in the protein sequence of human KIAA0226L in the
presence of cells expressing a MHC class I molecule of the HLA-A*0301 allele loaded with the
peptide of any one of claims 137 to 148, wherein X²³ is R in said peptide, under conditions
30 suitable for CD8 T lymphocyte expansion.

157. The method of any one of claims 152 to 156, wherein said subject is an allogeneic stem
cell transplantation (ASCT) recipient.

158. A method of expanding CD8 T lymphocytes for adoptive T-cell immunotherapy, said
method comprising:

(a) determining whether a candidate donor expresses a KIAA0226L nucleic acid comprising a G or A at a position corresponding to nucleotide 1059 in the nucleic acid sequence of human KIAA0226L (FIGs. 6A and 6B, NCBI Reference Sequence: NM_025113.2) and/or an KIAA0226L polypeptide having a glycine or an arginine at a position corresponding to residue
 5 152 in the KIAA0226L protein sequence (FIG. 6C, NCBI Reference Sequence: NP_079389.2) and

(b)(i) if said candidate donor expresses a KIAA0226L nucleic acid comprising G at a position corresponding to nucleotide 1059 in the nucleic acid sequence of human KIAA0226L and/or a KIAA0226L polypeptide comprising a glycine residue at a position corresponding to
 10 residue 152 in the protein sequence of human KIAA0226L, culturing CD8 T lymphocytes from said candidate donor in the presence of cells expressing a MHC class I molecule of the HLA-A*0301 allele loaded with the peptide of any one of claims 137 to 148, wherein X^{23} is R in said peptide, under conditions suitable for CD8 T lymphocyte expansion; or

(b)(ii) if said subject expresses a KIAA0226L nucleic acid comprising an A at a position
 15 corresponding to nucleotide 1059 in the nucleic acid sequence of human a KIAA0226L and/or a KIAA0226L polypeptide comprising an arginine residue at a position corresponding to residue 152 in the protein sequence of human KIAA0226L, culturing CD8 T lymphocytes from said candidate donor in the presence of cells expressing a MHC class I molecule of the HLA-A*0301 allele loaded with the peptide of any one of claims 137 to 148, wherein X^{23} is G in said peptide,
 20 under conditions suitable for CD8 T lymphocyte expansion.

159. A peptide of 50 amino acids or less comprising the sequence (VII)



wherein

Z^1 is an amino terminal modifying group or is absent;
 25 X^{25} is a sequence of 1 to 43 amino acids or is absent;
 X^{26} is an amino acid or is absent;
 X^{27} is an amino acid or is absent;
 X^{28} is K or N;
 X^{29} is a sequence of 1 to 43 amino acids or is absent;
 30 Z^2 is a carboxy terminal modifying group or is absent.

160. The peptide of claim 159, wherein said peptide has a length of 8 to 12 amino acids.

161. The peptide of claim 159 or 160, wherein X^{27} is an amino acid.

162. The peptide of claim 161, wherein X^{27} is methionine (M).

163. The peptide of any one of claims 159 to 162, wherein X^{26} is an amino acid.

164. The peptide of claim 141, wherein X^{26} is valine (V).
165. The peptide of any one of claims 159 to 163, wherein Z^1 is absent.
166. The peptide of any one of claims 159 to 164, wherein Z^2 is absent.
167. The peptide of any one of claims 159 to 166, wherein X^{28} is K.
- 5 168. The peptide of any one of claims 159 to 166, wherein X^{28} is N.
169. The peptide of claim 167, wherein said peptide is VMGNPGTFK (SEQ ID NO:27).
170. The peptide of claim 168, wherein said peptide is VMGNPGTFN (SEQ ID NO:28).
171. A nucleic acid encoding the peptide of any one of claims 159 to 170.
172. An isolated major histocompatibility complex (MHC) class I molecule of the HLA-A*0301
10 allele loaded with the peptide of any one of claims 159 to 170.
173. An isolated cell expressing at its surface a MHC class I molecule of the HLA-A*0301 allele loaded with the peptide of any one of claims 159 to 170.
174. A method of treating cancer, said method comprising administering to a subject in need thereof an effective amount of CD8 T lymphocytes recognizing a MHC class I molecule of the
15 HLA-A*0301 allele loaded with the peptide of any one of claims 159 to 170.
175. The method of claim 174, said method further comprising determining whether said subject expresses an RMDN1 nucleic acid comprising an A or C at a position corresponding to nucleotide 316 in the nucleic acid sequence of human RMDN1 (FIG. 7A, NCBI Reference Sequence: NM_016033.2) and/or an RMDN1 polypeptide having a lysine or an asparagine at a
20 position corresponding to residue 52 in the RMDN1 protein sequence (FIG. 7B, NCBI Reference Sequence: NP_057117.2), wherein
- (a) if said subject expresses an RMDN1 nucleic acid comprising an A at a position corresponding to nucleotide 316 in the nucleic acid sequence of human RMDN1 and/or an RMDN1 polypeptide comprising a lysine residue at a position corresponding to residue 52 in the
25 protein sequence of human RMDN1, X^{28} is K in said peptide;
- (b) if said subject expresses an RMDN1 nucleic acid comprising a C at a position corresponding to nucleotide 316 in the nucleic acid sequence of human RMDN1 and/or a RMDN1 polypeptide comprising an asparagine residue at a position corresponding to residue 52 in the protein sequence of human RMDN1, X^{28} is N in said peptide.

176. The method of claim 175, wherein said determining comprises sequencing a human KIAA0226L nucleic acid.

177. The method of any one of claims 174 to 176, wherein said CD8 T lymphocytes are *in vitro* expanded CD8 T lymphocytes.

5 178. The method of any one of claims 175 to 177, wherein said method further comprises:

(i) if said subject is the subject of (a), culturing CD8 T lymphocytes from a second subject comprising a C at a position corresponding to nucleotide 316 in the nucleic acid sequence of human RMDN1 and/or a RMDN1 polypeptide comprising an asparagine residue at a position corresponding to residue 52 in the protein sequence of human RMDN1 in the presence of cells expressing a MHC class I molecule of the HLA-A*0301 allele loaded with the peptide of any one of claims 159 to 170, wherein X²⁸ is K in said peptide, under conditions suitable for CD8 T lymphocyte expansion; or

(ii) if said subject is the subject of (b), culturing CD8 T lymphocytes from a second subject comprising an A at a position corresponding to nucleotide 316 in the nucleic acid sequence of human RMDN1 and/or a RMDN1 polypeptide comprising a lysine residue at a position corresponding to residue 52 in the protein sequence of human RMDN1 in the presence of cells expressing a MHC class I molecule of the HLA-A*0301 allele loaded with the peptide of any one of claims 159 to 170, wherein X²⁸ is N in said peptide, under conditions suitable for CD8 T lymphocyte expansion.

20 179. The method of any one of claims 174 to 178, wherein said subject is an allogeneic stem cell transplantation (ASCT) recipient.

180. A method of expanding CD8 T lymphocytes for adoptive T-cell immunotherapy, said method comprising:

(a) determining whether a candidate donor expresses an RMDN1 nucleic acid comprising an A or C at a position corresponding to nucleotide 316 in the nucleic acid sequence of human RMDN1 (FIG. 7A, NCBI Reference Sequence: NM_016033.2) and/or an RMDN1 polypeptide having a lysine or an asparagine at a position corresponding to residue 52 in the RMDN1 protein sequence (FIG. 7B, NCBI Reference Sequence: NP_057117.2), and

(b)(i) if said subject expresses an RMDN1 nucleic acid comprising an A at a position corresponding to nucleotide 316 in the nucleic acid sequence of human RMDN1 and/or an RMDN1 polypeptide comprising a lysine residue at a position corresponding to residue 52 in the protein sequence of human RMDN1, culturing CD8 T lymphocytes from said candidate donor in the presence of cells expressing a MHC class I molecule of the HLA-A*0301 allele loaded with

the peptide of any one of claims 159 to 170, wherein X^{28} is N in said peptide, under conditions suitable for CD8 T lymphocyte expansion; or

(b)(ii) if said subject expresses an RMDN1 nucleic acid comprising a C at a position corresponding to nucleotide 316 in the nucleic acid sequence of human RMDN1 and/or a
5 RMDN1 polypeptide comprising an asparagine residue at a position corresponding to residue 52 in the protein sequence of human RMDN1, culturing CD8 T lymphocytes from said candidate donor in the presence of cells expressing a MHC class I molecule of the HLA-A*0301 allele loaded with the peptide of any one of claims 159 to 170, wherein X^{28} is R in said peptide, under conditions suitable for CD8 T lymphocyte expansion.

10

15

1/11

```

1 gagaccagaa gcgggcgaat tgggcaccgg tggcggtgc gggcagtttg aattagactc
61 tgggctccag cccgccgaag ccgcgccaga actgtactct ccgagaggtc gttttcccg
121 ccccgagagc aagtttattht acaaatgttg gagtaataaa gaaggcagaa caaaatgagc
181 tgggcttttg aagaatggaa agaagggtcg cctacaagag ctcttcagaa aattcaagag
241 cttgaaggac agcttgacaa actgaagaag gaaaagcagc aaaggcagtt tcagcttgac
301 agtctcgagg ctgcgctgca gaagcaaaaa cagaagggtt aaaatgaaaa aaccgagggg
361 acaaacctga aaaggggagaa tcaaagattg atggaaatat gtgaaagtct ggagaaaact
421 aagcagaaga tttctcatga acttcaagtc aaggagtcac aagtgaattt ccaggaagga
481 caactgaatt caggcaaaaa acaaatagaa aaactggaac aggaacttaa aagggtgtaaa
541 tctgagcttg aaagaagcca acaagctgcg cagtctgcag atgtctctct gaatccatgc
601 aatacaccac aaaaaatttt tacaactcca ctaacaccaa gtcaatatta tagtggttcc
661 aagtatgaag atctaaaaga aaaatataat aaagagggtt aagaacgaaa aagattagag
721 gcagagggtta aagccttgca ggctaaaaaa gcaagccaga ctcttcacaa agccaccatg
781 aatcacccgc acattgcccg gcatcaggct tcatcatctg tgttctcatg gcagcaagag
841 aagaccccaa gtcatctttc atctaattct caaagaactc caattaggag agatttctct
901 gcatcttact tttctgggga acaagagggt actccaagtc gatcaacttt gcaaataagg
961 aaaagagatg ctaatagcag tttctttgac aattctagca gtctcatct tttggatcaa
1021 ttaaaagcgc agaatacaaga gctaagaaac aagattaatg agttggaact acgcctgcaa
1081 ggacatgaaa aagaaatgaa aggccaaagt aataagtttc aagaactcca actccaactg
1141 gagaaagcaa aagtggaatt aattgaaaaa gagaaagttt tgaacaaatg tagggatgaa
1201 ctagtgagaa caacagcaca atacgaccag gcgtcaacca agtatactgc attggaacaa
1261 aaactgaaaa aattgacgga agatttgagt tgtcagcgac aaaatgcaga aagtgccaga
1321 tgttctctgg aacagaaaaa taaggaaaaa gaaaaggagt ttcaagagga gctctcccg
1381 caacagcggt ctttccaaac actggaccag gagtgcattc agatgaaggc cagactcacc
1441 caggagttae agcaagccaa gaatatgcac aacgtcctgc aggtgaaact ggataaaact
1501 acatcagtaa agcaacagct agaaaaaat ttggaagagt ttaagcaaaa gttgtgcaga
1561 gctgaacagg cgttccaggc gagtcaagtc aaggagaatg agctgaggag aagcatggag
1621 gaaatgaaga aggaaaacaa cctccttaag agtcaactct agcaaaaagg cagagaagtc
1681 tgccacctgg aggcagaact caagaacatc aaacagtggt taaatcagag ccagaatttt
1741 gcagaagaaa tgaaagcgaa gaatacctct caggaaacca tgttaagaga tcttcaagaa
1801 aaaataaatc agcaagaaaa ctctttgact ttagaaaaac tgaagcttgc tgtggctgat
1861 ctggaaaagc agcgagattg ttctcaagac cttttgaaga aaagagaaca tcacattgaa
1921 caacttaatg ataagttaag caagacagag aaagagtcca aagccttgct gagtgcctta
1981 gagttaaaaa agaaagaata tgaagaattg aaagaagaga aaactctggt ttcttggtgg
2041 aaaagtgaia acgaaaaact tttaactcag atggaatcag aaaaggaaaa cttgcagagt
2101 aaaattaatc acttggaaac ttgtctgaag acacagcaaa taaaaagtca tgaatacaac
2161 gagagagtaa gaacgctgga gatggacaga gaaaacctaa gtgtcgagat cagaaacctt
2221 cacaacgtgt tagacagtaa gtcaagtggag gtagagaccc agaaactagc ttatatggag
2281 ctacagcaga aagctgagtt ctcaagatcag aaacatcaga aggaataaga aaatatgtgt
2341 ttgaagactt ctcaagcttae tgggcaagtt gaagatctag aacacaagct tcagttactg
2401 tcaaatgaaa taatggacaa agaccggtgt taccaagact tgcatgccga atatgagagc
2461 ctcaagggatc tgctaaaatc caaagatgct tctctggtga caaatgaaga tcatcagaga
2521 agtctttttg cttttgatca gcagcctgcc atgcatcatt cttttgcaaa tataattgga
2581 gaacaaggaa gcatgccttc agagaggagt gaatgtcgtt tagaagcaga ccaaagtcag
2641 aaaaattctg ccatcctaca aaatagagtt gattcaactg aattttcatt agagtctcaa
2701 aaacagatga actcagacct gcaaaaagcag tgtgaagagt tgggtgcaaat caaaggagaa
2761 atagaagaaa atctcatgaa agcagaacag atgcatcaaa gttttgtggc tgaacaagat
2821 cagcgcatta gtaagttaca ggaagacact tctgctcacc agaattgtgt tgctgaaacc
2881 ttaagtgcc cttgagaacaa ggaaaaagag ctgcaacttt taaatgataa ggtagaaact
2941 gagcaggcag agattcaaga attaaaaaag agcaaccatc tacttgaaga ctctctaaag
3001 gagctacaac ttttatccga aacctaaagc ttggagaaga aagaaatgag ttccatcatt
3061 tctctaaata aaagggaat tgaagagctg acccaagaga atgggactct taaggaaatt
3121 aatgcatcct taaatcaaga gaagatgaac ttaatccaga aaagtgaag ttttgcaaac
3181 tatatagatg aaaggggagaa aagcatttca gagttatctg atcagtacaa gcaagaaaaa
3241 cttattttac tacaaagatg tgaagaaacc ggaaatgcat atgaggatct tagtcaaaaa
3301 tacaaagcag cacaggaaaa gaattctaaa ttagaatgct tgctaaatga atgcactagt
3361 ctttgtgaaa ataggaaaaa tgagttggaa cagctaaagg aagcatttgc aaaggaacac
3421 caagaattct taacaaaatt agcatttgct gaagaaagaa atcagaatct gatgctagag
3481 ttggagacag tgcagcaagc tctgagatct gagatgacag ataaccaaaa caattctaag

```

FIG. 1A

2/11

```

3541 agcgaggctg gtggttttaa gcaagaaatc atgactttta aggaagaaca aaacaaaatg
3601 caaaaggaag ttaatgactt attacaagag aatgaacagc tgatgaaggt aatgaagact
3661 aaacatgaat gtcaaaatct agaatcagaa ccaattagga actctgtgaa agaaagagag
3721 agtgagagaa atcaatgtaa ttttaaacct cagatggatc ttgaagttaa agaaatttct
3781 ctagatagtt ataatgcgca gttggtgcaa ttagaagcta tgctaagaaa taaggaatta
3841 aaacttcagg aaagtgagaa ggagaaggag tgcctgcagc atgaattaca gacaattaga
3901 ggagatcttg aaaccagcaa tttgcaagac atgcagtcac aagaaattag tggccttaaa
3961 gactgtgaaa tagatgcgga agaaaagtat atttcagggc ctcatgagtt gtcaacaagt
4021 caaaacgaca atgcacacct tcagtgtctc ctgcaaaaca caatgaacaa gctgaatgag
4081 ctagagaaaa tatgtgaaat actgcaggct gaaaagtatg aactcgtaac tgagctgaat
4141 gattcaaggt cagaatgtat cacagcaact aggaaaatgg cagaagaggt agggaaacta
4201 ctaaatgaag ttaaaatatt aaatgatgac agtgggtctc tccatgggtga gttagtggaa
4261 gacataccag gaggtgaatt tgggtgaacaa ccaaatgaac agcaccctgt gtctttggct
4321 ccattggacg agagtaattc ctacgagcac ttgacattgt cagacaaaga agttcaaatg
4381 cactttgccg aattgcaaga gaaattctta tctttacaaa gtgaacacaa aattttacat
4441 gatcagcact gtcagatgag ctctaaaatg tcagagctgc agacctatgt tgactcatta
4501 aaggccgaaa atttggtctt gtcaacgaat ctgagaaaact ttcaaggtga cttggtgaag
4561 gagatgcagc tgggcttgga ggaggggctc gttccatccc tgtcatcctc ttgtgtgcct
4621 gacagctcta gtcttagcag tttgggagac tcctcctttt acagagctct tttagaacag
4681 acaggagata tgtctctttt gagtaattta gaaggggctg tttcagcaaa ccagtgcagt
4741 gtagatgaag tattttgcag cagtctgcag gaggagaatc tgaccaggaa agaaacctct
4801 tcggccccag cgaaggggtg tgaagagctt gagtccctct gtgaggtgta ccggcagtc
4861 ctcgagaagc tagaagagaa aatggaaagt caagggatta tgaaaaataa ggaaattcaa
4921 gagctcgagc agttattaag ttctgaaagg caagagcttg actgccttag gaagcagtat
4981 ttgtcagaaa atgaacagtg gcaacagaag ctgacaagcg tgactctgga gatggagtcc
5041 aagttggcgg cagaaaagaa acagacggaa caactgtcac ttgagctgga agtagcacga
5101 ctccagctac aaggtctgga cttaagttct cggctctttg tggcatcga cacagaagat
5161 gctattcaag gccgaaatga gagctgtgac atatcaaaag aacatacttc agaaactaca
5221 gaaagaacac caaagcatga tgttcatcag atttgtgata aagatgctca gcaggacctc
5281 aatctagaca ttgagaaaat aactgagact ggtgcagtga aaccacagag agagtgtctc
5341 ggggaacagt cccagatac caattatgag cctccagggg aagataaaac ccagggctct
5401 tcagaatgca tttctgaatt gtcattttct ggtcctaata ctttgggtacc tatggatttc
5461 ctggggaatc aggaagatat ccataatctt caactgcggg taaaagagac atcaaatgag
5521 aatttgagat tacttcatgt gatagaggac cgtgacagaa aagttgaaag tttgctaaat
5581 gaaatgaaag aattagactc aaaactccat ttacaggagg tacaactaat gacccaaaatt
5641 gaagcatgca tagaattgga aaaaatagtt ggggaactta agaaagaaaa ctgagattta
5701 agtgaaaaat tggaatatth ttcttgtgat caccaggagt tactccagag agtagaaaact
5761 tctgaaggcc tcaattctga tttagaaatg catgcagata aatcatcacg tgaagatatt
5821 ggagataatg tggccaaggt gaatgacagc tgggaaggaga gatttcttga tgtggaaaat
5881 gagctgagta ggatcagatc ggagaaagct agcattgagc atgaagccct ctacctggag
5941 gctgacttag aggtagttca aacagagaag ctatgttttag aaaaagacaa tgaaaataag
6001 cagaaggtta ttgtctgcct tgaagaagaa ctctcagtggt tcacaagtga gagaaaccag
6061 cttcgtggag aattagatac tatgtcaaaa aaaaccacgg cactggatca gttgtctgaa
6121 aaaatgaagg agaaaacaca agagcttgag tctcatcaaa gtgagtgtct ccattgcatt
6181 caggtggcag aggcagaggt gaaggaaaag acggaactcc ttcagacttt gtcctctgat
6241 gtgagtgagc tgttaaaaga caaaactcat ctccaggaaa agctgcagag tttggaaaag
6301 gactcacagg cactgtcttt gacaaaatgt gagctggaaa accaaattgc acaactgaat
6361 aaagagaaag aattgcttgt caaggaatct gaaagcctgc aggccagact gagtgaatca
6421 gattatgaaa agctgaatgt ctccaaggcc ttggaggccg cactggtgga gaaaggtgag
6481 ttcgcattga ggctgagctc aacacaggag gaagtgcac agctgagaag aggcacgag
6541 aaactgagag ttcgcattga ggccgatgaa aagaagcagc tgcacatcgc agagaaactg
6601 aaagaacgcg agcgggagaa tgattcactt aaggataaag ttgagaacct tgaaagggaa
6661 ttgcagatgt cagaagaaaa ccaggagcta gtgattcttg atgccgagaa ttccaaagca
6721 gaagtagaga ctctaaaaac acaaatagaa gagatggcca gaagcctgaa agtttttgaa
6781 ttagaccttg tcacgttaag gtctgaaaaa gaaaatctga caaaacaaat acaagaaaaa
6841 caaggtcagt tgtcagaact agacaagtta ctctcttcat ttaaaagtct gttagaagaa
6901 aaggagcaag cagagataca gatcaaagaa gaatctaaaa ctgcagtgga gatgcttcag
6961 aatcagttaa aggagctaaa tgaggcagta gcagccttgt gtggtgacca agaaattatg

```

FIG. 1B

3/11

```

7021 aaggccacag aacagagtct agaccaccca atagaggaag agcatcagct gagaaatagc
7081 attgaaaagc tgagagcccg cctagaagct gatgaaaaga agcagctctg tgtcttaca
7141 caactgaagg aaagtgagca tcatgcagat ttacttaagg gtagagtgga gaaccttgaa
7201 agagagctag agatagccag gacaaaccaa gagcatgcag ctcttgaggc agagaattcc
7261 aaagggagagg tagagaccct aaaagcaaaa atagaaggga tgacccaaag tctgagaggt
7321 ctggaattag atgttggttac tataaggtca gaaaaagaaa atctgacaaa tgaattacaa
7381 aaagagcaag agcgaatatc tgaattagaa ataataaatt catcatttga aaatattttg
7441 caagaaaaag agcaagagaa agtacagatg aaagaaaaat caagcactgc catggagatg
7501 cttcaaacac aattaaaaga gctcaatgag agagtggcag ccctgcataa tgaccaagaa
7561 gcctgtaagg ccaagagagca gaatcttagt agtcaagtag agtgtcttga acttgagaag
7621 gctcagttgc tacaaggcct tgatgaggcc aaaaataatt atattgtttt gcaatcttca
7681 gtgaatggcc tcattcaaga agtagaagat ggcaagcaga aactggagaa gaaggatgaa
7741 gaaatcagta gactgaaaaa tcaaattcaa gaccaagagc agcttgtctc taaactgtcc
7801 caggtggaag gagagcacca actttggaag gagcaaaact tagaactgag aaatctgaca
7861 gtggaattgg agcagaagat ccaagtgcta caatccaaaa atgcctcttt gcaggacaca
7921 ttagaagtgc tgcagagttc ttacaagaat ctagagaatg agcttgaatt gacaaaaatg
7981 gacaaaaatgt cctttgttga aaaagtaaac aaaatgactg caaaggaaac tgagctgcag
8041 agggaaatgc atgagatggc acagaaaaca gcagagctgc aagaagaact cagtggagag
8101 aaaaataggc tagctggaga gttgcagtta ctggttggaa aaataaagag cagcaaagat
8161 caattgaagg agctcacact agaaaatagt gaattgaaga agagcctaga ttgcatgcac
8221 aaagaccagg tggaaaagga agggaaagtg agagaggaaa tagctgaata tcagctacgg
8281 cttcatgaag ctgaaaagaa acaccaggct ttgcttttgg acacaaacaa acagtatgaa
8341 gtagaaatcc agacataccg agagaaattg acttctaaag aagaatgtct cagttcacag
8401 aagctggaga tagacctttt aaagtctagt aaagaagagc tcaataattc attgaaagct
8461 actactcaga ttttggaaga attgaagaaa accaagatgg acaatctaaa atatgtaaat
8521 cagttgaaga aggaaaatga acgtgcccg gggaaaatga agttgttgat caaatcctgt
8581 aaacagctgg aagaggaaaa ggagatactg cagaaagaac tctctcaact tcaagctgca
8641 caggagaagc agaaaacagg tactgttatg gataccaagg tcgatgaatt aacaactgag
8701 atcaaagaac tgaaagaaac tcttgaagaa aaaaccaagg aggcagatga atacttggat
8761 aagtactgtt ccttgcttat aagccatgaa aagttagaga aagctaaaga gatgttagag
8821 acacaagtgg cccatctgtg ttcacagcaa tctaacaag attcccagg gtctcctttg
8881 ctaggtccag ttgttccagg accatctcca atcccttctg ttactgaaaa gaggttatca
8941 tctggccaaa ataaagcttc aggcagagg caaagatcca gtggaatatg ggagaatggt
9001 agaggaccaa cacctgctac cccagagagc ttttctaaaa aaagcaagaa agcagtcag
9061 agtggatttc accctgcaga agacacggaa ggtactgagt ttgagccaga gggacttcca
9121 gaagttgtaa agaaagggtt tgctgacatc ccgacaggaa agactagccc atatatcctg
9181 cgaagaacaa ccatggcaac tcggaccagc cccgcctgg ctgcacagaa gttagcgcta
9241 tccccactga gtctcggaag agaaaatctt gcagagtcct ccaaaccaac agctggtggc
9301 agcagatcac aaaagggtcaa agttgctcag cggagcccag tagattcagg caccatcctc
9361 cgagaacca ccacgaaatc cgtcccagtc aataatcttc ctgagagaag tccgactgac
9421 agccccagag agggcctgag ggtcaagcga ggccgacttg tccccagccc caaagctgga
9481 ctggagtcca acggcagtga gaactgtaag gtccagtga ggcactttgt gtgtcagtac
9541 ccctgggagg tgccagtcac tgaatagata aggctgtgcc tacaggactt ctctttagtc
9601 agggcatgct ttattagtga ggagaaaaca attccttaga agtcttaaat atattgtact
9661 ctttagatct cccatgtgta ggtattgaaa aagtttgga gcaactgatc cctgttagca
9721 ttgccattcc tctactgcaa tgtaaatagt ataaagctat gtatataaag ctttttggt
9781 atatgttaca attaaaatga caagcactat atcacaatct ctgtttgtat gtgggtttta
9841 cactaaaaaa atgcaaaaaca cattttatcc ttctaattaa cagctcctag gaaaatgtag
9901 acttttgctt tatgatattc tatctgtagt atgaggcatg gaatagtttt gtatcgggaa
9961 tttctcagag ctgagtaaaa tgaaggaaaa gcatgttatg tgtttttaag gaaaatgtgc
10021 acacatatac atgtaggagt gtttatcttt ctcttacaat ctgtttttaga catctttgct
10081 tatgaaacct gtacatatgt gtgtgtgggt atgtgtttat ttccagtgag ggctgcaggc
10141 ttcctagagg tgtgtatatac catgcgtctg tcgttgtgct tttttctgtt ttagacca
10201 ttttttacag ttcttttggt agcattgtcg tatctggtga tggattaaca tatagccttt
10261 gttttctaataaaaatagtcg ccttcgtttt ctgtaaaaaa aaaaaaaaaa aaaaaa

```

FIG. 1C

4/11

```

1 mswaleewke glptralqki qelegqldkl kkekqqrqfq ldsleaalk qkqkvenekt
61 egtnlkrenq rlmeicesle ktkqkishel qvkesqvnfq egqlnsgkkq iekleqelkr
121 ckselersqq aaqsadvsln pcntpqkift tpltpsqqys gskyedlkek ynkeveerkr
181 leaevkalqa kkasqtlpqa tmnhrdiarh qasssvfswq qektpshlss nsqrtpirrd
241 fsasyfsgeq evtpsrstlq igkrdanssf fdnsssp hll dqlkaqnqel rnkinlelrl
301 lqghekemkg qvnkfqelql qlekakveli ekekvlncr delvrttaqy dqastkytal
361 eqklkklted lscqrqnaes arcsleqkik ekekefgeel srqgrsfqtl dqeci qmkar
421 ltqelqqakn mhnvlqaeld kltsvkqgle nnleefkqkl craeqafqas qikenelrrs
481 meemkkennl lkshseqkar evchleaelk nikqclnqsq nfaemkakn tsqetmlrdl
541 qekinqgens ltlekklav adlekqr dcs qdllkkrehh ieqlndklslsk tekeskalls
601 alelkkkeye elkeektlfs cwksenekll tqmesekenl qskinhletc lktqqikshe
661 ynervrtlem drenlsveir nlhnvldsk vevetqklay melqqkaefs dqkhqkeien
721 mclktstqltg qvedlehklq llsneimdkd rcyqdlhaey eslrdllksk daslvtnedh
781 qrsllafdqq pamhhsfani igeqgsm pse rsecrleadq spknsailqn rvdslfslsle
841 sqkqmnsdlq kqceelvqik geieenlmka eqmhqsfvae tsqrisklqe dtsahqn vva
901 etlsalenke kelqllndkv eteqaeiqel kksnhlleds lkelqllset lslekkemss
961 iisl nkreie eltqengtlk einaslnqek mnliqksesf anyidereks iselsdqykq
1021 eklillqrce etgnayedls qkykaa qekn sklecllnec tslcenrkne leqlkeafak
1081 ehqefltkla faernqn l m leletvqqal rsemt dnqnn skseagglkq eimtlkee qn
1141 kmqkevndll qeneqlmkvm ktkhecqnle sepirnsvke resernqcnf kpgmdlevke
1201 isldsynaql vqleamlrnk elklqeseke keclqhelqt irgdletsnl qdmqsqeisg
1261 lkdceidaee kyisgphels tsqndnah lq cslqttnmkl nele kiceil qaekyelvte
1321 lndsrsecit atrkmaeevg kllnevkiln dds gllhgel vedipggefg eqpneqhpvs
1381 lapldesnsy ehltlsdkev qmhfaelqek flslqse hki lhdqhcqmss kmselqtyvd
1441 slkaenlvls tnlnrfqgd l vkemqlglee glvpslsssc vpdssslssl gdssfyra ll
1501 eqtgdmslls nlegavs anq csvdevfcss lqeenltrke tpsapakgve eleslcevyr
1561 qslekleekm esqgimknke iqeleql lss ergeldclrk qylseneqwq qkltsvtlem
1621 esklaaekkq teqlslelev arlqlqgd l ssrsllgidt edaiqgrnes cdiskehtse
1681 ttertpkhdv hqicdkdaq q dlnldiekit etgavkptge csgeqsp dtn yep pgedktq
1741 gssecisels fsgpnalvpm dflgnqedih nlqlrvkets nenlrllhvi edrdrkvesl
1801 lnemkeldsk lhlqevqlmt kieacielek ivgelkkens dlsekleyfs cdhqellqrv
1861 etseglnsdl emhadkssre digdnvakvn dswkerfldv enelsrirse kasiehealy
1921 leadlevvqt eklclekdne nkqkvivcle eelsvtser nqlrgeldtm skkttaldql
1981 sekmkektqe leshqseclh ciqvaaevk ektellqtls sdvsellkdk thlqeklqsl
2041 ekdsqalslt kcelenqiaq lnkekellvk eseslqarls esdyeklnvs kaleaalvek
2101 gefalrlsst qeevhqlrrg ieklr vriea dekkqlhia e klkererend slkdkvenle
2161 relqmseenq elvildaens kaevetlktq ieemarslkv feldlvtlrs ekenltkqiq
2221 ekqgqlseld kllssfksll eekeqaeiqi keesktavem lqnqlkelne avaalcgdqe
2281 imkateqsld ppieehqlr nsieklrarl eadekkqlcv lqqlkesehh adllkgrven
2341 lereleiart ngehaaleae nskgevetlk akiegmtqsl rgleldvvti rsekenltne
2401 lqkeqerise leiinssfen ilqekeqekv qmkeksstam emlqtqlkel nervaalhnd
2461 qeackakeqn lssqveclel ekaql lqgd eaknnyivlq ssvngliqev edgkqklekk
2521 deeisrlknq iqdqeqlvsk lsqvegehql wkeqnlelrn ltveleqkiq vlqsknaslq
2581 dtlevlqssy knlenelelt kmdkmsfvek vnkmtakete lqremhemaq ktaelqeels
2641 geknrlagel qlllleeikss kdqlkel tle nselkksldc mhkdqvekeg kvreeiaeyq
2701 lrlheaekkh qallldtnkq yeveiqtyre kltskeecsl sqkleidl l sskeelnnsl
2761 kattqileel kktkmdnlky vnqlkkener aqgkmllik sckqleeeke ilqkelsqlq
2821 aagekqktgt vmdtkvdelt teikelketl eektkeadey ld ky csllis heklekakem
2881 letqvahlcs qqskqdsrgs pll gpvvp gsp ipsvtekr lssgqnkasg krqrssgiwe
2941 ngrgptpatp esfskkskka vm sgihp aed tegtefepeg lpevvkkgfa diptgktspy
3001 ilr rttmatr tsprlaaqkl alsplslgke nlaesskpta ggsrsqkvkv aqrspvds gt
3061 ilrepttksv pvnnlpersp tdspreglrv krgrlvpspk aglesngsen ckvq

```

FIG. 1D

5/11

1 gattgtggga aggcagctga actcggcgcc tggaaagatg gaggcagcgg agacagaggc
61 ggaagctgca gccctagagg tcctggctga ggtggcaggc atcttggaac ctgtaggcct
121 gcaggaggag gcagaactgc cagccaagat cctggttgag tttgtggtgg actctcagaa
181 gaaagacaag ctgctctgca gccagcttca ggtagcggat ttctgcaga acatcctggc
241 tcaggaggac actgctaagg gtctcgaccc cttggcttct gaagacacga gccgacagaa
301 ggcaattgca gctaaggaac aatggaaaga gctgaaggcc acctacaggg agcacgtaga
361 ggccatcaaa attggcctca ccaaggccct gactcagatg gaggaagccc agaggaaacg
421 gacacaactc cgggaagcct ttgagcagct ccaggccaag aaacaaatgg ccatggagaa
481 acgcagagca gtccagaacc agtggcagct acaacaggag aagcatctgc agcatctggc
541 ggaggtttct gcagaggtga gggagcgtaa gacagggact cagcaggagc ttgacagggt
601 gtttcagaaa cttggaaacc tgaagcagca ggcagaacag gagcgggaca agctgcagag
661 gtatcagacc ttctccagc ttctgtatac cctgcagggt aagctgttgt tccctgaggc
721 tgaggctgag gcagagaatc ttccagatga taaaccccag cagccgactc gaccccagga
781 gcagagtaca ggagacacca tggggagaga ccctgggtgtg tccttcaagg ctggttggtct
841 acaacctgct ggagatgtaa atttgccatg acttcctgga ggacagcagc atggagaaag
901 atcctagaaa aggcctctga cttccctcac ctcccaacca tcattacagg aaagactgtg
961 aactcctgag ttcagcttga tttctgacta catcccagca agctctggca tctgtggatt
1021 aaaatccctg gatctctctc agttgtgtat ttgttcatct tcatatgctg gcaggaacaa
1081 ctattaatac agatactcag aagccaataa catgacagga gctgggactg gtttgaacac
1141 aggggtgtgca gatggggagg gggactggc cttgggcctc ctatgatgca gacatggtga
1201 atttaattca aggaggagga gaatgtttta ggcagggtgt tatatgtggg aagataattt
1261 tattcatgga tccaaatgtt tgttgagtcc tttctttgtg ctaaggttct tgcggtgaac
1321 cagaattata acagtgagct catctgactg ttttaggatg tacagcctag tgttaacatt
1381 cttggtatct ttttgtgcct tatctaaaac atttctcgat cactggtttc agatgttcat
1441 ttattatatt cttttcaaag attcagagat tggcttttgt catccactat tgtatgtttt
1501 gtttcattga cctctagtga taccttgatc tttcccaact tctgttttcg gattggagaa
1561 gatgtacctt ttttgtcaac tcttactttt atcagatgat caactcacgt atttggatct
1621 ttatttgttt tctcaaataa atattttaagg ttatacatth aaaaaaaaaa aaaaaaaaaa
1681 aaaaaaa

FIG. 2A

1 meaaeteaea aalevlaeva gilepvglqe eaelpakilv efvvdsgkkd kllcsqlqva
61 dflqnilage dtakgldpla sedtsrqkai aakeqwkclk atyrehveai kigltkaltq
121 meeagrkrqt lreafeqlqa kkqmamekrr avqnqwwlqq ekhlqhlaev saevrerktg
181 **tqgeldrvfq** klgnlkqqae qerdklqryq tflqllytlq gklfpeaea eaenlpddkp
241 qqptrpqs tgdtmgrdpq vsfkavglqp agdvnlp

FIG. 2B

6/11

```

1  gagegcctgc  cgtttctcgg  ggcgggacgg  gggggcgggg  actgggcgga  gaggcgcgtg
61  ctgctgcgtg  cgtgcgcgcg  cgccgcgggc  gggccagtga  aaccggcggc  cctggcacgt
121  gacctaggac  cggtcacccg  ggctcgcttg  tggctccgtc  tgtctgtccg  tccgcccgcg
181  ggtgccatca  tggcggacgc  ggccagtcag  gtgctcctgg  gctccggtct  caccatcctg
241  tcccagccgc  tcatgtacgt  gaaagtgtc  atccagggtg  gatatgagcc  tcttcctcca
301  acaataggac  gaaatatttt  tgggcggcaa  gtgtgtcagc  ttctggtct  ctttagttat
361  gctcagcaca  ttgccagtat  cgatgggagg  cgcggttgt  tcacaggctt  aactccaaga
421  ctgtgttcgg  gagtcccttg  aactgtggtc  catggtaaag  ttttacagca  ttaccaggag
481  agtgacaagg  gtgaggagtt  aggacctgga  aatgtacaga  aagaagtctc  atcttccttt
541  gaccacgtta  tcaaggagac  aactcgagag  atgatcgctc  gttctgctgc  taccctcatc
601  acacatccct  tccatgtgat  cactctgaga  tctatggtac  agttcattgg  cagagaatcc
661  aagtactgtg  gactttgtga  ttccataata  accatctatc  gggaagaggg  cattctagga
721  tttttcgcg  gtcttggtcc  tcgccttcta  ggtgacatcc  tttctttgtg  gctgtgtaac
781  tcaactggcct  acctcgtcaa  tacctatgca  ctggacagtg  gggtttctac  catgaatgaa
841  atgaagagtt  attctcaagc  tgtcacagga  ttttttgca  gtatgttgac  ctatcccttt
901  gtgcttgtct  ccaatcttat  ggctgtcaac  aactgtggtc  ttgctggtgg  atgccctcct
961  tactcccaa  tatatacgtc  ttggatagac  tgttggtgca  tgctacaaaa  agaggggaat
1021  atgagccgag  gaaatagctt  atttttccgg  aaggtccccct  ttgggaagac  ttattgttgt
1081  gacctgaaaa  tgtaaatttg  aagatgtggg  gcagggacag  tgacatttct  gtagtcccag
1141  atgcacagaa  ttatgggaga  gaatgttgat  ttctatacag  tgtggcgcg  ttttttaata
1201  atcatttaat  cttgggaaaa  ttcagggtgt  tgggtgtctg  cttttttgtt  cttttttcca
1261  gcacaacata  acttaccact  gatactcccc  ctttagttat  tctgaattag  gatatttttg
1321  ctccaaattc  ttattttact  taaccagaag  ggaaaaaaag  ttgtattttc  ctgaagctac
1381  aggcactttg  tcatgtgatt  tttgagtctc  aatttaaggc  tttgtaaaat  gaagagtaga
1441  attccaagaa  aaatgagaaa  taattttgta  aaacttaaca  aaatcactaa  attaaactat
1501  atgggaggtt  atgaattact  ttttcttggg  tagaccctaa  aatgtcagta  gcatgcacca
1561  gaatctgact  ccattatgc  ttctaagcac  atttcattga  ccttgtctct  catacttcaa
1621  gaaaaggaca  gtacattgct  acattaccct  agaaagtctg  tgtgaggatc  tgccccttca
1681  gtctgttatt  gcaaagtaat  aaaatgtcac  ctacagggag  cctctgagcc  tactctagtt
1741  caagaggcta  cctgaaaaaa  aataaataag  ataaagggtc  agcaacaaca  aagaaaaaga
1801  caattacaga  aaataagcaa  gatttggaag  ggaagtataa  tggcactttt  ttcctcaaag
1861  gaagtctctg  ttttcacata  aaatatgaaa  agcagatcct  gcaggagtaa  ccccttctt
1921  taagagccaa  gtatttgcca  gtgcttaaat  tacaccatac  agttctaatt  atatataatc
1981  ttttgttctt  cagttttttg  ttttgtttcc  tttttgttat  tgttgccgaa  ggtgagtagt
2041  tttgcatttc  tgatgacagc  cttggaaagt  atatttgtaa  ctccatgtct  ggtaatgcca
2101  acccaagtcg  acatgggtct  taggacactg  accacctcac  atgccatacc  ctcagttaag
2161  catgttaaca  tttataggag  gaaaaaaatc  actttgggag  aaaataaaat  tcaactcaag
2221  cataaagctt  ctgtttactc  aggccttcta  aaaagcaggt  taaaatgctc  taaaatgaga
2281  aagcctgtgg  tttcacttat  ttatataact  cactgggaca  ttgccaaatg  agtaagcact
2341  taattcgctg  cttctgagac  ttctctgtca  aaacagcccc  actgataata  ttagacagaa
2401  cgagaatgca  ggggtctctt  cctcccctg  gggtttagga  agctcatgag  gagctcggct
2461  taaaatgtct  ttgatgtctc  ttcctttgtc  tcaaaaagta  atgtcaattt  tatatactat
2521  ttcaatatta  ctatctgcat  ttgttttaat  ataaaaatgt  ttgctgccta  cttttttctc
2581  ccaaaaaatc  tttaagtaaa  gatgatctgg  gaaaatgtgc  catgttta

```

FIG. 3A

```

1  madaasqvll  gsgltilsqp  lmyvkvliqv  gyeplpptig  rnifgrqvcq  lpglfsyaqh
61  iasidgrrgl  ftgltprlcs  gvlgtvvhvk  vlqhyqesdk  geelgpqnvq  kevsssfdhv
121  ikettremia  rsaatlithp  fhvitlrsmv  qfigreskyc  glcdsiitiy  reegilgffa
181  glvprllgdi  lslwlcnsla  ylvntyalds  gvstmnemks  ysqavtgffa  smlypfvlv
241  snlmavnncg  laggcppysp  iytswidcwc  mlqkegnmsr  gnslffrkvp  fgktyccdlk
301  mli

```

FIG. 3B

7/11

```

1  agaaacactt ggtttcatgt atgactcata agctgaagca cagacaccac ttccccaatc
61  tacaggagcc attttaacag ctaaaacttg tcggattgct ttttattttc aagctcaaaa
121 gacgatagag aaagaatact tgaaggccaa gaagcttgag agaagaaaaa tttcagaaaa
181 attgtctcaa tttgactaga atatcaatga accaggaaaa ctgaagcacc ttccctaaag
241 aaaacttggg tatacaatta ctccacagac agagctgagg gttttttacc caaatcagtc
301 actggatttt gctgcctgat acgtgaatct tcttggaatt tttctcatgt ggatctaagg
361 ggaatgcttt attatggctg ctggtgtcca acagaacgac ctagtatttg aatttgctag
421 taacgtcatg gaggatgaac gacagcttgg tgatccagct atttttcctg ccgtaattgt
481 ggaacatgtt cctgggtgctg atattctcaa tagttatgcc ggtctagcct gtgtggaaga
541 gccaatgac atgattactg agagttcact ggatgttgct gaagaagaaa tcatagacga
601 tgatgatgat gacatcacc ttacagttga agcttcctgt catgacgggg atgaaacaat
661 tgaaactatt gaggtgctg aggcactcct caatatggat tcccctggcc ctatgctgga
721 tgaaaaacga ataaataata atatatattag ttcacctgaa gatgacatgg ttggtgcccc
781 agtcacccat gtgtccgtca cattagatgg gattcctgaa gtgatggaaa cacagcaggt
841 gcaagaaaaa tatgcagact caccgggagc ctcatcacca gaacagccta agaggaaaaa
901 aggaagaaaa actaaaccac cacgaccaga ttccccagcc actacgcca atatatctgt
961 gaagaagaaa acaaagatg gaaagggaaa cacaatttat ctttgggagt ttttactggc
1021 actgctccag gacaaggcta cttgtcctaa atacatcaag tggaccagc gagagaaaag
1081 catttttaaa ttggtggatt ctaaagcagt gtccaggttg tgggggaagc aaaaaacaa
1141 acctgatatg aattatgaga ccatgggaag agcactcagg tactattacc aaaggggtat
1201 tctggcaaaa gtggaaggtc agcgcttggg gtatcagttt aaagaaatgc caaagatct
1261 tatatatata aatgatgagg atccaagttc cagcatagag tcttcagatc catcgctatc
1321 ttcacagcc acttcaaata ggaatcaaac cagccggtcg agagtatctt caagtcaggy
1381 ggtaaaagga ggagccacta cagttctaaa accagggaat tctaaagctg caaaacccaa
1441 agatcctgtg gaagttgcac aaccatcaga agttttgagg acagtgcagc ccacgcagtc
1501 tccatatcct acccagctct tccggactgt tcatgtagta cagccagtac aggctgtccc
1561 agagggagaa gcagctagaa ccagtaccat gcaggatgaa acattaaatt ctccgttca
1621 gagtattagg actatacagg ctccaaccca agttccagtg gttgtgtctc ctaggaatca
1681 gcagttgcat acagtaacac tccaaacagt gccactcaca acagttatag ccagcacaga
1741 tccatcagca ggtactggat ctcagaagtt tattttacaa gccattccat catcacagcc
1801 catgacagta ctgaaagaaa atgtcatget gcagtcacaa aaggcgggct ctctccttc
1861 aattgtcttg ggccctgccc aggttcagca ggtccttact agcaatgttc agaccatttg
1921 caatggaacc gtcagtggtg cttcctctcc atccttcagt gctactgcac ctgtggtgac
1981 cttttctcct cgcagttcac agctggttgc tcacccacct ggcactgtaa tcacttcagt
2041 tatcaaaact caagaaacaa aaactcttac acaggaagta gagaaaaagg aatctgaaga
2101 tcatttgaaa gagaacactg agaaaacgga gcagcagcca cagccttatg tgatggtagt
2161 gtccagttcc aatggattta cttctcaggt agctatgaaa caaaacgaac tgctggaacc
2221 caactctttt tagttaatat accaaagctt atgaataatt gtttgtaaat tgaacatttt
2281 caattatatg cagactgact gattctaaga taaattctaa ggaggtttct aattttgtaa
2341 ttgttaaaaa tagagttaat tttgactttg ttagatgagg gaggaaaact caactgtttc
2401 tctttgttat ctaaatgttt cagaattcaa tcgtgaagga acaggcattt tacactatga
2461 agacattctt ttgagatttt tatttcagtt gctatatcat aagcattttt aaagtctctt
2521 ttctaatttt acattgtatt agattttctg attcctttgt aaatacagaa cttaaataga
2581 aggcaacagg aaatttatat aggaactatt ttcattccac ttgtgtaagt taagtcttga
2641 ctctttcaaa tgcaaaaaac ctattttatg ctttggttaa attatggtgt cacttagatt
2701 gacttttagt gactgcacta tataatatag aactatgaat atgtagaata acatgaaaaa
2761 ttggaggtgc tgggtggtat gctgaccctg tttcagaagc aggatagtat aaaagcatca
2821 gcctaagaat ggcactccca ctaactagct atgtaatctt gacctctttg ggctttagtt
2881 cctctcataa aaggaagaga tgtattggat tagactagat gatcaccact ttctcttcta
2941 gttctaattt ttttaattct aatacctata ttttcaagtt atgtcaatta aatcattatc
3001 aggttatttc ctaatgtaag aatagctaaa atggtgcaga gaaataagtg acccaacaaa
3061 atttattcat ctgttatggg taagatctgc cataaattct tcctaataaa tttgtttact
3121 aactctttag gccactgtgc tttgcgggtc attagtaaac ttgtgttgct aagtgctaaa
3181 cagaatactg ctattttgag agagtcaaga ctctttctta agggccaaga aagcaacttg
3241 agccttgggc taatctggct gagtagtcag ttataaaaagc ataattgctt tatatttttg
3301 atcatttttt actgggggcg gacttggggg ggggtgcata caaagataac atatatatcc
3361 aactttctga aatgaaatgt ttttagatta ctttttcaac tgtaaataat gtacatttaa
3421 tgtcacaaga aaaaaatgtc ttctgcaaat tttctagtat aacagaaatt tttgtagatg
3481 aaaaaaatca ttatgtttag aggtctaatt ctatgttttc atattacaga gtgaatttgt

```

FIG. 4A

8/11

3541 atttaaaca aaattttaaat tttggaatcc tctaaacatt tttgtatcct taattggttt
 3601 attattaaat aaatcatata aaaattctca gtgtctgttt tcaggcaaaa gtttcttaaa
 3661 gaataagtgt gcagagaata ttactagaac atcagcatta cttaatgttt ataaataaat
 3721 ttcattagtc agaattgcaa aaaaaaaaaa aaaaaaaaaa

FIG. 4B

1 maavvqqndl vfefasnvme derqlgdpai fpavivehvp gadilnsyag lacveepndm
 61 itessldvae eeiiddddd itltveasch dgdetietie aaeallnmds pgpmldekri
 121 nnnifssped dmvvapvthv svtldgipev metqqvqeky adspgasspe qpkrkkgrkt
 181 kpprpdspat tpnisvkkkn kdgkgntiyl wefllallqd katcpkyikw tqrekgifkl
 241 vdkavsrllw gkhknkpdmn yetmgralry yyqrgilakv egqrlvyqfk empkdliyin
 301 dedpssies sdpslsssat snrnqtsrsr vssspgvkkg at**tvlpkpgns** kaakpkdpve
 361 vaqpsevlrt vqptqspypt qlfrtvhvqv pvqavpegea artstmqdet lnssvqsirt
 421 iqaptqvpvv vsprnqqlht vtlqtvpltt viastdpsag tgsqkfilqa ipssqpmtvl
 481 kenvmlqsqk agsppsivlg paqvqqvltv nvqticngtv svasspsfsa tapvvtfspr
 541 ssqvlvahppg tvitsviktt etktltqeve kkesedhlke ntekteqqpq pyvmvsssn
 601 gftsqvamkq nellepnsf

FIG. 4C

1 ccgcccttgt aggctgtcca cctcaaacgg gccggacagg atatataaga gagaatgcac
 61 cgtgcactac acacgcgact cccacaaggt tgcagccgga gccgcccagc tcaccgagag
 121 cctagtccg gccagggtcg ccccggaac cacgagccca gccaatcagc gccccggact
 181 gcaccagagc catggtcggc agaagagcac tgatcgtact ggctcactca gagaggacgt
 241 ccttcaacta tgccatgaag gaggctgctg cagcggcctt gaagaagaaa ggatgggagg
 301 tggtagagtc ggacctctat gccatgaact tcaatcccat catttccaga aaggacatca
 361 caggtaaact gaaggaccct gcgaactttc agtatcctgc cgagtctgtt ctggcttata
 421 aagaaggcca tctgagccca gatattgtgg ctgaacaaaa gaagctggaa gccgcagacc
 481 ttgtgatatt ccagttcccc ctgcagtggt ttggagtccc tgccattctg aaaggctggt
 541 ttgagcgagt gttcatagga gagtttgctt acacttacgc tgccatgtat gacaaaggac
 601 ccttccggag taagaaggca gtgctttcca tcaccactgg tggcagtggc tccatgtact
 661 ctctgcaagg gatccacggg gacatgaatg tcattctctg gccaatcag agtggcattc
 721 tgcatctctg tggcttccaa gtcttagaac ctcaactgac atatagcatt gggcacactc
 781 cagcagacgc ccgaattcaa atcctggaag gatggaagaa acgcctggag aatatttggg
 841 atgagacacc actgtatatt gctccaagca gcctctttga cctaaacttc caggcaggat
 901 tcttaatgaa aaaagaggta caggatgagg agaaaaacaa gaaatttggc ctttctgtgg
 961 gccatcactt gggcaagtcc atcccaactg acaaccagat caaagctaga aatgagatt
 1021 ccttagcctg gatctccttc taacatgtta tcaaactctg gtatctttcc aggttccct
 1081 gacttgcttt agtttttaag atttgtgttt ttctttttcc acaaggaata aatgagaggg
 1141 aatcgactgt attcgtgcat ttttgatca tttttaactg attcttatga ttactatcat
 1201 ggcatataac caaaatccga ctgggctcaa gaggccactt agggaaagat gtagaaagat
 1261 gctagaaaaa tgttctttaa aggcactctac acaatttaac tcctcttttt agggctaaag
 1321 ttttagggta cagtttggct aggtatcatt caactctcca atgttctatt aatcacctct
 1381 ctgtagttta tggcagaagg gaattgctca gagaaggaaa agactgaatc tacctgccct
 1441 aagggactta acttgtttg tagttagcca tctaagtctt gtttatgata tttcttgctt
 1501 tcaattacaa agcagttact aatatgccta gcacaagtac cactcttggc cagcttttgt
 1561 tgtttatata cagtacacag ataccttgaa aggaagagct aataaatctc ttctttgctg
 1621 cagtcactca cttttttttt aattaaaaaa aatttttttt tgaagcagtc ttgctctgtt
 1681 acccaggctg gaggcagtg gtgtgatctc ggctcactgc aacctctgcc tcccaggctc
 1741 cagcaattct cctgcctcag cctccctagt agctgggatg acaggcgcct gccatcatgc
 1801 ctgactaatt tttgtatttt tagtagagac ggcgtttcac catgttggcc aggttggtct

FIG. 5A

9/11

```

1861 caaactcctg acctcaggtg atccgcctac ctcagcctcc caaagtgctg ggattacagg
1921 cgtgatccac cacacctggc ccttgcaatc ttctacttta aggtttgcag agataaacca
1981 ataaatccac accgtacatc tgcaatatga attcaagaaa ggaaatagta ccttcaatac
2041 ttaaaaaatag tcttccacaa aaaatacttt atttctgata tatacaaatt ttcagaaggt
2101 tattttcttt atcattgcta aactgatgac ttactatggg atgggggtcca gtcccatgac
2161 cttgggggtac aattgtaaac ctagagtttt atcaactttg gtgaacagtt ttggcataat
2221 agtcaatttc tacttctgga agtcatctca ttccactggt ggtattatat aattcaagga
2281 gaatatgata aaacactgcc ctcttggtgt gcatlgaaag aagagatgag aatgatgaa
2341 aaggttgcct gaaaaatggg agacagcctc ttacttgcca agaaaatgaa gggattggac
2401 cgagctggaa aacctccttt accagatgct gactggcact ggtggttttt gctctcgaca
2461 gtatccacaa tagctgacgg ctgggtgttt cagtttgaaa atattttgtt gccttcatct
2521 tcaactgcaat tttgtgtaaa tttctcaaag atctgaatta aataaataaa attcatttct
2581 acagaccac aaaaaaaaaa a

```

FIG. 5B

```

1 mvgrralivl ahsertsfny amkeaaaaal kkkgwevves dlyamfnpi isrkditgkl
61 kdpanfqypa esvlaykegh lspdivaeqk kleaadlvif qfplqwfvgp ailkgwferv
121 figefaytya amydkgpfrs kkavlsittg gsgmyslqg ihgdmnvilw piqsgilhfc
181 gfvleplqly ysightpada riqilegwkk rleniwdetp lyfapsslfd lnfqagflmk
241 kevqdeeknk kfglsvghhl gksiptdnqi kark

```

FIG. 5C

```

1 ggattttctag gaggaccggc agaggcgcgc ataggtgcgt ggtgctgggc ccgggcgcgcg
61 cggcaccggt gtaggagcgc gcatctccag agtttcttcc atctgggcga cgtctcgggtg
121 cctgcggcgg gaacggcgct ttgcttccct gaggagcttc tagagagcta cgggtggcccc
181 cgtgtgggag gcggggggcg tggcggcgtc ggggcgtcgc tgtcccctcc tcggtagctc
241 tcctcccctcc cttttctgct gttaccggga gcgcgggtggc cacggaacgc tgcccggagc
301 cgcgcgaggg aggacccgac gcgcggcggt taccagcgc agcgttccac cgctcgggtt
361 tggctggata aaataaaaaa tggggatatt gacctcctgt cactactgca tggactttga
421 tggtttccaa tcattacttt ctctctgtgt tcaatctgcc tcttcgagaa attcatactc
481 ctgaatagct ctccagaccc ccagctggcc atgtggtgag ttcagggccc aaatcaagta
541 gtaccagcaa tcagggaact cctatctggt ttgaatggat tcacaccagc cacaagcctg
601 gaaagatggt gtcacaatct acagtcaggg aggattctcc tgtggagccc tgggaagggga
661 tcagcgatca ctctggcatt attgatggtt cgcccagact cctgaacact gaccatcctc
721 cttgccaatt agacatcagg ctcatgaggg acaaagctgt ctggattaac ccccaggatg
781 tgcagcaaca gccgcaggac ttgcaatctc agggtgccagc agcagggaac agtgggaccc
841 attttgtgac agatgctgcc tctccctcag gcccttcacc ttcgtgcctc ggggactccc
901 tggcagagac aacgttgtct gaggatacca cagactccgt tggcagcgct tctccccatg
961 gctcgagtga aaagagtagc agcttctctc tgtcctcaac agaggtacac atggtccgcc
1021 caggatactc tcatcgggtg tctctgccc caagccctgg gattttggcc acctccccat
1081 atcctgagac tgacagtgt ttttttgagc cttcccattc gacatctgct gctgatgaag
1141 gtgctgttca agtcagtaga agaaccattt cttcgaattc cttctacca gaggtatttg
1201 tgctgcctgt tgatgtagaa aaggaaaatg cccactttta tgttgagat atgattatat
1261 cagcaatgga gaaaatgaag tgtaacattc tgagtcaaca gcagacagag agctggagta
1321 aagaagt cag tgggttactt gggagtgatc agcctgactc tgaaatgact tttgatacca
1381 acataaagca agagtctggg tcttctactt cttcatacag tggctatgaa ggttgtgctg
1441 tgttacaggt cagcccagtg actgaaacac gtacttacca tgatgtgaaa gagatttgca
1501 aatgcgatgt tgatgaattt gttatttttag agcttgagaa ttttaatgat atcacagaaa
1561 cctgtagctg ttcctgcagc tcctctaaga gtgtcactta tgagccagac ttcaattctg
1621 cagaactatt agccaaagag ctgtaccgcg tgttccagaa gtgctggata ctgtcagtag
1681 ttaattctca gctggcaggt tcctgagtg cagctggctc gatagtcgta aatgaagagt
1741 gtgtccgaaa agactttgaa tccagtatga atgtagtaca ggaaattaaa tttaagtcta

```

FIG. 6A

10/11

```

1801 ggatcagagg gactgaagac tgggctcctc ctagatttca aatcatattht aatattcatc
1861 caccactcaa gagggacctt gtggtggcag cccagaattt tttctgtgcc ggctgtggaa
1921 ctccagtaga gcctaagttt gtgaagcggc tccggtactg cgaataaccta gggaagtatt
1981 tctgtgactg ctgccactca tatgcagagt cgtgcatccc tgcccgaatc ctgatgatgt
2041 gggacttcaa gaagtactac gtacagcaatt tctccaaaca gctgctcgac agcatatggc
2101 accagcccat tttcaatttg ctgagcatcg gccaaagcct gtatgcgaaa gccaaggagc
2161 tggacagagt gaaggaaatt caggagcagc tcttccatat caagaagctg ttgaagacct
2221 gtaggtttgc taacagtgc ttaaaggagt tccagcaggt gccgggacac ttgactgatg
2281 agctccacct gttctccctt gaggacctgg tcaggatcaa gaaagggctg ctggcacctc
2341 tactcaagga cattctgaaa gcttcccttg cacatgtggc tggctgtgag ctgtgtcaag
2401 gaaagggctt ttttgtgaa ttttgccaga atacgactgt catcttccca tttcagacag
2461 caacatgtag aagatgttca gcgtgcaggg cttgctttca caaacagtgc ttccagtcct
2521 ccgagtggcc ccggtgtgcg aggatcacag cgaggagaaa acttctggaa agtgtggcct
2581 ctgcagcaac atgatgcccc tgagtactgt gaaaaagact gttcaacatg ccttatgata
2641 acaccgattt gtgtctatta ttggtgacat tgttttagat attgggtatt gtatattaag
2701 gaaaaagatg gtctatatcc tctttattgc atatacttaa tgtttcaaaa gaatgcagat
2761 tctgtgttta agcacagggc tgatagttgt ggttttgttt acaaagtgtc tgttttggct
2821 gctattggtt ttttaaagag gttttttata cttttgtatt tgaatagtta tgtttcactg
2881 atgctgagcc agtttgtatg tgtgtgcata tatgtgaact gtaactgaca agatgaatta
2941 ctgagtttct ctttctctaa agcttgtttg atgaaactgg ttggtccttt cagtgaacaa
3001 aaatatgacc ccaaactctgt ttgctctggc ttttatthtct tcaggaagca gacttccact
3061 taaatgccat tttgtgattg tgtcaatcat acacatttta tttacttcag agtttgaata
3121 gagagtacac atttcttctg cagatttatt tcatgatgag tttgagttgc ttagcagggc
3181 gtgtgggtcc cgttgaagtg cagtttgaag caactgcttc tagatggcac tctttcaggt
3241 ggcacaaatt gaacctgtat ttgtcatctc tgttccacac actgcaatgt caagggatgc
3301 agaagtgagt agaattccat ccctgccctt gaggatcttg ctttaacaga tgtaaaactg
3361 aacataaggt atttgcagat ttaaacgaac tgggggaaat aatgaacagt gtgattctag
3421 taataacatt aaaatcatag acattgacta ataaggttaa atgaatcaca aaacctttat
3481 gaatttcttt tttctaatag ttcttatatg ttttcctgaa acatgtgagc ctattctttt
3541 ttcttctact ttctatatac tttctcccac ttgagaaagg ggccttgagg ctgggtccct
3601 tcatggtata ccttttagact gaacggtttg caacctaggg cttgggcatt acattccctg
3661 ggattcacat gccctaacta aacctacctt gattttctca gacagcacag gcaggcaata
3721 aagcgtcaca gattgtcccc taaccccatc cagccatgtg tatgagtgtg ttttattcaa
3781 tgggatagta ctgagcacat gaaagaaatg aatgacttct gtcaatctct tttcattcag
3841 tcttctcatt ctgtcaattg ttttctcatc cgcagtgcct ctgccagAAC tgtgctcaca
3901 tccattattht aagccagatc ttttctaagt attatagaag tgtagaggca catagaataa
3961 ataaaaccag acttcaaaaa aaaaaaaaaa aaa

```

FIG. 6B

```

1 mvsqstvrqd spvepwegis dhsgiidgsp rllntdhppc qldirlmrhk avwinpqdvq
61 qqpqdlqsqv paagnsgthf vtdaaspsgp spscldgsla ettlsettdt svgsasphgs
121 seksssfsls stevhmvrpg yshrvslpts pgilatspyp etdsaffeps hltsaadega
181 vqvsrrtiss nsfspevfvl pvdvekenah fyvadmiisa mekmkcnils qqqtewske
241 vsllgsdqp dsemtfdtni kqesgsstss ysgyegcavl qvspvtetrt yhdvkeickc
301 dvdefvilel gdfnditetc scscsssksv tyepdfnsae llakelyrvf qkcwilsvvn
361 sqlagslsaa gsivvneecv rkdfessmnv vqeikfksri rgtedwappr fqiifnihpp
421 lkrdlvvaaq nffcagcgtg vepkfvkrllr yceylgkyfc dcchsyaesc iparilmmwd
481 fkkyyvsnfs kqlldsiwhq pifnllsigq slyakakeld rvkeiqeqlf hikkllktr
541 fansalkefe qvpghltdel hlfsledlvr ikkgllapl1 kdilkaslah vagcelcggk
601 gficefcqnt tvifpfqtat crrcsacrac fhkqcfqsse cprcaritar rkllsvasa
661 at

```

FIG. 6C

11/11

```

1 aaaccatggg ggcggaagag gtgctgggca ggaggcgga gaggtgctgt gcaggaggcg
61 ggcgggcgcg gttctttccg gaaggattga atctccttta gccccgccc cctccgtagc
121 tgctgaagt agtgcagggt cagcccgcga gttgcaggtc atggcgctgg ctgctcgact
181 gtggcgccct ctgcctttcc gacgtggagc cgccccgggg tctcgtctcc ctgcggggac
241 ttcgggcagc cgcgggcatt gcgccccctg tcgattccgc ggcttcgagg taatgggaaa
301 cccaggaact ttcaaaagag gccttttact ctgagctttg tcgtatttgg gttttgaaac
361 ttaccagggt atctctcagg ctgctgtggt tcatgccaca gccaaagtgt aagaaatact
421 tgaacaagca gactacctgt atgaaagcgg agaaacagaa aaactttatc agttgctaac
481 ccaatacaag gaaagtgaag atgcagagtt actgtggcgt ttggcacggg catcacgtga
541 tgtagctcag cttagcagaa cctcagaaga ggagaaaaag ctattggtgt atgaagccct
601 agagtatgca aaaagagcac tagaaaaaaa tgaatcaagt ttgcatctc ataagtggta
661 tgcaatctgc cttagtgatg ttggagatta tgaaggcatc aaggctaaaa ttgcaaatgc
721 atatatcatc aaggagcatt ttgagaaagc aattgaactg aaccctaaag atgctacttc
781 aattcacctt atgggtattt ggtgctatac atttgccgaa atgccttggg atcaaagaag
841 aattgctaaa atgctgtttg caactcctcc tagttccacc tatgagaagg ccttaggcta
901 ctttcacagg gcagaacaag tggatccaaa cttctacagc aaaaacttac ttcttttagg
961 aaagacatac ttgaaactac acaacaaaaa gcttgctgct ttctggctaa tgaaagccaa
1021 ggactatcca gcacacacag aggaggataa acagatacag acagaagctg ctgagttgct
1081 tacaagtttc agtgagaaga attgagaact tttcagagaa gatttatgaa atagctaata
1141 aacattgcct tttcttttaa ttctaaactt aatatatgaa ctataactgt tctacggctt
1201 tttaaatggt gtgaccatth aaccgtgtaa atataaaata ttctaggctt cttcaciaat
1261 aatagggtaa aataaataat cgccataaga gtggtagaaa taaatctcca tggctcaggc
1321 aaagagatta ttttgcatcc tggataccag caatgcaaaa tggtagaga tttctaagga
1381 ttgatcacat tgggatggga gatcaagcaa agaaatatth gttagaggag ggaaatggat
1441 ctatagggga tatacagggg gatggattth caaattggat tgattctaag ttgaaatctt
1501 gaagagaagg tgtggtgaca gtggttagga tgttggtggg tcctgacata aagtagttaa
1561 atgatatatc ttggagctaa cctgtgtaag taaagaacta agtaaggaga tgactaaaaa
1621 tggagtagtt tcctttttta tttttttgag acagagtctc actttgtttc ccaggctggt
1681 gtgcagtggc acaatctcgg cccactgcag cctccgcctc cggggttcaa gtgattctcc
1741 tgcttagcc tcctgagtgg ctgggattac aggttggtac caccacactc ggctaactth
1801 tgtatthttg gtagagatgg ggttttgcca tgttggttag gctggtctca aactcctggc
1861 ctcaagtgat ctgcccgcct tggcctccca aattgctggg attacaggcg tgagccaccg
1921 cacctggcca gtttacttta aatgtggtgt agtctcatgg taaactgaat ttgtcatcag
1981 atgcaaagtt ctattcccta atggaatgga aggaacacaa aacttaagag tgaaatggaa
2041 tactaagatg tttttaaata ggcaggacta tgctactcac ttgaggctgg agtgccacca
2101 ctgcaaaatc tttttaagtt ttgtaaaaag gagcatcttg aatccactta gataaagaca
2161 gactgtgtgt gtaggtggat ttttcccaaa ggatttgggg attgtaatgt tacaatgaac
2221 tgtatggata tgtttgtcat gtacatthtc aaacaaaaag gaaaactgaa agtagtgatc
2281 tttgtatacc catctcttag attcagtgat tttgctatat aggttggtga tcccttatct
2341 gaaatacttg ggactagtag aagcatcttg gatttgggat gtttttccaa attttggaat
2401 acctgcatac acacaataag atatcttgga gatgggaccc aagtttaaac acaaattcac
2461 ttgtttcata tataccttat gcacatagct tgaaggtaac tttatataac attatthtta
2521 ataathttgt gcattgagac caagtttgca taccttgaac catcagaaag caaagggtgc
2581 attatctcag ccactcatgt gggtaathtt tgggttggtt atgtcaccat cattcctgac
2641 tgaatgtata tgctaccaat aagcagttat tttcttatac ttattcatgc ataagtactt
2701 aacagtaaaa aatatgacat aactcgcaca ggaacaagga tggcaaaaaa aaaaatatga
2761 cacaccactg atacagtga aataaatgt ggtcagggtg gctaggcaac agtagcatca
2821 ccagaaacct gtatcagctg ttaaaccggc acaacaatgg caggctttca gtttccact
2881 taatgatgct gtatthttaa aggttattgt atactgtaat tttatthttg taggtgaaga
2941 gaaacagaag cagctgaagg gccaggaagt gggctthttc agggatgtgg cattctgctg
3001 gatggcttht taaatgggt ttttctctt agggagaccg aataaactgt gttgtgcacc
3061 tgca

```

FIG. 7A

```

1 malaarlwrl lpfrrgaapg srlpagtsgr rghcgpcrfr gfevmgnpgt fkrglllsal
61 sylgfetyqv isqaavvhat akveeileqa dylYESgete klyqltqyk esedaellwr
121 larasrdvaq lsrtseeekk llvyealeya kralekness fashkwyaic lsdvgdyegi
181 kakianayii kehfekaiel npkdatsihl mgiwcytfae mpwyqrriak mlfatppsst
241 yekalgyfhr aeqvdpnfys knllllgkty lklhnkklaa fwlmkakdyp ahteedkqiq
301 teaaqltsf sekn

```

FIG. 7B