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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
13 November 2014 (13.11.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/180490 Al

(51) International Patent Classification:
A61K 39/00 (2006.01) *G06F 19/18* (201 1.01)

(21) International Application Number:
PCT/EP20 13/00 1400

(22) International Filing Date:
10 May 2013 (10.05.2013)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicants: **BIONTECH AG** [DE/DE]; An der Goldgrube 12, 55131 Mainz (DE). **TRON - Translationale Onkologie an der Johannes Gutenberg-Universitat Mainz gemeinnutzige GmbH** [DE/DE]; Langenbeckstrasse 1, 551 31 Mainz (DE). **UNIVERSITÄTSMEDIZIN DER JOHANNES GUTENBERG-UNIVERSITAT MAINZ** [DE/DE]; Langenbeckstrasse 1, 55131 Mainz (DE).

(72) Inventors: **SAHIN, Ugur;** Philipp-von-Zabern-Platz 1, 55116 Mainz (DE). **TADMOR, Arbel David;** Am Anger 10, 55294 Bodenheim (DE). **CASTLE, John Christoph¬er;** Rudolf Diesel Strasse 40, 551 31 Mainz (DE). **BOEGEL, Sebastian;** Bergstrasse 13, 67823 Obermoschel (DE). **LÖWER, Martin;** Freseniusstrasse 19, 65193 Wiesbaden (DE).

(74) Agents: **SCHNAPPAUF, Georg** et al; ZSP Patentanwalte, Partnerschaftsgesellschaft, Radlkoferstrasse 2, 81373 Munchen (DE).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available):* AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, 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, KM, 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, ML, MR, NE, SN, TD, TG).

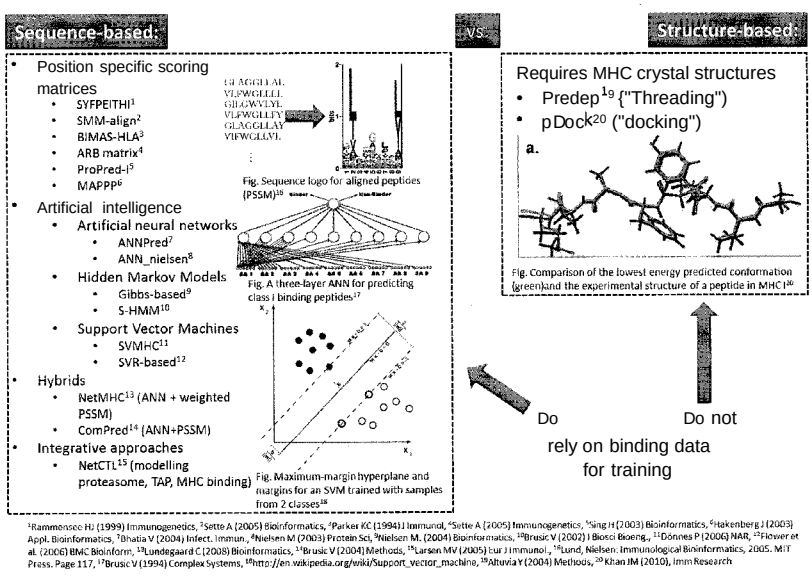
Published:

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

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

(54) Title: PREDICTING IMMUNOGENICITY OF T CELL EPITOPES

Fig. 1



[1]Rammensee HJ (1999) Immunogenetics, [2]Sette A (2005) Bioinformatics, [3]Parker KC (1994) J Immunol, [4]Sette A (2005) Immunogenetics, [5]Singh H (2003) Bioinformatics, [6]Hakenberg J (2003) Appl. Bioinformatics, [7]Dhatia V (2004) Infect. Immun., [8]Nielsen M (2003) Protein Sci, [9]Nielsen M. (2004) Bioinformatics, [10]Brusic V (2002) J Biosci Bioeng., [11]Donnes P (2006) NAR, [12]Flower et al. (2006) BMC Bioinform. [13]Lundegaard C (2008) Bioinformatics, [14]Brusic V (2004) Methods, [15]Larsen MV (2005) Eur J Immunol., [16]Lund, Nielsen: Immunological Bioinformatics, 2005. MIT Press. Page 117, [17]Brusic V (1994) Complex Systems, [18]http://en.wikipedia.org/wiki/Support_vector_machine, [19]Altuvia Y (2004) Methods, [20] Khan JM (2010), Imm Research

(57) **Abstract:** The present invention relates to methods for predicting T cell epitopes. In particular, the present invention relates to methods for predicting whether modifications in peptides or polypeptides such as tumor-associated neoantigens are immunogenic or not. The methods of the invention are useful, in particular, for the provision of vaccines which are specific for a patient's tumor and, thus, in the context of personalized cancer vaccines.

WO 2014/180490 A1

PREDICTING IMMUNOGENICITY OF T CELL EPITOPES

TECHNICAL FIELD OF THE INVENTION

The present invention relates to methods for predicting T cell epitopes. In particular, the present invention relates to methods for predicting whether modifications in peptides or polypeptides such as tumor-associated neoantigens are immunogenic or not. The methods of the invention are useful, in particular, for the provision of vaccines which are specific for a patient's tumor and, thus, in the context of personalized cancer vaccines.

BACKGROUND OF THE INVENTION

Personalized cancer vaccines are therapeutic vaccines custom tailored to target tumor-specific mutations that are unique to a given patient. Such a treatment offers great hope for cancer patients as it does not harm healthy cells and has the potential to provide life-long remission. Yet not every mutation expressed by the tumor can be used as a target for a vaccine. In fact, most cancer somatic mutations will not lead to an immune response when vaccinated against (J. C. Castle et al., Exploiting the mutanome for tumor vaccination. Cancer Research 72, 1081 (2012)). Since tumors can encode as many as 100,000 somatic mutations (M. R. Stratton, Science Signalling 331, 1553 (201 1)) whereas vaccines target only a handful of epitopes, it is apparent that a critical goal of cancer immunotherapy is to identify which mutations are likely to be immunogenic.

From a biological perspective, in order for a somatic mutation to generate an immune response several criteria need to be satisfied: the allele containing the mutation should be expressed by the cell, the mutation should be in a protein coding region and nonsynonymous, the translated protein should be cleaved by the proteasome and an epitope containing the mutation should be presented by the MHC complex, the presented epitope should be recognized by a T cell receptor (TCR) and, finally, the TCR-pMHC complex should launch a signaling cascade that activates the T cell (S. Whelan, N. Goldman, Molecular biology and evolution 18, 691 (2001)). Thus far no algorithm has been put forth that is capable of predicting with a high degree of certainty which

mutations are likely to fulfill all these criteria. In the present report we consider several factors that may contribute to immunogenicity, compare these factors against experimental data and propose a simple model for identifying immunogenic mutations.

MHC binding prediction: State of the art

Over 20 years ago, it was established, that there are positions in a MHC binding peptide, which contribute more to the binding capability then others (e.g., (A. Sette et al., Proceedings of the National Academy of Sciences 86, 3296 (1989))). The identification and description of those anchor positions enabled finding patterns of MHC binding peptides and thus were the basis for developing methods for predicting. In recent years significant developments in the field of ***in silico*** models of the Antigen Processing Machinery were achieved. The two pioneering approaches, which were developed in the late 1990's, BIMAS (K. C. Parker, M. A. Bednarek, J. E. Coligan, The Journal of Immunology 152, 163 (1994)) and SYFPEITHI (H.-G. Rammensee, J. Bachmann, N. P. N. Emmerich, O. A. Bachor, S. Stevanovic, rmmunogenetics 50, 213 (1999)), were based on the knowledge of anchor positions and on the derived allele-specific motifs. As more and more experimental MHC peptide-binding data became available, more tools have been developed using a wide variety of statistical and computational techniques (see Fig. 1 for an overview). The so-called matrix-based methods use position-specific scoring matrices to determine if a peptide sequence matches the binding motif of particular MHC allele. Another class of MHC binding prediction methods use machine learning techniques, such as Artificial Neural Networks or Support Vector Machines (see Fig. 1). The performance of these algorithms strongly depends on quantity and quality of the available training dataset for each allele model (e.g. "HLA-A*02:01", "H2-Db" etc.) to "learn" underlying patterns /features, which have prediction capability for binding. Recently, structure-based methods are emerging, which circumvent the bottleneck of having a large training set, as they solely rely on peptide-MHC crystal structures and scoring functions (e.g. different energy functions) to predict peptide-MHC interactions by, e.g. energy minimization (see Fig. 1). However, the accuracy of those approaches is still far behind the sequence-based methods. Benchmarking studies shows that the artificial neural network based tool NetMHC (C. Lundegaard et al., Nucleic Acids Research 36, W509 (2008)) and the matrix based algorithm SMM (B. Peters, A. Sette, BMC bioinformatics 6, 132 (2005)) perform best on the tested evaluation data (B. Peters, A. Sette, BMC bioinformatics

6, 132 (2005); H. H. Lin, S. Ray, S. Tongchusak, E. L. Reinherz, V. Brusic, BMC immunology 9, 8 (2008)). Both approaches are integrated in the so-called IEDB consensus methods, available at the Immune Epitope Database (Y. Kim et al., Nucleic Acids Research 40, W525 (2012)). Modeling interactions of peptide-MHC II binding is far more complex than for MHC I, as MHC II molecules possess a binding groove with open ends at either side, allowing binding of peptides of different lengths. Whereas peptides binding to MHC I is restricted to mainly 8-12 amino acids, this length can differ for MHC II peptides dramatically (9-30 amino acids). A recent benchmarking study shows, that the available MHC II predictions methods offer a limited accuracy compared to MHC I prediction (H. H. Lin, S. Ray, S. Tongchusak, E. L. Reinherz, V. Brusic, BMC immunology 9, 8 (2008)).

The first large-scale and systematic use of those algorithms to find T cell epitopes was undertaken by Moutaftsi *et al.* (M. Moutaftsi et al., Nature Biotechnology 24, 817 (2006)), where different tools were combined to predict possible vaccine candidates of vaccinia virus infected C57BL76 mice, extracted spleenocytes and measured CD8+ T cell responses against the top 1% of the predicted peptides. They identified 49 (out of 2256) peptides, that induced a T cell response. Since then many studies have been published using various MHC binding prediction tools to search for T cell epitopes as candidates for a vaccine, mainly for pathogens, e.g., *Leishmania major* (C. Herrera-Najera, R. Pifia-Aguilar, F. Xacur-Garcia, M. J. Ramirez-Sierra, E. Dumonteil, Proteomics 9, 1293 (2009)). However, to use solely MHC I binding prediction tools for prediction of immunogenicity is misleading, as those tools are trained to predict whether a given peptide has the potential to bind to a given MHC allele. The rationale of using MHC binding predictions for predicting immunogenicity is the assumption that peptides binding with high affinity a respective MHC allele is more likely to be immunogenic (A. Sette et al., The Journal of Immunology 153, 5586 (1994)). However, there are numerous studies indicating, that also low MHC binding affinity can result in high immunogenicity (M. C. Feltkamp, M. P. Vierboom, W. M. Kast, C. J. Melief, Molecular Immunology 31, 1391 (1994)) and that peptide-MHC stability might be a better predictor for immunogenicity than peptide affinity (M. Harndahl et al., European Journal of Immunology 42, 1405 (2012)). For that reason, immunogenicity prediction was not very accurate so far, which is mirrored in the low success rates for predicting immunogenicity. Nevertheless, peptide binding is a necessary but not sufficient condition of T

cell epitope recognition, and efficient prediction can dramatically reduce the number of peptides to be tested experimentally.

It is clear that the development of a model that predicts immunogenicity needs also to take into account the recognition of the T cell receptor (TCR) as well as central tolerance, i.e., the negative and positive selection of T cells during development in the thymus.

There is a need for a predictive model, which is capable to model all the aspects mentioned above to accurately predict immunogenicity of an epitope, rather than only binding.

DESCRIPTION OF INVENTION

SUMMARY OF THE INVENTION

In one aspect, the present invention relates to a method for predicting immunogenic amino acid modifications, the method comprising the steps:

a) ascertaining a score for binding of a modified peptide to one or more MHC molecules, and

b) ascertaining a score for binding of the non-modified peptide to one or more MHC molecules, and/or

c) ascertaining a score for binding of the modified peptide when present in a MHC-peptide complex to one or more T cell receptors.

In one embodiment, the modified peptide comprises a fragment of a modified protein, said fragment comprising the modification(s) present in the protein. In one embodiment, the non-modified peptide or protein has the germline amino acid at the position(s) corresponding to the position(s) of the modification(s) in the modified peptide or protein.

In one embodiment, the non-modified peptide or protein and modified peptide or protein are identical but for the modification(s). Preferably, the non-modified peptide or protein and

modified peptide or protein have the same length/or and sequence (except for the modification(s)).

In one embodiment, the non-modified peptide and modified peptide are 8 to 15, preferably 8 to 12 amino acids in length.

In one embodiment, the one or more MHC molecules comprise different MHC molecule types, in particular different MHC alleles. In one embodiment, the one or more MHC molecules are MHC class I molecules and/or MHC class II molecules. In one embodiment, the one or more MHC molecules comprise a set of MHC alleles such as a set of MHC alleles of an individual or a subset thereof.

In one embodiment, the score for binding to one or more MHC molecules is ascertained by a process comprising a sequence comparison with a database of MHC-binding motifs.

In one embodiment, step a) comprises ascertaining whether said score satisfies a pre-determined threshold for binding to one or more MHC molecules and/or step b) comprises ascertaining whether said score satisfies a pre-determined threshold for binding to one or more MHC molecules. In one embodiment, the threshold applied in step a) is different to the threshold applied in step b). In one embodiment, the pre-determined threshold for binding to one or more MHC molecules reflects a probability for binding to one or more MHC molecules.

In one embodiment, the one or more T cell receptors comprise a set of T cell receptors such a set of T cell receptors of an individual or a subset thereof. In one embodiment, step c) comprises assuming that said set of **T** cell receptors does not include **T** cell receptors which bind to the non-modified peptide when present in a MHC-peptide complex and/or does not include T cell receptors which bind to the non-modified peptide when present in a MHC-peptide complex with high affinity.

In one embodiment, step c) comprises ascertaining a score for the chemical and physical similarities between the non-modified and modified amino acids. In one embodiment, step c)

comprises ascertaining whether said score satisfies a pre-determined threshold for the chemical and physical similarities between amino acids. In one embodiment, said pre-determined threshold for the chemical and physical similarities between amino acids reflects a probability for amino acids being chemically and physically similar. In one embodiment, the score for the chemical and physical similarities is ascertained on the basis of the probability of amino acids being interchanged in nature. In one embodiment, the more frequently amino acids are interchanged in nature the more similar the amino acids are considered and *vice versa.* In one embodiment, the chemical and physical similarities are determined using evolutionary based log-odds matrices,

In one embodiment, if the non-modified peptide has a score for binding to one or more MHC molecules satisfying a threshold indicating binding to one or more MHC molecules and the modified peptide has a score for binding to one or more MHC molecules satisfying a threshold indicating binding to one or more MHC molecules, the modification or modified peptide is predicted as immunogenic if the non-modified and modified amino acids have a score for the chemical and physical similarities satisfying a threshold indicating chemical and physical dissimilarity.

In one embodiment, if the non-modified peptide binds to one or more MHC molecules or has a probability for binding to one or more MHC molecules and the modified peptide binds to one or more MHC molecules or has a probability for binding to one or more MHC molecules, the modification or modified peptide is predicted as immunogenic if the non-modified and modified amino acids are chemically and physically dissimilar or have a probability of being chemically and physically dissimilar.

In one embodiment, the modification is not in an anchor position for binding to one or more MHC molecules.

In one embodiment, if the non-modified peptide has a score for binding to one or more MHC molecules satisfying a threshold indicating no binding to one or more MHC molecules and the modified peptide has a score for binding to one or more MHC molecules satisfying a threshold

indicating binding to one or more MHC molecules, the modification or modified peptide is predicted as immunogenic.

In one embodiment, if the non-modified peptide does not bind to one or more MHC molecules or has a probability for not binding to one or more MHC molecules and the modified peptide binds to one or more MHC molecules or has a probability for binding to one or more MHC molecules, the modification or modified peptide is predicted as immunogenic.

In one embodiment, the modification is in an anchor position for binding to one or more MHC molecules.

In one embodiment, the method of the invention comprises performing step a) on two or more different modified peptides, said two or more different modified peptides comprising the same modification(s). In one embodiment, the two or more different modified peptides comprising the same modification(s) comprise different fragments of a modified protein, said different fragments comprising the same modification(s) present in the protein. In one embodiment, the two or more different modified peptides comprising the same modification(s) comprise all potential MHC binding fragments of a modified protein, said fragments comprising the same modification(s) present in the protein. In one embodiment, the method of the invention further comprises selecting (the) modified peptide(s) from the two or more different modified peptides comprising the same modification(s) having a probability or having the highest probability for binding to one or more MHC molecules. In one embodiment, the two or more different modified peptides comprising the same modification(s) differ in length and/or position of the modification(s).

In one embodiment, the method of the invention comprises performing step a) and optionally one or both of steps b) and c) on two or more different modified peptides. In one embodiment, said two or more different modified peptides comprise the same modification(s) and/or comprise different modifications. In one embodiment, the different modifications are present in the same and/or in different proteins. The set of two or more different modified peptides used in step a) and optionally one or both of steps b) and c) may be the same or different. In one embodiment,

the set of two or more different modified peptides used in step b) and/or step c) is a subset of the set of two or more different modified peptides used in step a). Preferably, said subset includes the peptide(s) scoring best in step a).

In one embodiment, the method of the invention comprises comparing the scores of two or more of said different modified peptides. In one embodiment, the method of the invention comprises ranking two or more of said different modified peptides. In one embodiment, a score for binding of the modified peptide to one or more MHC molecules is weighted higher than a score for binding of the modified peptide when present in a MHC-peptide complex to one or more T cell receptors, preferably a score for the chemical and physical similarities between the non-modified and modified amino acids and a score for binding of the modified peptide when present in a MHC-peptide complex to one or more T cell receptors, preferably a score for the chemical and physical similarities between the non-modified and modified amino acids is weighted higher than a score for binding of the non-modified peptide to one or more MHC molecules.

In one embodiment, the method of the invention further comprises identifying non-synonymous mutations in one or more protein-coding regions.

In one embodiment, modifications are identified according to the invention by partially or completely sequencing the genome or transcriptome of one or more cells such as one or more cancer cells and optionally one or more non-cancerous cells and identifying mutations in one or more protein-coding regions.

In one embodiment, said mutations are somatic mutations. In one embodiment, said mutations are cancer mutations.

In one embodiment, the method of the invention is used in the manufacture of a vaccine. In one embodiment, the vaccine is derived from (a) modification(s) or (a) modified peptide(s) predicted as immunogenic by the methods of the invention.

In a further aspect, the present invention provides a method for providing a vaccine comprising the step:
identifying (a) modification(s) or (a) modified peptide(s) predicted as immunogenic by the methods of the invention.

In one embodiment, the method further comprises the step:
providing a vaccine comprising a peptide or polypeptide comprising the modification(s) or modified peptide(s) predicted as immunogenic, or a nucleic acid encoding the peptide or polypeptide.

In a further aspect, the present invention provides a vaccine which is obtainable using the methods according to the invention. Preferred embodiments of such vaccines are described herein.

A vaccine provided according to the invention may comprise a pharmaceutically acceptable carrier and may optionally comprise one or more adjuvants, stabilizers etc. The vaccine may in the form of a therapeutic or prophylactic vaccine.

Another aspect relates to a method for inducing an immune response in a patient, comprising administering to the patient a vaccine provided according to the invention.

Another aspect relates to a method of treating a cancer patient comprising the steps:
(a) providing a vaccine by the methods according to the invention; and
(b) administering said vaccine to the patient.

Another aspect relates to a method of treating a cancer patient comprising administering the vaccine according to the invention to the patient.

In further aspects, the invention provides the vaccines described herein for use in the methods of treatment described herein, in particular for use in treating or preventing cancer.

The treatments of cancer described herein can be combined with surgical resection and/or radiation and/or traditional chemotherapy.

Other features and advantages of the instant invention will be apparent from the following detailed description and claims.

DETAILED DESCRIPTION OF THE INVENTION

Although the present invention is described in detail below, it is to be understood that this invention is not limited to the particular methodologies, protocols and reagents described herein as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims. Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art.

In the following, the elements of the present invention will be described. These elements are listed with specific embodiments, however, it should be understood that they may be combined in any manner and in any number to create additional embodiments. The variously described examples and preferred embodiments should not be construed to limit the present invention to only the explicitly described embodiments. This description should be understood to support and encompass embodiments which combine the explicitly described embodiments with any number of the disclosed and/or preferred elements. Furthermore, any permutations and combinations of all described elements in this application should be considered disclosed by the description of the present application unless the context indicates otherwise.

Preferably, the terms used herein are defined as described in "A multilingual glossary of biotechnological terms: (IUPAC Recommendations)", *H.G. W. Leuenberger, B. Nagel, and H. Kolbl, Eds., (1995) Helvetica Chimica Acta, CH-4010 Basel, Switzerland.*

The practice of the present invention will employ, unless otherwise indicated, conventional methods of biochemistry, cell biology, immunology, and recombinant DNA techniques which are explained in the literature in the field (cf, e.g., *Molecular Cloning: A Laboratory Manual*, 2nd Edition, J. Sambrook et al. eds., Cold Spring Harbor Laboratory Press, Cold Spring Harbor 1989).

Throughout this specification and the claims which follow, unless the context requires otherwise, the word "comprise", and variations such as "comprises" and "comprising", will be understood to imply the inclusion of a stated member, integer or step or group of members, integers or steps but not the exclusion of any other member, integer or step or group of members, integers or steps although in some embodiments such other member, integer or step or group of members, integers or steps may be excluded, i.e. the subject-matter consists in the inclusion of a stated member, integer or step or group of members, integers or steps. The terms "a" and "an" and "the" and similar reference used in the context of describing the invention (especially in the context of the claims) are to be construed to cover both the singular and the plural, unless otherwise indicated herein or clearly contradicted by context. Recitation of ranges of values herein is merely intended to serve as a shorthand method of referring individually to each separate value falling within the range. Unless otherwise indicated herein, each individual value is incorporated into the specification as if it were individually recited herein.

All methods described herein can be performed in any suitable order unless otherwise indicated herein or otherwise clearly contradicted by context. The use of any and all examples, or exemplary language (e.g., "such as"), provided herein is intended merely to better illustrate the invention and does not pose a limitation on the scope of the invention otherwise claimed. No language in the specification should be construed as indicating any non-claimed element essential to the practice of the invention.

Several documents are cited throughout the text of this specification. Each of the documents cited herein (including all patents, patent applications, scientific publications, manufacturer's specifications, instructions, etc.), whether *supra* or *infra*, are hereby incorporated by reference in

their entirety. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

According to the present invention, the term "peptide" refers to substances comprising two or more, preferably 3 or more, preferably 4 or more, preferably 6 or more, preferably 8 or more, preferably 10 or more, preferably 13 or more, preferably 16 more, preferably 21 or more and up to preferably 8, 10, 20, 30, 40 or 50, in particular 100 amino acids joined covalently by peptide bonds. The term "polypeptide" or "protein" refers to large peptides, preferably to peptides with more than 100 amino acid residues, but in general the terms "peptide", "polypeptide" and "protein" are synonyms and are used interchangeably herein.

According to the invention, the term "modification" with respect to peptides, polypeptides or proteins relates to a sequence change in a peptide, polypeptide or protein compared to a parental sequence such as the sequence of a wildtype peptide, polypeptide or protein. The term includes amino acid insertion variants, amino acid addition variants, amino acid deletion variants and amino acid substitution variants, preferably amino acid substitution variants. All these sequence changes according to the invention may potentially create new epitopes.

Amino acid insertion variants comprise insertions of single or two or more amino acids in a particular amino acid sequence.

Amino acid addition variants comprise amino- and/or carboxy-terminal fusions of one or more amino acids, such as 1, 2, 3, 4 or 5, or more amino acids.

Amino acid deletion variants are characterized by the removal of one or more amino acids from the sequence, such as by removal of 1, 2, 3, 4 or 5, or more amino acids.

Amino acid substitution variants are characterized by at least one residue in the sequence being removed and another residue being inserted in its place.

According to the invention, a modification or modified peptide used for testing in the methods of the invention may be derived from a protein comprising a modification.

The term "derived" means according to the invention that a particular entity, in particular a particular peptide sequence, is present in the object from which it is derived. In the case of amino acid sequences, especially particular sequence regions, "derived" in particular means that the relevant amino acid sequence is derived from an amino acid sequence in which it is present.

A protein comprising a modification from which a modification or modified peptide used for testing in the methods of the invention may be derived may be a neoantigen.

According to the invention, the term "neoantigen" relates to a peptide or protein including one or more amino acid modifications compared to the parental peptide or protein. For example, the neoantigen may be a tumor-associated neoantigen, wherein the term "tumor-associated neoantigen" includes a peptide or protein including amino acid modifications due to tumor-specific mutations.

According to the invention, the term "tumor-specific mutation" or "cancer-specific mutation" relates to a somatic mutation that is present in the nucleic acid of a tumor or cancer cell but absent in the nucleic acid of a corresponding normal, i.e. non-tumorous or non-cancerous, cell. The terms "tumor-specific mutation" and "tumor mutation" and the terms "cancer-specific mutation" and "cancer mutation" are used interchangeably herein.

The term "immune response" refers to an integrated bodily response to a target such as an antigen and preferably refers to a cellular immune response or a cellular as well as a humoral immune response. The immune response may be protective/preventive/prophylactic and/or therapeutic.

"Inducing an immune response" may mean that there was no immune response before induction, but it may also mean that there was a certain level of immune response before induction and after induction said immune response is enhanced. Thus, "inducing an immune response" also

includes "enhancing an immune response". Preferably, after inducing an immune response in a subject, said subject is protected from developing a disease such as a cancer disease or the disease condition is ameliorated by inducing an immune response. For example, an immune response against a tumor-expressed antigen may be induced in a patient having a cancer disease or in a subject being at risk of developing a cancer disease. Inducing an immune response in this case may mean that the disease condition of the subject is ameliorated, that the subject does not develop metastases, or that the subject being at risk of developing a cancer disease does not develop a cancer disease.

The terms "cellular immune response" and "cellular response" or similar terms refer to an immune response directed to cells characterized by presentation of an antigen with class I or class II MHC involving T cells or T-lymphocytes which act as either "helpers" or "killers". The helper T cells (also termed $CD4^+$ T cells) play a central role by regulating the immune response and the killer cells (also termed cytotoxic T cells, cytolytic T cells, $CD8^+$ T cells or CTLs) kill diseased cells such as cancer cells, preventing the production of more diseased cells. In preferred embodiments, the present invention involves the stimulation of an anti-tumor CTL response against tumor cells expressing one or more tumor-expressed antigens and preferably presenting such tumor-expressed antigens with class I MHC.

An "antigen" according to the invention covers any substance, preferably a peptide or protein, that is a target of and/or induces an immune response such as a specific reaction with antibodies or T-lymphocytes (T cells). Preferably, an antigen comprises at least one epitope such as a T cell epitope. Preferably, an antigen in the context of the present invention is a molecule which, optionally after processing, induces an immune reaction, which is preferably specific for the antigen (including cells expressing the antigen). The antigen or a T cell epitope thereof is preferably presented by a cell, preferably by an antigen presenting cell which includes a diseased cell, in particular a cancer cell, in the context of MHC molecules, which results in an immune response against the antigen (including cells expressing the antigen).

In one embodiment, an antigen is a tumor antigen (also termed tumor-expressed antigen herein), i.e., a part of a tumor cell such as a protein or peptide expressed in a tumor cell which may be

derived from the cytoplasm, the cell surface or the cell nucleus, in particular those which primarily occur intracellularly or as surface antigens of tumor cells. For example, tumor antigens include the carcinoembryonal antigen, a 1-fetoprotein, isoferritin, and fetal sulphoglycoprotein, a2-H-ferroprotein and γ-fetoprotein. According to the present invention, a tumor antigen preferably comprises any antigen which is expressed in and optionally characteristic with respect to type and/or expression level for tumors or cancers as well as for tumor or cancer cells, i.e. a tumor-associated antigen. In one embodiment, the term "tumor-associated antigen" relates to proteins that are under normal conditions specifically expressed in a limited number of tissues and/or organs or in specific developmental stages, for example, the tumor-associated antigens may be under normal conditions specifically expressed in stomach tissue, preferably in the gastric mucosa, in reproductive organs, e.g., in testis, in trophoblastic tissue, e.g., in placenta, or in germ line cells, and are expressed or aberrantly expressed in one or more tumor or cancer tissues. In this context, "a limited number" preferably means not more than 3, more preferably not more than 2. The tumor antigens in the context of the present invention include, for example, differentiation antigens, preferably cell type specific differentiation antigens, i.e., proteins that are under normal conditions specifically expressed in a certain cell type at a certain differentiation stage, cancer/testis antigens, i.e., proteins that are under normal conditions specifically expressed in testis and sometimes in placenta, and germ line specific antigens. Preferably, the tumor antigen or the aberrant expression of the tumor antigen identifies cancer cells. In the context of the present invention, the tumor antigen that is expressed by a cancer cell in a subject, e.g., a patient suffering from a cancer disease, is preferably a self-protein in said subject. In preferred embodiments, the tumor antigen in the context of the present invention is expressed under normal conditions specifically in a tissue or organ that is non-essential, i.e., tissues or organs which when damaged by the immune system do not lead to death of the subject, or in organs or structures of the body which are not or only hardly accessible by the immune system.

According to the invention, the terms "tumor antigen", "tumor-expressed antigen", "cancer antigen" and "cancer-expressed antigen" are equivalents and are used interchangeably herein.

The term "immunogenicity" relates to the relative effectivity to induce an immune response that

is preferably associated with therapeutic treatments, such as treatments against cancers. As used herein, the term "immunogenic" relates to the property of having immunogenicity. For example, the term "immunogenic modification" when used in the context of a peptide, polypeptide or protein relates to the effectivity of said peptide, polypeptide or protein to induce an immune response that is caused by and/or directed against said modification. Preferably, the non-modified peptide, polypeptide or protein does not induce an immune response, induces a different immune response or induces a different level, preferably a lower level, of immune response.

The terms "major histocompatibility complex" and the abbreviation "MHC" include MHC class I and MHC class II molecules and relate to a complex of genes which occurs in all vertebrates. MHC proteins or molecules are important for signaling between lymphocytes and antigen presenting cells or diseased cells in immune reactions, wherein the MHC proteins or molecules bind peptides and present them for recognition by T cell receptors. The proteins encoded by the MHC are expressed on the surface of cells, and display both self antigens (peptide fragments from the cell itself) and non-self antigens (e.g., fragments of invading microorganisms) to a T cell.

The MHC region is divided into three subgroups, class I, class II, and class III. MHC class I proteins contain an a-chain and p2-microglobulin (not part of the MHC encoded by chromosome 15). They present antigen fragments to cytotoxic T cells. On most immune system cells, specifically on antigen-presenting cells, MHC class II proteins contain a- and β-chains and they present antigen fragments to T-helper cells. MHC class III region encodes for other immune components, such as complement components and some that encode cytokines.

The MHC is both polygenic (there are several MHC class I and MHC class II genes) and polymorphic (there are multiple alleles of each gene).

As used herein, the term "haplotype" refers to the HLA alleles found on one chromosome and the proteins encoded thereby. Haplotype may also refer to the allele present at any one locus within the MHC. Each class of MHC is represented by several loci: e.g., HLA-A (Human Leukocyte

Antigen-A), HLA-B, HLA-C, HLA-E, HLA-F, HLA-G, HLA-H, HLA-J, HLA-K, HLA-L, HLA-P and HLA-V for class I and HLA-DRA, HLA-DRB1-9, HLA-, HLA-DQA1, HLA-DQB1, HLA-DPA1 , HLA-DPB1, HLA-DMA, HLA-DMB, HLA-DOA, and HLA-DOB for class II. The terms "HLA allele" and "MHC allele" are used interchangeably herein.

The MHCs exhibit extreme polymorphism: within the human population there are, at each genetic locus, a great number of haplotypes comprising distinct alleles. Different polymorphic MHC alleles, of both class I and class II, have different peptide specificities: each allele encodes proteins that bind peptides exhibiting particular sequence patterns.

In one preferred embodiment of all aspects of the invention an MHC molecule is an HLA molecule.

In the context of the present invention, the term "MHC binding peptide" includes MHC class I and/or class II binding peptides or peptides that can be processed to produce MHC class I and/or class II binding peptides. In the case of class I MHC/peptide complexes, the binding peptides are typically 8-12, preferably 8-10 amino acids long although longer or shorter peptides may be effective. In the case of class II MHC/peptide complexes, the binding peptides are typically 9-30, preferably 10-25 amino acids long and are in particular 13-18 amino acids long, whereas longer and shorter peptides may be effective.

If a peptide is to be presented directly, i.e., without processing, in particular without cleavage, it has a length which is suitable for binding to an MHC molecule, in particular a class I MHC molecule, and preferably is 7-30 amino acids in length such as 7-20 amino acids in length, more preferably 7-12 amino acids in length, more preferably 8-1 1 amino acids in length, in particular 9 or 10 amino acids in length.

If a peptide is part of a larger entity comprising additional sequences, e.g. of a vaccine sequence or polypeptide, and is to be presented following processing, in particular following cleavage, the peptide produced by processing has a length which is suitable for binding to an MHC molecule, in particular a class I MHC molecule, and preferably is 7-30 amino acids in length such as 7-20

amino acids in length, more preferably 7-12 amino acids in length, more preferably 8-1 1 amino acids in length, in particular 9 or 10 amino acids in length. Preferably, the sequence of the peptide which is to be presented following processing is derived from the amino acid sequence of an antigen or polypeptide used for vaccination, i.e., its sequence substantially corresponds and is preferably completely identical to a fragment of the antigen or polypeptide.

Thus, an MHC binding peptide in one embodiment comprises a sequence which substantially corresponds and is preferably completely identical to a fragment of an antigen.

The term "epitope" refers to an antigenic determinant in a molecule such as an antigen, i.e., to a part in or fragment of the molecule that is recognized by the immune system, for example, that is recognized by a T cell, in particular when presented in the context of MHC molecules. An epitope of a protein such as a tumor antigen preferably comprises a continuous or discontinuous portion of said protein and is preferably between 5 and 100, preferably between 5 and 50, more preferably between 8 and 30, most preferably between 10 and 25 amino acids in length, for example, the epitope may be preferably 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 amino acids in length. It is particularly preferred that the epitope in the context of the present invention is a T cell epitope.

According to the invention an epitope may bind to MHC molecules such as MHC molecules on the surface of a cell and thus, may be a "MHC binding peptide".

As used herein the term "neo-epitope" refers to an epitope that is not present in a reference such as a normal non-cancerous or germline cell but is found in cancer cells. This includes, in particular, situations wherein in a normal non-cancerous or germline cell a corresponding epitope is found, however, due to one or more mutations in a cancer cell the sequence of the epitope is changed so as to result in the neo-epitope.

As used herein, the term "T cell epitope" refers to a peptide which binds to a MHC molecule in a configuration recognized by a T cell receptor. Typically, T cell epitopes are presented on the surface of an antigen-presenting cell.

As used herein, the term "predicting T cell epitopes" refers to a prediction whether a peptide will bind to a MHC molecule and will be recognized by a T cell receptor. The term "predicting T cell epitopes" is essentially synonymous with the phrase "predicting whether a peptide is immunogenic".

According to the invention, a T cell epitope may be present in a vaccine as a part of a larger entity such as a vaccine sequence and/or a polypeptide comprising more than one T cell epitope. The presented peptide or T cell epitope is produced following suitable processing.

T cell epitopes may be modified at one or more residues that are not essential for TCR recognition or for binding to MHC. Such modified T cell epitopes may be considered immunologically equivalent.

Preferably a T cell epitope when presented by MHC and recognized by a T cell receptor is able to induce in the presence of appropriate co-stimulatory signals, clonal expansion of the T cell carrying the T cell receptor specifically recognizing the peptide/MHC-complex.

Preferably, a T cell epitope comprises an amino acid sequence substantially corresponding to the amino acid sequence of a fragment of an antigen. Preferably, said fragment of an antigen is an MHC class I and/or class II presented peptide.

A T cell epitope according to the invention preferably relates to a portion or fragment of an antigen which is capable of stimulating an immune response, preferably a cellular response against the antigen or cells characterized by expression of the antigen and preferably by presentation of the antigen such as diseased cells, in particular cancer cells. Preferably, a T cell epitope is capable of stimulating a cellular response against a cell characterized by presentation of an antigen with class I MHC and preferably is capable of stimulating an antigen-responsive cytotoxic T-lymphocyte (CTL).

"Antigen processing" or "processing" refers to the degradation of a peptide, polypeptide or

protein into procession products, which are fragments of said peptide, polypeptide or protein (e.g., the degradation of a polypeptide into peptides) and the association of one or more of these fragments (e.g., via binding) with MHC molecules for presentation by cells, preferably antigen presenting cells, to specific T cells.

"Antigen presenting cells" (APC) are cells which present peptide fragments of protein antigens in association with MHC molecules on their cell surface. Some APCs may activate antigen specific T cells.

Professional antigen-presenting cells are very efficient at internalizing antigen, either by phagocytosis or by receptor-mediated endocytosis, and then displaying a fragment of the antigen, bound to a class II MHC molecule, on their membrane. The T cell recognizes and interacts with the antigen-class II MHC molecule complex on the membrane of the antigen-presenting cell. An additional co-stimulatory signal is then produced by the antigen-presenting cell, leading to activation of the T cell. The expression of co-stimulatory molecules is a defining feature of professional antigen-presenting cells.

The main types of professional antigen-presenting cells are dendritic cells, which have the broadest range of antigen presentation, and are probably the most important antigen-presenting cells, macrophages, B-cells, and certain activated epithelial cells. Dendritic cells (DCs) are leukocyte populations that present antigens captured in peripheral tissues to T cells via both MHC class II and I antigen presentation pathways. It is well known that dendritic cells are potent inducers of immune responses and the activation of these cells is a critical step for the induction of antitumoral immunity. Dendritic cells are conveniently categorized as "immature" and "mature" cells, which can be used as a simple way to discriminate between two well characterized phenotypes. However, this nomenclature should not be construed to exclude all possible intermediate stages of differentiation. Immature dendritic cells are characterized as antigen presenting cells with a high capacity for antigen uptake and processing, which correlates with the high expression of Fey receptor and mannose receptor. The mature phenotype is typically characterized by a lower expression of these markers, but a high expression of cell surface molecules responsible for T cell activation such as class I and class II MHC, adhesion

molecules (e. g. CD54 and CD1 1) and costimulatory molecules (e. g., CD40, CD80, CD86 and 4-1 BB). Dendritic cell maturation is referred to as the status of dendritic cell activation at which such antigen-presenting dendritic cells lead to T cell priming, while presentation by immature dendritic cells results in tolerance. Dendritic cell maturation is chiefly caused by biomolecules with microbial features detected by innate receptors (bacterial DNA, viral RNA, endotoxin, etc.), pro-inflammatory cytokines (TNF, IL-1, IFNs), ligation of CD40 on the dendritic cell surface by CD40L, and substances released from cells undergoing stressful cell death. The dendritic cells can be derived by culturing bone marrow cells *in vitro* with cytokines, such as granulocyte-macrophage colony-stimulating factor (GM-CSF) and tumor necrosis factor alpha.

Non-professional antigen-presenting cells do not constitutively express the MHC class II proteins required for interaction with naive T cells; these are expressed only upon stimulation of the non-professional antigen-presenting cells by certain cytokines such as IFNy.

Antigen presenting cells can be loaded with MHC class I presented peptides by transducing the cells with nucleic acid, preferably RNA, encoding a peptide or polypeptide comprising the peptide to be presented, e.g. a nucleic acid encoding an antigen or polypeptide used for vaccination.

In some embodiments, a pharmaceutical composition or vaccine comprising a nucleic acid delivery vehicle that targets a dendritic or other antigen presenting cell may be administered to a patient, resulting in transfection that occurs *in vivo. In vivo* transfection of dendritic cells, for example, may generally be performed using any methods known in the art, such as those described in WO 97/24447, or the gene gun approach described by Mahvi et al., Immunology and cell Biology 75: 456-460, 1997.

According to the invention, the term "antigen presenting cell" also includes target cells.

"Target cell" shall mean a cell which is a target for an immune response such as a cellular immune response. Target cells include cells that present an antigen, i.e. a peptide fragment derived from an antigen, and include any undesirable cell such as a cancer cell. In preferred

embodiments, the target cell is a cell expressing an antigen as described herein and preferably presenting said antigen with class I MHC.

The term "portion" refers to a fraction. With respect to a particular structure such as an amino acid sequence or protein the term "portion" thereof may designate a continuous or a discontinuous fraction of said structure. Preferably, a portion of an amino acid sequence comprises at least 1%, at least 5%, at least 10%, at least 20%, at least 30%, preferably at least 40%, preferably at least 50%, more preferably at least 60%, more preferably at least 70%, even more preferably at least 80%, and most preferably at least 90% of the amino acids of said amino acid sequence. Preferably, if the portion is a discontinuous fraction said discontinuous fraction is composed of 2, 3, 4, 5, 6, 7, 8, or more parts of a structure, each part being a continuous element of the structure. For example, a discontinuous fraction of an amino acid sequence may be composed of 2, 3, 4, 5, 6, 7, 8, or more, preferably not more than 4 parts of said amino acid sequence, wherein each part preferably comprises at least 5 continuous amino acids, at least 10 continuous amino acids, preferably at least 20 continuous amino acids, preferably at least 30 continuous amino acids of the amino acid sequence.

The terms "part" and "fragment" are used interchangeably herein and refer to a continuous element. For example, a part of a structure such as an amino acid sequence or protein refers to a continuous element of said structure. A portion, a part or a fragment of a structure preferably comprises one or more functional properties of said structure. For example, a portion, a part or a fragment of an epitope, peptide or protein is preferably immunologically equivalent to the epitope, peptide or protein it is derived from. In the context of the present invention, a "part" of a structure such as an amino acid sequence preferably comprises, preferably consists of at least 10%, at least 20%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 92%, at least 94%, at least 96%, at least 98%, at least 99% of the entire structure or amino acid sequence.

The term "immunoreactive cell" in the context of the present invention relates to a cell which exerts effector functions during an immune reaction. An "immunoreactive cell" preferably is capable of binding an antigen or a cell characterized by presentation of an antigen or a peptide

fragment thereof (e.g. a T cell epitope) and mediating an immune response. For example, such cells secrete cytokines and/or chemokines, secrete antibodies, recognize cancerous cells, and optionally eliminate such cells. For example, immunoreactive cells comprise T cells (cytotoxic T cells, helper T cells, tumor infiltrating T cells), B cells, natural killer cells, neutrophils, macrophages, and dendritic cells. Preferably, in the context of the present invention, "immunoreactive cells" are T cells, preferably $CD4^+$ and/or $CD8^+$ T cells.

Preferably, an "immunoreactive cell" recognizes an antigen or a peptide fragment thereof with some degree of specificity, in particular if presented in the context of MHC molecules such as on the surface of antigen presenting cells or diseased cells such as cancer cells. Preferably, said recognition enables the cell that recognizes an antigen or a peptide fragment thereof to be responsive or reactive. If the cell is a helper T cell ($CD4^+$ T cell) bearing receptors that recognize an antigen or a peptide fragment thereof in the context of MHC class II molecules such responsiveness or reactivity may involve the release of cytokines and/or the activation of $CD8^+$ lymphocytes (CTLs) and/or B-cells. If the cell is a CTL such responsiveness or reactivity may involve the elimination of cells presented in the context of MHC class I molecules, i.e., cells characterized by presentation of an antigen with class I MHC, for example, via apoptosis or perforin-mediated cell lysis. According to the invention, CTL responsiveness may include sustained calcium flux, cell division, production of cytokines such as IFN-γ and TNF-a, up-regulation of activation markers such as CD44 and CD69, and specific cytolytic killing of antigen expressing target cells. CTL responsiveness may also be determined using an artificial reporter that accurately indicates CTL responsiveness. Such CTL that recognize an antigen or an antigen fragment and are responsive or reactive are also termed "antigen-responsive CTL" herein. If the cell is a B cell such responsiveness may involve the release of immunoglobulins.

The terms "T cell" and "T lymphocyte" are used interchangeably herein and include T helper cells (CD4+ T cells) and cytotoxic T cells (CTLs, CD8+ T cells) which comprise cytolytic T cells.

T cells belong to a group of white blood cells known as lymphocytes, and play a central role in cell-mediated immunity. They can be distinguished from other lymphocyte types, such as B cells

and natural killer cells by the presence of a special receptor on their cell surface called T cell receptor (TCR). The thymus is the principal organ responsible for the maturation of T cells. Several different subsets of T cells have been discovered, each with a distinct function.

T helper cells assist other white blood cells in immunologic processes, including maturation of **B** cells into plasma cells and activation of cytotoxic T cells and macrophages, among other functions. These cells are also known as CD4+ T cells because they express the CD4 protein on their surface. Helper T cells become activated when they are presented with peptide antigens by MHC class II molecules that are expressed on the surface of antigen presenting cells (APCs). Once activated, they divide rapidly and secrete small proteins called cytokines that regulate or assist in the active immune response.

Cytotoxic T cells destroy virally infected cells and tumor cells, and are also implicated in transplant rejection. These cells are also known as CD8+ T cells since they express the CD8 glycoprotein at their surface. These cells recognize their targets by binding to antigen associated with MHC class I, which is present on the surface of nearly every cell of the body.

A majority of T cells have a T cell receptor (TCR) existing as a complex of several proteins. The actual T cell receptor is composed of two separate peptide chains, which are produced from the independent T cell receptor alpha and beta (TCRa and TCRP) genes and are called a- and β-TCR chains. γδ T cells (gamma delta T cells) represent a small subset of T cells that possess a distinct T cell receptor (TCR) on their surface. However, in γδ T cells, the TCR is made up of one γ-chain and one δ-chain. This group of T cells is much less common **(2%** of total T cells) than the αβ T cells.

The first signal in activation of T cells is provided by binding of the T cell receptor to a short peptide presented by the MHC on another cell. This ensures that only a T cell with a TCR specific to that peptide is activated. The partner cell is usually an antigen presenting cell such as a professional antigen presenting cell, usually a dendritic cell in the case of naive responses, although **B** cells and macrophages can be important APCs.

According to the present invention, a molecule is capable of binding to a target if it has a significant affinity for said predetermined target and binds to said predetermined target in standard assays. "Affinity" or "binding affinity" is often measured by equilibrium dissociation constant (K_D). A molecule is not (substantially) capable of binding to a target if it has no significant affinity for said target and does not bind significantly to said target in standard assays.

Cytotoxic T lymphocytes may be generated *in vivo* by incorporation of an antigen or a peptide fragment thereof into antigen-presenting cells *in vivo.* The antigen or a peptide fragment thereof may be represented as protein, as DNA (e.g. within a vector) or as RNA. The antigen may be processed to produce a peptide partner for the MHC molecule, while a fragment thereof may be presented without the need for further processing. The latter is the case in particular, if these can bind to MHC molecules. In general, administration to a patient by intradermal injection is possible. However, injection may also be carried out intranodally into a lymph node (Maloy et al. (2001), Proc Natl Acad Sci USA PS:3299-303). The resulting cells present the complex of interest and are recognized by autologous cytotoxic T lymphocytes which then propagate.

Specific activation of CD4+ or CD8+ T cells may be detected in a variety of ways. Methods for detecting specific T cell activation include detecting the proliferation of T cells, the production of cytokines (e.g., lymphokines), or the generation of cytolytic activity. For CD4+ T cells, a preferred method for detecting specific T cell activation is the detection of the proliferation of T cells. For CD8+ T cells, a preferred method for detecting specific T cell activation is the detection of the generation of cytolytic activity.

By "cell characterized by presentation of an antigen" or "cell presenting an antigen" or similar expressions is meant a cell such as a diseased cell, e.g. a cancer cell, or an antigen presenting cell presenting the antigen it expresses or a fragment derived from said antigen, e.g. by processing of the antigen, in the context of MHC molecules, in particular MHC Class I molecules. Similarly, the terms "disease characterized by presentation of an antigen" denotes a disease involving cells characterized by presentation of an antigen, in particular with class I MHC. Presentation of an antigen by a cell may be effected by transfecting the cell with a nucleic acid such as RNA encoding the antigen.

By "fragment of an antigen which is presented" or similar expressions is meant that the fragment can be presented by MHC class I or class II, preferably MHC class I, e.g. when added directly to antigen presenting cells. In one embodiment, the fragment is a fragment which is naturally presented by cells expressing an antigen.

The term "immunologically equivalent" means that the immunologically equivalent molecule such as the immunologically equivalent amino acid sequence exhibits the same or essentially the same immunological properties and/or exerts the same or essentially the same immunological effects, e.g., with respect to the type of the immunological effect such as induction of a humoral and/or cellular immune response, the strength and/or duration of the induced immune reaction, or the specificity of the induced immune reaction. In the context of the present invention, the term "immunologically equivalent" is preferably used with respect to the immunological effects or properties of a peptide used for immunization. For example, an amino acid sequence is immunologically equivalent to a reference amino acid sequence if said amino acid sequence when exposed to the immune system of a subject induces an immune reaction having a specificity of reacting with the reference amino acid sequence.

The term "immune effector functions" in the context of the present invention includes any functions mediated by components of the immune system that result, for example, in the killing of tumor cells, or in the inhibition of tumor growth and/or inhibition of tumor development, including inhibition of tumor dissemination and metastasis. Preferably, the immune effector functions in the context of the present invention are T cell mediated effector functions. Such functions comprise in the case of a helper T cell ($CD4^+$ T cell) the recognition of an antigen or an antigen fragment in the context of MHC class II molecules by T cell receptors, the release of cytokines and/or the activation of $CD8^+$ lymphocytes (CTLs) and/or B-cells, and in the case of CTL the recognition of an antigen or an antigen fragment in the context of MHC class I molecules by T cell receptors, the elimination of cells presented in the context of MHC class I molecules, i.e., cells characterized by presentation of an antigen with class I MHC, for example, via apoptosis or perforin-mediated cell lysis, production of cytokines such as IFN-γ and TNF-a, and specific cytolytic killing of antigen expressing target cells.

According to the invention, the term "score" relates to a result, usually expressed numerically, of a test or examination. Terms such as "score better" or "score best" relate to a better result or the best result of a test or examination.

Terms such as "predict", "predicting" or "prediction" relate to the determination of a likelihood.

According to the invention, ascertaining a score for binding of a peptide to one or more MHC molecules includes determining the likelihood of binding of a peptide to one or more MHC molecules.

A score for binding of a peptide to one or more MHC molecules may be ascertained by using any peptide:MHC binding predictive tools. For example, the immune epitope database analysis resource (IEDB-AR: http://tools.iedb.org) may be used.

Predictions are usually made against a set of MHC molecules such as a set of different MHC alleles such as all possible MHC alleles or a set or subset of MHC alleles found in a patient preferably having the modification(s) the immunogenicity of which is to be determined according to the invention.

According to the invention, ascertaining a score for binding of a modified peptide when present in a MHC-peptide complex to one or more T cell receptors includes determining the likelihood of binding of a peptide when present in a complex with an MHC molecule to T cell receptors.

Predictions may be made against one T cell receptor such as a T cell receptor found in a patient or preferably against a set of T cell receptors such as an unknown set of different T cell receptors or a set or subset of T cell receptors found in a patient preferably having the modiflcation(s) the immunogenicity of which is to be determined according to the invention.

Furthermore, predictions are usually made against a set of MHC molecules such as a set of different MHC alleles such as all possible MHC alleles or a set or subset of MHC alleles found

in a patient preferably having the modification(s) the immunogenicity of which is to be determined according to the invention.

A score for binding of a modified peptide when present in a MHC-peptide complex to one or more T cell receptors may be ascertained by estimating the effect of the modification on the binding of a T cell receptor-peptide-MHC complex given an (unknown) T cell receptor repertoire. The score for binding of a modified peptide when present in a MHC-peptide complex to one or more T cell receptors may generally be defined as a kind of a proxy for the recognition a given peptide-MHC molecule to a matching T cell receptor.

The score for binding of a modified peptide when present in a MHC-peptide complex to one or more T cell receptors may be ascertained by ascertaining the physico-chemical differences between the modified and the non-modified amino acid. For example, substitution matrices may be used. Such matrices describe the rate at which one amino acid in a sequence changes to other amino acid states over time.

For example log-odds matrices such as evolutionary based log-odds matrices may be used: a substitution with a low log odds score has a better chance of finding a matching T cell receptor from the pool of (unknown) T cell receptor molecules than a substitution with a high log odds score (due to negative selection of T cell receptor matching non-modified peptides). However there are other ways of ascertaining this score. For example, considering the position of the mutation in the peptide (some positions may have a lower impact on binding than others), taking into account the nearest neighbors of the substituted amino acid (which could impact the secondary structure of the substituted amino acid), taking into account the entire peptide sequence, taking into account the complete structural information of the peptide in the MHC molecule, an so on. Ascertaining the score could also involve determination of a T cell receptor repertoire (such as the T cell receptor repertoire of a patient or a subset thereof) e.g. via NGS and performing docking simulations of T cell receptor-peptide-MHC complexes.

The present invention also may comprise performing the method of the invention on different peptides comprising the same modification(s) and/or different modifications.

The term "different peptides comprising the same modification(s)" in one embodiment relates to peptides comprising or consisting of different fragments of a modified protein, said different fragments comprising the same modification(s) present in the protein but differing in length and/or position of the modification(s). If a protein has a modification at position x, two or more fragments of said protein each comprising a different sequence window of said protein covering said position x are considered different peptides comprising the same modification(s).

The term "different peptides comprising different modifications" in one embodiment relates to peptides either of the same and/or differing lengths comprising different modifications of either of the same and/or different proteins. If a protein has modifications at positions x and y, two fragments of said protein each comprising a sequence window of said protein covering either position x or position y are considered different peptides comprising different modifications.

The present invention also may comprise breaking of protein sequences having modifications the immunogenicity of which is to be determined according to the invention into appropriate peptide lengths for MHC binding and ascertaining scores for binding to one or more MHC molecules of different modified peptides comprising the same and/or different modifications of either the same and/or different proteins. Outputs may be ranked and may consist of a list of peptides and their predicted scores, indicating their likelihood of binding.

The step of ascertaining a score for binding of the non-modified peptide to one or more MHC molecules and/or the step of ascertaining a score for binding of the modified peptide when present in a MHC-peptide complex to one or more T cell receptors may subsequently performed with all different modified peptides comprising the same and/or different modifications, a subset thereof, e.g. those modified peptides comprising the same and/or different modifications scoring best for binding to one or more MHC molecules, or only with the one modified peptide scoring best for binding to one or more MHC molecules.

Following said further steps, the results may be ranked and may consist of a list of peptides and their predicted scores, indicating their likelihood of being immunogenic.

Preferably, in such ranking, a score for binding of the modified peptide to one or more MHC molecules is weighted higher than a score for binding of the modified peptide when present in a MHC-peptide complex to one or more T cell receptors, preferably a score for the chemical and physical similarities between the non-modified and modified amino acids and a score for binding of the modified peptide when present in a MHC-peptide complex to one or more T cell receptors, preferably a score for the chemical and physical similarities between the non-modified and modified amino acids is weighted higher than a score for binding of the non-modified peptide to one or more MHC molecules.

The amino acid modifications the immunogenicity of which is to be determined according to the present invention may result from mutations in the nucleic acid of a cell. Such mutations may be identified by known sequencing techniques.

In one embodiment, the mutations are cancer specific somatic mutations in a tumor specimen of a cancer patient which may be determined by identifying sequence differences between the genome, exome and/or transcriptome of a tumor specimen and the genome, exome and/or transcriptome of a non-tumorigenous specimen.

According to the invention a tumor specimen relates to any sample such as a bodily sample derived from a patient containing or being expected of containing tumor or cancer cells. The bodily sample may be any tissue sample such as blood, a tissue sample obtained from the primary tumor or from tumor metastases or any other sample containing tumor or cancer cells. Preferably, a bodily sample is blood and cancer specific somatic mutations or sequence differences are determined in one or more circulating tumor cells (CTCs) contained in the blood. In another embodiment, a tumor specimen relates to one or more isolated tumor or cancer cells such as circulating tumor cells (CTCs) or a sample containing one or more isolated tumor or cancer cells such as circulating tumor cells (CTCs).

A non-tumorigenous specimen relates to any sample such as a bodily sample derived from a patient or another individual which preferably is of the same species as the patient, preferably a

healthy individual not containing or not being expected of containing tumor or cancer cells. The bodily sample may be any tissue sample such as blood or a sample from a non-tumorigenous tissue.

The invention may involve the determination of the cancer mutation signature of a patient. The term "cancer mutation signature" may refer to all cancer mutations present in one or more cancer cells of a patient or it may refer to only a portion of the cancer mutations present in one or more cancer cells of a patient. Accordingly, the present invention may involve the identification of all cancer specific mutations present in one or more cancer cells of a patient or it may involve the identification of only a portion of the cancer specific mutations present in one or more cancer cells of a patient. Generally, the methods of the invention provides for the identification of a number of mutations which provides a sufficient number of modifications or modified peptides to be included in the methods of the invention.

Preferably, the mutations identified according to the present invention are non-synonymous mutations, preferably non-synonymous mutations of proteins expressed in a tumor or cancer cell.

In one embodiment, cancer specific somatic mutations or sequence differences are determined in the genome, preferably the entire genome, of a tumor specimen. Thus, the invention may comprise identifying the cancer mutation signature of the genome, preferably the entire genome of one or more cancer cells. In one embodiment, the step of identifying cancer specific somatic mutations in a tumor specimen of a cancer patient comprises identifying the genome-wide cancer mutation profile.

In one embodiment, cancer specific somatic mutations or sequence differences are determined in the exome, preferably the entire exome, of a tumor specimen. Thus, the invention may comprise identifying the cancer mutation signature of the exome, preferably the entire exome of one or more cancer cells. In one embodiment, the step of identifying cancer specific somatic mutations in a tumor specimen of a cancer patient comprises identifying the exome-wide cancer mutation profile.

In one embodiment, cancer specific somatic mutations or sequence differences are determined in the transcriptome, preferably the entire transcriptome, of a tumor specimen. Thus, the invention may comprise identifying the cancer mutation signature of the transcriptome, preferably the entire transcriptome of one or more cancer cells. In one embodiment, the step of identifying cancer specific somatic mutations in a tumor specimen of a cancer patient comprises identifying the transcriptome- wide cancer mutation profile.

In one embodiment, the step of identifying cancer specific somatic mutations or identifying sequence differences comprises single cell sequencing of one or more, preferably 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or even more cancer cells. Thus, the invention may comprise identifying a cancer mutation signature of said one or more cancer cells. In one embodiment, the cancer cells are circulating tumor cells. The cancer cells such as the circulating tumor cells may be isolated prior to single cell sequencing.

In one embodiment, the step of identifying cancer specific somatic mutations or identifying sequence differences involves using next generation sequencing (NGS).

In one embodiment, the step of identifying cancer specific somatic mutations or identifying sequence differences comprises sequencing genomic DNA and/or RNA of the tumor specimen.

To reveal cancer specific somatic mutations or sequence differences the sequence information obtained from the tumor specimen is preferably compared with a reference such as sequence information obtained from sequencing nucleic acid such as DNA or RNA of normal non-cancerous cells such as germline cells which may either be obtained from the patient or a different individual. In one embodiment, normal genomic germline DNA is obtained from peripheral blood mononuclear cells (PBMCs)

The term "genome" relates to the total amount of genetic information in the chromosomes of an organism or a cell.

The term "exome" refers to part of the genome of an organism formed by exons, which are coding portions of expressed genes. The exome provides the genetic blueprint used in the synthesis of proteins and other functional gene products. It is the most functionally relevant part of the genome and, therefore, it is most likely to contribute to the phenotype of an organism. The exome of the human genome is estimated to comprise 1.5% of the total genome ***(Ng, PC et al, PLoS Gen., 4(8): 1-15, 2008).***

The term "transcriptome" relates to the set of all RNA molecules, including mRNA, rRNA, tRNA, and other non-coding RNA produced in one cell or a population of cells. In context of the present invention the transcriptome means the set of all RNA molecules produced in one cell, a population of cells, preferably a population of cancer cells, or all cells of a given individual at a certain time point.

A "nucleic acid" is according to the invention preferably deoxyribonucleic acid (DNA) or ribonucleic acid (RNA), more preferably RNA, most preferably ***in vitro*** transcribed RNA (IVT RNA) or synthetic RNA. Nucleic acids include according to the invention genomic DNA, cDNA, mRNA, recombinantly produced and chemically synthesized molecules. According to the invention, a nucleic acid may be present as a single-stranded or double-stranded and linear or covalently circularly closed molecule. A nucleic acid can, according to the invention, be isolated. The term "isolated nucleic acid" means, according to the invention, that the nucleic acid (i) was amplified ***in vitro,*** for example via polymerase chain reaction (PCR), (ii) was produced recombinantly by cloning, (iii) was purified, for example, by cleavage and separation by gel electrophoresis, or (iv) was synthesized, for example, by chemical synthesis. A nucleic can be employed for introduction into, i.e. transfection of, cells, in particular, in the form of RNA which can be prepared by ***in vitro*** transcription from a DNA template. The RNA can moreover be modified before application by stabilizing sequences, capping, and polyadenylation.

The term "genetic material" refers to isolated nucleic acid, either DNA or RNA, a section of a double helix, a section of a chromosome, or an organism's or cell's entire genome, in particular its exome or transcriptome.

The term "mutation" refers to a change of or difference in the nucleic acid sequence (nucleotide substitution, addition or deletion) compared to a reference. A "somatic mutation" can occur in any of the cells of the body except the germ cells (sperm and egg) and therefore are not passed on to children. These alterations can (but do not always) cause cancer or other diseases. Preferably a mutation is a non-synonymous mutation. The term "non-synonymous mutation" refers to a mutation, preferably a nucleotide substitution, which does result in an amino acid change such as an amino acid substitution in the translation product.

According to the invention, the term "mutation" includes point mutations, Indels, fusions, chromothripsis and RNA edits.

According to the invention, the term "Indel" describes a special mutation class, defined as a mutation resulting in a colocalized insertion and deletion and a net gain or loss in nucleotides. In coding regions of the genome, unless the length of an indel is a multiple of **3,** they produce a frameshift mutation. Indels can be contrasted with a point mutation; where an Indel inserts and deletes nucleotides from a sequence, a point mutation is a form of substitution that replaces one of the nucleotides.

Fusions can generate hybrid genes formed from two previously separate genes. It can occur as the result of a translocation, interstitial deletion, or chromosomal inversion. Often, fusion genes are oncogenes. Oncogenic fusion genes may lead to a gene product with a new or different function from the two fusion partners. Alternatively, a proto-oncogene is fused to a strong promoter, and thereby the oncogenic function is set to function by an upregulation caused by the strong promoter of the upstream fusion partner. Oncogenic fusion transcripts may also be caused by trans-splicing or read-through events.

According to the invention, the term "chromothripsis" refers to a genetic phenomenon by which specific regions of the genome are shattered and then stitched together via a single devastating event.

According to the invention, the term "RNA edit" or "RNA editing" refers to molecular processes

in which the information content in an RNA molecule is altered through a chemical change in the base makeup. RNA editing includes nucleoside modifications such as cytidine (C) to uridine (U) and adenosine (A) to inosine (I) deaminations, as well as non-templated nucleotide additions and insertions. RNA editing in mRNAs effectively alters the amino acid sequence of the encoded protein so that it differs from that predicted by the genomic DNA sequence.

The term "cancer mutation signature" refers to a set of mutations which are present in cancer cells when compared to non-cancerous reference cells.

According to the invention, a "reference" may be used to correlate and compare the results obtained in the methods of the invention from a tumor specimen. Typically the "reference" may be obtained on the basis of one or more normal specimens, in particular specimens which are not affected by a cancer disease, either obtained from a patient or one or more different individuals, preferably healthy individuals, in particular individuals of the same species. A "reference" can be determined empirically by testing a sufficiently large number of normal specimens.

Any suitable sequencing method can be used according to the invention for determining mutations, Next Generation Sequencing (NGS) technologies being preferred. Third Generation Sequencing methods might substitute for the NGS technology in the future to speed up the sequencing step of the method. For clarification purposes: the terms "Next Generation Sequencing" or "NGS" in the context of the present invention mean all novel high throughput sequencing technologies which, in contrast to the "conventional" sequencing methodology known as Sanger chemistry, read nucleic acid templates randomly in parallel along the entire genome by breaking the entire genome into small pieces. Such NGS technologies (also known as massively parallel sequencing technologies) are able to deliver nucleic acid sequence information of a whole genome, exome, transcriptome (all transcribed sequences of a genome) or methylome (all methylated sequences of a genome) in very short time periods, e.g. within 1-2 weeks, preferably within 1-7 days or most preferably within less than 24 hours and allow, in principle, single cell sequencing approaches. Multiple NGS platforms which are commercially available or which are mentioned in the literature can be used in the context of the present invention e.g. those described in detail in *Zhang et al. 2011: The impact of next-generation*

sequencing on genomics. J. Genet Genomics 38 (3), 95-109; or in Voelkerding et al. 2009: Next generation sequencing: From basic research to diagnostics. Clinical chemistry 55, 641-658. Non-limiting examples of such NGS technologies/platforms are

1) The sequencing-by-synthesis technology known as pyrosequencing implemented e.g. in the GS-FLX 454 Genome Sequencer ™ of Roche-associated company 454 Life Sciences (Branford, Connecticut), first described in *Ronaghi et al. 1998: A sequencing method based on real-time pyrophosphate". Science 281 (5375), 363-365.* This technology uses an emulsion PCR in which single-stranded DNA binding beads are encapsulated by vigorous vortexing into aqueous micelles containing PCR reactants surrounded by oil for emulsion PCR amplification. During the pyrosequencing process, light emitted from phosphate molecules during nucleotide incorporation is recorded as the polymerase synthesizes the DNA strand.

2) The sequencing-by-synthesis approaches developed by Solexa (now part of Illumina Inc., San Diego, California) which is based on reversible dye-terminators and implemented e.g. in the Illumina/Solexa Genome Analyzer™ and in the Illumina HiSeq 2000 Genome Analyzer™. In this technology, all four nucleotides are added simultaneously into oligo-primed cluster fragments in flow-cell channels along with DNA polymerase. Bridge amplification extends cluster strands with all four fluorescently labeled nucleotides for sequencing.

3) Sequencing-by-ligation approaches, e.g. implemented in the SOLid™ platform of Applied Biosystems (now Life Technologies Corporation, Carlsbad, California). In this technology, a pool of all possible oligonucleotides of a fixed length are labeled according to the sequenced position. Oligonucleotides are annealed and ligated; the preferential ligation by DNA ligase for matching sequences results in a signal informative of the nucleotide at that position. Before sequencing, the DNA is amplified by emulsion PCR. The resulting bead, each containing only copies of the same DNA molecule, are deposited on a glass slide. As a second example, he Polonator™ G.007 platform of Dover Systems (Salem, New Hampshire) also employs a sequencing-by-ligation approach by using a randomly arrayed, bead-based, emulsion PCR to amplify DNA fragments for parallel sequencing.

4) Single-molecule sequencing technologies such as e.g. implemented in the PacBio RS system of Pacific Biosciences (Menlo Park, California) or in the HeliScope™ platform of Helicos Biosciences (Cambridge, Massachusetts). The distinct characteristic of this technology is its ability to sequence single DNA or RNA molecules without amplification, defined as Single-Molecule Real Time (SMRT) DNA sequencing. For example, HeliScope uses a highly sensitive fluorescence detection system to directly detect each nucleotide as it is synthesized. A similar approach based on fluorescence resonance energy transfer (FRET) has been developed from Visigen Biotechnology (Houston, Texas). Other fluorescence-based single-molecule techniques are from U.S. Genomics (GeneEngine™) and Genovoxx (AnyGene™).

5) Nano-technologies for single-molecule sequencing in which various nanostructures are used which are e.g. arranged on a chip to monitor the movement of a polymerase molecule on a single strand during replication. Non-limiting examples for approaches based on nano-technologies are the GridON™ platform of Oxford Nanopore Technologies (Oxford, UK), the hybridization-assisted nano-pore sequencing (HANS™) platforms developed by Nabsys (Providence, Rhode Island), and the proprietary ligase-based DNA sequencing platform with DNA nanoball (DNB) technology called combinatorial probe-anchor ligation (cPAL™).

6) Electron microscopy based technologies for single-molecule sequencing, e.g. those developed by LightSpeed Genomics (Sunnyvale, California) and Halcyon Molecular (Redwood City, California)

7) Ion semiconductor sequencing which is based on the detection of hydrogen ions that are released during the polymerisation of DNA. For example, Ion Torrent Systems (San Francisco, California) uses a high-density array of micro-machined wells to perform this biochemical process in a massively parallel way. Each well holds a different DNA template. Beneath the wells is an ion-sensitive layer and beneath that a proprietary Ion sensor.

Preferably, DNA and RNA preparations serve as starting material for NGS. Such nucleic acids can be easily obtained from samples such as biological material, e.g. from fresh, flash-frozen or formalin-fixed paraffin embedded tumor tissues (FFPE) or from freshly isolated cells or from

CTCs which are present in the peripheral blood of patients. Normal non-mutated genomic DNA or RNA can be extracted from normal, somatic tissue, however germline cells are preferred in the context of the present invention. Germline DNA or RNA may be extracted from peripheral blood mononuclear cells (PBMCs) in patients with non-hematological malignancies. Although nucleic acids extracted from FFPE tissues or freshly isolated single cells are highly fragmented, they are suitable for NGS applications.

Several targeted NGS methods for exome sequencing are described in the literature (for review see e.g. ***Teer and Mullikin 2010: Human Mol Genet 19 (2), R145-51),*** all of which can be used in conjunction with the present invention. Many of these methods (described e.g. as genome capture, genome partitioning, genome enrichment etc.) use hybridization techniques and include array-based (e.g. ***Hodges et al. 2007: Nat. Genet. 39, 1522-1527)*** and liquid-based (e.g. ***Choi et al. 2009: Proc. Natl. Acad. Sci USA 106, 19096-19101)*** hybridization approaches. Commercial kits for DNA sample preparation and subsequent exome capture are also available: for example, Illumina Inc. (San Diego, California) offers the TruSeq™ DNA Sample Preparation Kit and the Exome Enrichment Kit TruSeq™ Exome Enrichment Kit.

In order to reduce the number of false positive findings in detecting cancer specific somatic mutations or sequence differences when comparing e.g. the sequence of a tumor sample to the sequence of a reference sample such as the sequence of a germ line sample it is preferred to determine the sequence in replicates of one or both of these sample types. Thus, it is preferred that the sequence of a reference sample such as the sequence of a germ line sample is determined twice, three times or more. Alternatively or additionally, the sequence of a tumor sample is determined twice, three times or more. It may also be possible to determine the sequence of a reference sample such as the sequence of a germ line sample and/or the sequence of a tumor sample more than once by determining at least once the sequence in genomic DNA and determining at least once the sequence in RNA of said reference sample and/or of said tumor sample. For example, by determining the variations between replicates of a reference sample such as a germ line sample the expected rate of false positive (FDR) somatic mutations as a statistical quantity can be estimated. Technical repeats of a sample should generate identical results and any detected mutation in this "same vs. same comparison" is a false positive. In

particular, to determine the false discovery rate for somatic mutation detection in a tumor sample relative to a reference sample, a technical repeat of the reference sample can be used as a reference to estimate the number of false positives. Furthermore, various quality related metrics (e.g. coverage or SNP quality) may be combined into a single quality score using a machine learning approach. For a given somatic variation all other variations with an exceeding quality score may be counted, which enables a ranking of all variations in a dataset.

In the context of the present invention, the term "RNA" relates to a molecule which comprises at least one ribonucleotide residue and preferably being entirely or substantially composed of ribonucleotide residues. "Ribonucleotide" relates to a nucleotide with a hydroxyl group at the 2'-position of a β-D-ribofuranosyl group. The term "RNA" comprises double-stranded RNA, single-stranded RNA, isolated RNA such as partially or completely purified RNA, essentially pure RNA, synthetic RNA, and recombinantly generated RNA such as modified RNA which differs from naturally occurring RNA by addition, deletion, substitution and/or alteration of one or more nucleotides. Such alterations can include addition of non-nucleotide material, such as to the end(s) of a RNA or internally, for example at one or more nucleotides of the RNA. Nucleotides in RNA molecules can also comprise non-standard nucleotides, such as non-naturally occurring nucleotides or chemically synthesized nucleotides or deoxynucleotides. These altered RNAs can be referred to as analogs or analogs of naturally-occurring RNA.

According to the present invention, the term "RNA" includes and preferably relates to "mRNA". The term "mRNA" means "messenger-RNA" and relates to a "transcript" which is generated by using a DNA template and encodes a peptide or polypeptide. Typically, an mRNA comprises a 5'-UTR, a protein coding region, and a 3'-UTR. mRNA only possesses limited half-life in cells and ***in vitro.*** In the context of the present invention, mRNA may be generated by ***in vitro*** transcription from a DNA template. The ***in vitro*** transcription methodology is known to the skilled person. For example, there is a variety of ***in vitro*** transcription kits commercially available.

According to the invention, the stability and translation efficiency of RNA may be modified as required. For example, RNA may be stabilized and its translation increased by one or more

modifications having a stabilizing effects and/or increasing translation efficiency of RNA. Such modifications are described, for example, in PCT/EP2006/009448 incorporated herein by reference. In order to increase expression of the RNA used according to the present invention, it may be modified within the coding region, i.e. the sequence encoding the expressed peptide or protein, preferably without altering the sequence of the expressed peptide or protein, so as to increase the GC-content to increase mRNA stability and to perform a codon optimization and, thus, enhance translation in cells.

The term "modification" in the context of the RNA used in the present invention includes any modification of an RNA which is not naturally present in said RNA.

In one embodiment of the invention, the RNA used according to the invention does not have uncapped 5'-triphosphates. Removal of such uncapped 5'-triphosphates can be achieved by treating RNA with a phosphatase.

The RNA according to the invention may have modified ribonucleotides in order to increase its stability and/or decrease cytotoxicity. For example, in one embodiment, in the RNA used according to the invention 5-methylcytidine is substituted partially or completely, preferably completely, for cytidine. Alternatively or additionally, in one embodiment, in the RNA used according to the invention pseudouridine is substituted partially or completely, preferably completely, for uridine.

In one embodiment, the term "modification" relates to providing an RNA with a 5'-cap or 5'-cap analog. The term "5'-cap" refers to a cap structure found on the 5'-end of an mRNA molecule and generally consists of a guanosine nucleotide connected to the mRNA via an unusual 5' to 5' triphosphate linkage. In one embodiment, this guanosine is methylated at the 7-position. The term "conventional 5'-cap" refers to a naturally occurring RNA 5'-cap, preferably to the 7-methylguanosine cap (m^7G). In the context of the present invention, the term "5'-cap" includes a 5'-cap analog that resembles the RNA cap structure and is modified to possess the ability to stabilize RNA and/or enhance translation of RNA if attached thereto, preferably *in vivo* and/or in a cell.

Providing an RNA with a 5'-cap or 5'-cap analog may be achieved by ***in vitro*** transcription of a DNA template in presence of said 5'-cap or 5'-cap analog, wherein said 5'-cap is co-transcriptionally incorporated into the generated RNA strand, or the RNA may be generated, for example, by ***in vitro*** transcription, and the 5'-cap may be attached to the RNA post-transcriptionally using capping enzymes, for example, capping enzymes of vaccinia virus.

The RNA may comprise further modifications. For example, a further modification of the RNA used in the present invention may be an extension or truncation of the naturally occurring poly(A) tail or an alteration of the 5'- or 3'-untranslated regions (UTR) such as introduction of a UTR which is not related to the coding region of said RNA, for example, the exchange of the existing 3'-UTR with or the insertion of one or more, preferably two copies of a 3'-UTR derived from a globin gene, such as alpha2-globin, alpha 1-globin, beta-globin, preferably beta-globin, more preferably human beta-globin.

RNA having an unmasked poly-A sequence is translated more efficiently than RNA having a masked poly-A sequence. The term "poly(A) tail" or "poly-A sequence" relates to a sequence of adenyl (A) residues which typically is located on the 3'-end of a RNA molecule and "unmasked poly-A sequence" means that the poly-A sequence at the 3' end of an RNA molecule ends with an A of the poly-A sequence and is not followed by nucleotides other than A located at the 3' end, i.e. downstream, of the poly-A sequence. Furthermore, a long poly-A sequence of about 120 base pairs results in an optimal transcript stability and translation efficiency of RNA.

Therefore, in order to increase stability and/or expression of the RNA used according to the present invention, it may be modified so as to be present in conjunction with a poly-A sequence, preferably having a length of 10 to 500, more preferably 30 to 300, even more preferably 65 to 200 and especially 100 to 150 adenosine residues. In an especially preferred embodiment the poly-A sequence has a length of approximately 120 adenosine residues. To further increase stability and/or expression of the RNA used according to the invention, the poly-A sequence can be unmasked.

In addition, incorporation of a 3'-non translated region (UTR) into the 3'-non translated region of an RNA molecule can result in an enhancement in translation efficiency. A synergistic effect may be achieved by incorporating two or more of such 3'-non translated regions. The 3'-non translated regions may be autologous or heterologous to the RNA into which they are introduced. In one particular embodiment the 3'-non translated region is derived from the human β-globin gene.

A combination of the above described modifications, i.e. incorporation of a poly-A sequence, unmasking of a poly-A sequence and incorporation of one or more 3'-non translated regions, has a synergistic influence on the stability of RNA and increase in translation efficiency.

The term "stability" of RNA relates to the "half-life" of RNA. "Half-life" relates to the period of time which is needed to eliminate half of the activity, amount, or number of molecules. In the context of the present invention, the half-life of an RNA is indicative for the stability of said RNA. The half-life of RNA may influence the "duration of expression" of the RNA. It can be expected that RNA having a long half-life will be expressed for an extended time period.

Of course, if according to the present invention it is desired to decrease stability and/or translation efficiency of RNA, it is possible to modify RNA so as to interfere with the function of elements as described above increasing the stability and/or translation efficiency of RNA.

The term "expression" is used according to the invention in its most general meaning and comprises the production of RNA and/or peptides, polypeptides or proteins, e.g. by transcription and/or translation. With respect to RNA, the term "expression" or "translation" relates in particular to the production of peptides, polypeptides or proteins. It also comprises partial expression of nucleic acids. Moreover, expression can be transient or stable.

According to the invention, the term expression also includes an "aberrant expression" or "abnormal expression". "Aberrant expression" or "abnormal expression" means according to the invention that expression is altered, preferably increased, compared to a reference, e.g. a state in a subject not having a disease associated with aberrant or abnormal expression of a certain

protein, e.g., a tumor antigen. An increase in expression refers to an increase by at least 10%, in particular at least 20%, at least 50% or at least 100%, or more. In one embodiment, expression is only found in a diseased tissue, while expression in a healthy tissue is repressed.

The term "specifically expressed" means that a protein is essentially only expressed in a specific tissue or organ. For example, a tumor antigen specifically expressed in gastric mucosa means that said protein is primarily expressed in gastric mucosa and is not expressed in other tissues or is not expressed to a significant extent in other tissue or organ types. Thus, a protein that is exclusively expressed in cells of the gastric mucosa and to a significantly lesser extent in any other tissue, such as testis, is specifically expressed in cells of the gastric mucosa. In some embodiments, a tumor antigen may also be specifically expressed under normal conditions in more than one tissue type or organ, such as in 2 or 3 tissue types or organs, but preferably in not more than 3 different tissue or organ types. In this case, the tumor antigen is then specifically expressed in these organs. For example, if a tumor antigen is expressed under normal conditions preferably to an approximately equal extent in lung and stomach, said tumor antigen is specifically expressed in lung and stomach.

In the context of the present invention, the term "transcription" relates to a process, wherein the genetic code in a DNA sequence is transcribed into RNA. Subsequently, the RNA may be translated into protein. According to the present invention, the term "transcription" comprises "***in vitro*** transcription", wherein the term "***in vitro*** transcription" relates to a process wherein RNA, in particular mRNA, is ***in vitro*** synthesized in a cell-free system, preferably using appropriate cell extracts. Preferably, cloning vectors are applied for the generation of transcripts. These cloning vectors are generally designated as transcription vectors and are according to the present invention encompassed by the term "vector". According to the present invention, the RNA used in the present invention preferably is ***in vitro*** transcribed RNA (IVT-RNA) and may be obtained by ***in vitro*** transcription of an appropriate DNA template. The promoter for controlling transcription can be any promoter for any RNA polymerase. Particular examples of RNA polymerases are the T7, T3, and SP6 RNA polymerases. Preferably, the ***in vitro*** transcription according to the invention is controlled by a T7 or SP6 promoter. A DNA template for ***in vitro*** transcription may be obtained by cloning of a nucleic acid, in particular cDNA, and introducing

it into an appropriate vector for *in vitro* transcription. The cDNA may be obtained by reverse transcription of RNA.

The term "translation" according to the invention relates to the process in the ribosomes of a cell by which a strand of messenger RNA directs the assembly of a sequence of amino acids to make a peptide, polypeptide or protein.

Expression control sequences or regulatory sequences, which according to the invention may be linked functionally with a nucleic acid, can be homologous or heterologous with respect to the nucleic acid. A coding sequence and a regulatory sequence are linked together "functionally" if they are bound together covalently, so that the transcription or translation of the coding sequence is under the control or under the influence of the regulatory sequence. If the coding sequence is to be translated into a functional protein, with functional linkage of a regulatory sequence with the coding sequence, induction of the regulatory sequence leads to a transcription of the coding sequence, without causing a reading frame shift in the coding sequence or inability of the coding sequence to be translated into the desired protein or peptide.

The term "expression control sequence" or "regulatory sequence" comprises, according to the invention, promoters, ribosome-binding sequences and other control elements, which control the transcription of a nucleic acid or the translation of the derived RNA. In certain embodiments of the invention, the regulatory sequences can be controlled. The precise structure of regulatory sequences can vary depending on the species or depending on the cell type, but generally comprises 5'-untranscribed and 5'- and 3'-untranslated sequences, which are involved in the initiation of transcription or translation, such as TATA-box, capping-sequence, CAAT-sequence and the like. In particular, 5'-untranscribed regulatory sequences comprise a promoter region that includes a promoter sequence for transcriptional control of the functionally bound gene. Regulatory sequences can also comprise enhancer sequences or upstream activator sequences.

Preferably, according to the invention, RNA to be expressed in a cell is introduced into said cell. In one embodiment of the methods according to the invention, the RNA that is to be introduced into a cell is obtained by *in vitro* transcription of an appropriate DNA template.

According to the invention, terms such as "RNA capable of expressing" and "RNA encoding" are used interchangeably herein and with respect to a particular peptide or polypeptide mean that the RNA, if present in the appropriate environment, preferably within a cell, can be expressed to produce said peptide or polypeptide. Preferably, RNA according to the invention is able to interact with the cellular translation machinery to provide the peptide or polypeptide it is capable of expressing.

Terms such as "transferring", "introducing" or "transfecting" are used interchangeably herein and relate to the introduction of nucleic acids, in particular exogenous or heterologous nucleic acids, in particular RNA into a cell. According to the present invention, the cell can form part of an organ, a tissue and/or an organism. According to the present invention, the administration of a nucleic acid is either achieved as naked nucleic acid or in combination with an administration reagent. Preferably, administration of nucleic acids is in the form of naked nucleic acids. Preferably, the RNA is administered in combination with stabilizing substances such as RNase inhibitors. The present invention also envisions the repeated introduction of nucleic acids into cells to allow sustained expression for extended time periods.

Cells can be transfected with any carriers with which RNA can be associated, e.g. by forming complexes with the RNA or forming vesicles in which the RNA is enclosed or encapsulated, resulting in increased stability of the RNA compared to naked RNA. Carriers useful according to the invention include, for example, lipid-containing carriers such as cationic lipids, liposomes, in particular cationic liposomes, and micelles, and nanoparticles. Cationic lipids may form complexes with negatively charged nucleic acids. Any cationic lipid may be used according to the invention.

Preferably, the introduction of RNA which encodes a peptide or polypeptide into a cell, in particular into a cell present ***in vivo,*** results in expression of said peptide or polypeptide in the cell. In particular embodiments, the targeting of the nucleic acids to particular cells is preferred. In such embodiments, a carrier which is applied for the administration of the nucleic acid to a cell (for example, a retrovirus or a liposome), exhibits a targeting molecule. For example, a

molecule such as an antibody which is specific for a surface membrane protein on the target cell or a ligand for a receptor on the target cell may be incorporated into the nucleic acid carrier or may be bound thereto. In case the nucleic acid is administered by liposomes, proteins which bind to a surface membrane protein which is associated with endocytosis may be incorporated into the liposome formulation in order to enable targeting and/or uptake. Such proteins encompass capsid proteins of fragments thereof which are specific for a particular cell type, antibodies against proteins which are internalized, proteins which target an intracellular location etc.

The term "cell" or "host cell" preferably is an intact cell, i.e. a cell with an intact membrane that has not released its normal intracellular components such as enzymes, organelles, or genetic material. An intact cell preferably is a viable cell, i.e. a living cell capable of carrying out its normal metabolic functions. Preferably said term relates according to the invention to any cell which can be transformed or transfected with an exogenous nucleic acid. The term "cell" includes according to the invention prokaryotic cells (e.g., E. coli) or eukaryotic cells (e.g., dendritic cells, **B** cells, CHO cells, COS cells, K562 cells, HEK293 cells, HELA cells, yeast cells, and insect cells). The exogenous nucleic acid may be found inside the cell (i) freely dispersed as such, (ii) incorporated in a recombinant vector, or (iii) integrated into the host cell genome or mitochondrial DNA. Mammalian cells are particularly preferred, such as cells from humans, mice, hamsters, pigs, goats, and primates. The cells may be derived from a large number of tissue types and include primary cells and cell lines. Specific examples include keratinocytes, peripheral blood leukocytes, bone marrow stem cells, and embryonic stem cells. In further embodiments, the cell is an antigen-presenting cell, in particular a dendritic cell, a monocyte, or macrophage.

A cell which comprises a nucleic acid molecule preferably expresses the peptide or polypeptide encoded by the nucleic acid.

The term "clonal expansion" refers to a process wherein a specific entity is multiplied. In the context of the present invention, the term is preferably used in the context of an immunological response in which lymphocytes are stimulated by an antigen, proliferate, and the specific

lymphocyte recognizing said antigen is amplified. Preferably, clonal expansion leads to differentiation of the lymphocytes.

Terms such as "reducing" or "inhibiting" relate to the ability to cause an overall decrease, preferably of 5% or greater, 10% or greater, 20% or greater, more preferably of 50% or greater, and most preferably of 75% or greater, in the level. The term "inhibit" or similar phrases includes a complete or essentially complete inhibition, i.e. a reduction to zero or essentially to zero.

Terms such as "increasing", "enhancing", "promoting" or "prolonging" preferably relate to an increase, enhancement, promotion or prolongation by about at least 10%, preferably at least 20%, preferably at least 30%, preferably at least 40%, preferably at least 50%, preferably at least 80%, preferably at least 100%, preferably at least 200% and in particular at least 300%. These terms may also relate to an increase, enhancement, promotion or prolongation from zero or a non-measurable or non-detectable level to a level of more than zero or a level which is measurable or detectable.

The present invention provides vaccines such as cancer vaccines designed on the basis of amino acid modifications or modified peptides predicted as being immunogenic by the methods of the present invention.

According to the invention, the term "vaccine" relates to a pharmaceutical preparation (pharmaceutical composition) or product that upon administration induces an immune response, in particular a cellular immune response, which recognizes and attacks a pathogen or a diseased cell such as a cancer cell. A vaccine may be used for the prevention or treatment of a disease. The term "personalized cancer vaccine" or "individualized cancer vaccine" concerns a particular cancer patient and means that a cancer vaccine is adapted to the needs or special circumstances of an individual cancer patient.

In one embodiment, a vaccine provided according to the invention may comprise a peptide or polypeptide comprising one or more amino acid modifications or one or more modified peptides

predicted as being immunogenic by the methods of the invention or a nucleic acid, preferably RNA, encoding said peptide or polypeptide.

The cancer vaccines provided according to the invention when administered to a patent provide one or more T cell epitopes suitable for stimulating, priming and/or expanding T cells specific for the patient's tumor. The T cells are preferably directed against cells expressing antigens from which the T cell epitopes are derived. Thus, the vaccines described herein are preferably capable of inducing or promoting a cellular response, preferably cytotoxic T cell activity, against a cancer disease characterized by presentation of one or more tumor-associated neoantigens with class I MHC. Since a vaccine provided according to the present invention will target cancer specific mutations it will be specific for the patient's tumor.

A vaccine provided according to the invention relates to a vaccine which when administered to a patent preferably provides one or more T cell epitopes, such as 2 or more, 5 or more, 10 or more, 15 or more, 20 or more, 25 or more, 30 or more and preferably up to 60, up to 55, up to 50, up to 45, up to 40, up to 35 or up to 30 T cell epitopes, incorporating amino acid modifications or modified peptides predicted as being immunogenic by the methods of the invention. Such T cell epitopes are also termed "neo-epitopes" herein. Presentation of these epitopes by cells of a patient, in particular antigen presenting cells, preferably results in T cells targeting the epitopes when bound to MHC and thus, the patient's tumor, preferably the primary tumor as well as tumor metastases, expressing antigens from which the T cell epitopes are derived and presenting the same epitopes on the surface of the tumor cells.

The methods of the invention may comprise the further step of determining the usability of the identified amino acid modifications or modified peptides for cancer vaccination. Thus further steps can involve one or more of the following: (i) assessing whether the modifications are located in known or predicted MHC presented epitopes, (ii) *in vitro* and/or *in silico* testing whether the modifications are located in MHC presented epitopes, e.g. testing whether the modifications are part of peptide sequences which are processed into and/or presented as MHC presented epitopes, and (iii) *in vitro* testing whether the envisaged modified epitopes, in particular when present in their natural sequence context, e.g. when flanked by amino acid

sequences also flanking said epitopes in the naturally occurring protein, and when expressed in antigen presenting cells are able to stimulate T cells such as T cells of the patient having the desired specificity. Such flanking sequences each may comprise 3 or more, 5 or more, 10 or more, 15 or more, 20 or more and preferably up to 50, up to 45, up to 40, up to 35 or up to 30 amino acids and may flank the epitope sequence N-terminally and/or C-terminally.

Modified peptides determined according to the invention may be ranked for their usability as epitopes for cancer vaccination. Thus, in one aspect, the method of the invention comprises a manual or computer-based analytical process in which the identified modified peptides are analyzed and selected for their usability in the respective vaccine to be provided. In a preferred embodiment, said analytical process is a computational algorithm-based process. Preferably, said analytical process comprises determining and/or ranking epitopes according to a prediction of their capacity of being immunogenic.

The neo-epitopes identified according to the invention and provided by a vaccine of the invention are preferably present in the form of a polypeptide comprising said neo-epitopes such as a polyepitopic polypeptide or a nucleic acid, in particular RNA, encoding said polypeptide. Furthermore, the neo-epitopes may be present in the polypeptide in the form of a vaccine sequence, i.e. present in their natural sequence context, e.g. flanked by amino acid sequences also flanking said epitopes in the naturally occurring protein. Such flanking sequences each may comprise 5 or more, 10 or more, 15 or more, 20 or more and preferably up to 50, up to 45, up to 40, up to 35 or up to 30 amino acids and may flank the epitope sequence N-terminally and/or C-terminally. Thus, a vaccine sequence may comprise 20 or more, 25 or more, 30 or more, 35 or more, 40 or more and preferably up to 50, up to 45, up to 40, up to 35 or up to 30 amino acids. In one embodiment, the neo-epitopes and/or vaccine sequences are lined up in the polypeptide head-to-tail.

In one embodiment, the neo-epitopes and/or vaccine sequences are spaced by linkers, in particular neutral linkers. The term "linker" according to the invention relates to a peptide added between two peptide domains such as epitopes or vaccine sequences to connect said peptide domains. There is no particular limitation regarding the linker sequence. However, it is preferred

that the linker sequence reduces steric hindrance between the two peptide domains, is well translated, and supports or allows processing of the epitopes. Furthermore, the linker should have no or only little immunogenic sequence elements. Linkers preferably should not create non-endogenous neo-epitopes like those generated from the junction suture between adjacent neo-epitopes, which might generate unwanted immune reactions. Therefore, the polyepitopic vaccine should preferably contain linker sequences which are able to reduce the number of unwanted MHC binding junction epitopes. ***Hoyt et al. (EMBO J. 25(8), 1720-9, 2006)*** and ***Zhang et al. (J. Biol. Chem., 279(10), 8635-41, 2004)*** have shown that glycine-rich sequences impair proteasomal processing and thus the use of glycine rich linker sequences act to minimize the number of linker-contained peptides that can be processed by the proteasome. Furthermore, glycine was observed to inhibit a strong binding in MHC binding groove positions ***(Abastado et al., J. Immunol. 151(7), 3569-75, 1993). Schlessinger et al. (Proteins, 61(1), 115-26, 2005)*** had found that amino acids glycine and serine included in an amino acid sequence result in a more flexible protein that is more efficiently translated and processed by the proteasome, enabling better access to the encoded neo-epitopes. The linker each may comprise 3 or more, 6 or more, 9 or more, 10 or more, 15 or more, 20 or more and preferably up to 50, up to 45, up to 40, up to 35 or up to 30 amino acids. Preferably the linker is enriched in glycine and/or serine amino acids. Preferably, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or at least 95% of the amino acids of the linker are glycine and/or serine. In one preferred embodiment, a linker is substantially composed of the amino acids glycine and serine. In one embodiment, the linker comprises the amino acid sequence $\mathbf{(GGS)_a(GSS)_b(GGG)_c(SSG)d(GSG)_e}$ wherein a, b, c, d and e is independently a number selected from 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 and wherein a + b + c + d + e are different from 0 and preferably are 2 or more, 3 or more, 4 or more or 5 or more. In one embodiment, the linker comprises a sequence as described herein including the linker sequences described in the examples such as the sequence **GGSGGGGSG.**

In one particularly preferred embodiment, a polypeptide incorporating one or more neo-epitopes such as a polyepitopic polypeptide according to the present invention is administered to a patient in the form of a nucleic acid, preferably RNA such as ***in vitro*** transcribed or synthetic RNA, which may be expressed in cells of a patient such as antigen presenting cells to produce the

polypeptide. The present invention also envisions the administration of one or more multiepitopic polypeptides which for the purpose of the present invention are comprised by the term "polyepitopic polypeptide", preferably in the form of a nucleic acid, preferably RNA such as *in vitro* transcribed or synthetic RNA, which may be expressed in cells of a patient such as antigen presenting cells to produce the one or more polypeptides. In the case of an administration of more than one multiepitopic polypeptide the neo-epitopes provided by the different multiepitopic polypeptides may be different or partially overlapping. Once present in cells of a patient such as antigen presenting cells the polypeptide according to the invention is processed to produce the neo-epitopes identified according to the invention. Administration of a vaccine provided according to the invention may provide MHC class Il-presented epitopes that are capable of eliciting a CD4+ helper T cell response against cells expressing antigens from which the MHC presented epitopes are derived. Alternatively or additionally, administration of a vaccine provided according to the invention may provide MHC class I-presented epitopes that are capable of eliciting a CD8+ T cell response against cells expressing antigens from which the MHC presented epitopes are derived. Furthermore, administration of a vaccine provided according to the invention may provide one or more neo-epitopes (including known neo-epitopes and neo-epitopes identified according to the invention) as well as one or more epitopes not containing cancer specific somatic mutations but being expressed by cancer cells and preferably inducing an immune response against cancer cells, preferably a cancer specific immune response. In one embodiment, administration of a vaccine provided according to the invention provides neo-epitopes that are MHC class Il-presented epitopes and/or are capable of eliciting a CD4+ helper T cell response against cells expressing antigens from which the MHC presented epitopes are derived as well as epitopes not containing cancer-specific somatic mutations that are MHC class I-presented epitopes and/or are capable of eliciting a CD8+ T cell response against cells expressing antigens from which the MHC presented epitopes are derived. In one embodiment, the epitopes not containing cancer-specific somatic mutations are derived from a tumor antigen. In one embodiment, the neo-epitopes and epitopes not containing cancer-specific somatic mutations have a synergistic effect in the treatment of cancer. Preferably, a vaccine provided according to the invention is useful for polyepitopic stimulation of cytotoxic and/or helper T cell responses.

The vaccine provided according to the invention may be a recombinant vaccine.

The term "recombinant" in the context of the present invention means "made through genetic engineering". Preferably, a "recombinant entity" such as a recombinant polypeptide in the context of the present invention is not occurring naturally, and preferably is a result of a combination of entities such as amino acid or nucleic acid sequences which are not combined in nature. For example, a recombinant polypeptide in the context of the present invention may contain several amino acid sequences such as neo-epitopes or vaccine sequences derived from different proteins or different portions of the same protein fused together, e.g., by peptide bonds or appropriate linkers.

The term "naturally occurring" as used herein refers to the fact that an object can be found in nature. For example, a peptide or nucleic acid that is present in an organism (including viruses) and can be isolated from a source in nature and which has not been intentionally modified by man in the laboratory is naturally occurring.

Agents, compositions and methods described herein can be used to treat a subject with a disease, e.g., a disease characterized by the presence of diseased cells expressing an antigen and presenting a fragment thereof. Particularly preferred diseases are cancer diseases. Agents, compositions and methods described herein may also be used for immunization or vaccination to prevent a disease described herein.

According to the invention, the term "disease" refers to any pathological state, including cancer diseases, in particular those forms of cancer diseases described herein.

The term "normal" refers to the healthy state or the conditions in a healthy subject or tissue, i.e., non-pathological conditions, wherein "healthy" preferably means non-cancerous.

"Disease involving cells expressing an antigen" means according to the invention that expression of the antigen in cells of a diseased tissue or organ is detected. Expression in cells of a diseased tissue or organ may be increased compared to the state in a healthy tissue or organ. An increase

refers to an increase by at least 10%, in particular at least 20%, at least 50%, at least 100%, at least *200%,* at least 500%, at least 1000%, at least 10000% or even more. In one embodiment, expression is only found in a diseased tissue, while expression in a healthy tissue is repressed. According to the invention, diseases involving or being associated with cells expressing an antigen include cancer diseases.

According to the invention, the term "tumor" or "tumor disease" refers to an abnormal growth of cells (called neoplastic cells, tumorigenous cells or tumor cells) preferably forming a swelling or lesion. By "tumor cell" is meant an abnormal cell that grows by a rapid, uncontrolled cellular proliferation and continues to grow after the stimuli that initiated the new growth cease. Tumors show partial or complete lack of structural organization and functional coordination with the normal tissue, and usually form a distinct mass of tissue, which may be either benign, pre-malignant or malignant.

Cancer (medical term: malignant neoplasm) is a class of diseases in which a group of cells display uncontrolled growth (division beyond the normal limits), invasion (intrusion on and destruction of adjacent tissues), and sometimes metastasis (spread to other locations in the body via lymph or blood). These three malignant properties of cancers differentiate them from benign tumors, which are self-limited, and do not invade or metastasize. Most cancers form a tumor but some, like leukemia, do not. Malignancy, malignant neoplasm, and malignant tumor are essentially synonymous with cancer.

Neoplasm is an abnormal mass of tissue as a result of neoplasia. Neoplasia (new growth in Greek) is the abnormal proliferation of cells. The growth of the cells exceeds, and is uncoordinated with that of the normal tissues around it. The growth persists in the same excessive manner even after cessation of the stimuli. It usually causes a lump or tumor. Neoplasms may be benign, pre-malignant or malignant.

"Growth of a tumor" or "tumor growth" according to the invention relates to the tendency of a tumor to increase its size and/or to the tendency of tumor cells to proliferate.

For purposes of the present invention, the terms "cancer" and "cancer disease" are used interchangeably with the terms "tumor" and "tumor disease".

Cancers are classified by the type of cell that resembles the tumor and, therefore, the tissue presumed to be the origin of the tumor. These are the histology and the location, respectively.

The term "cancer" according to the invention comprises carcinomas, adenocarcinomas, blastomas, leukemias, seminomas, melanomas, teratomas, lymphomas, neuroblastomas, gliomas, rectal cancer, endometrial cancer, kidney cancer, adrenal cancer, thyroid cancer, blood cancer, skin cancer, cancer of the brain, cervical cancer, intestinal cancer, liver cancer, colon cancer, stomach cancer, intestine cancer, head and neck cancer, gastrointestinal cancer, lymph node cancer, esophagus cancer, colorectal cancer, pancreas cancer, ear, nose and throat (ENT) cancer, breast cancer, prostate cancer, cancer of the uterus, ovarian cancer and lung cancer and the metastases thereof. Examples thereof are lung carcinomas, mamma carcinomas, prostate carcinomas, colon carcinomas, renal cell carcinomas, cervical carcinomas, or metastases of the cancer types or tumors described above. The term cancer according to the invention also comprises cancer metastases and relapse of cancer.

By "metastasis" is meant the spread of cancer cells from its original site to another part of the body. The formation of metastasis is a very complex process and depends on detachment of malignant cells from the primary tumor, invasion of the extracellular matrix, penetration of the endothelial basement membranes to enter the body cavity and vessels, and then, after being transported by the blood, infiltration of target organs. Finally, the growth of a new tumor, i.e. a secondary tumor or metastatic tumor, at the target site depends on angiogenesis. Tumor metastasis often occurs even after the removal of the primary tumor because tumor cells or components may remain and develop metastatic potential. In one embodiment, the term "metastasis" according to the invention relates to "distant metastasis" which relates to a metastasis which is remote from the primary tumor and the regional lymph node system.

The cells of a secondary or metastatic tumor are like those in the original tumor. This means, for example, that, if ovarian cancer metastasizes to the liver, the secondary tumor is made up of

abnormal ovarian cells, not of abnormal liver cells. The tumor in the liver is then called metastatic ovarian cancer, not liver cancer.

The term "circulating tumor cells" or "CTCs" relates to cells that have detached from a primary tumor or tumor metastases and circulate in the bloodstream. CTCs may constitute seeds for subsequent growth of additional tumors (metastasis) in different tissues. Circulating tumor cells are found in frequencies in the order of 1-10 CTC per mL of whole blood in patients with metastatic disease. Research methods have been developed to isolate CTC. Several research methods have been described in the art to isolate CTCs, e.g. techniques which use of the fact that epithelial cells commonly express the cell adhesion protein EpCAM, which is absent in normal blood cells. Immunomagnetic bead-based capture involves treating blood specimens with antibody to EpCAM that has been conjugated with magnetic particles, followed by separation of tagged cells in a magnetic field. Isolated cells are then stained with antibody to another epithelial marker, cytokeratin, as well as a common leukocyte marker CD45, so as to distinguish rare CTCs from contaminating white blood cells. This robust and semi-automated approach identifies CTCs with an average yield of approximately 1 CTC/mL and a purity of 0.1% ***(Allard et ai, 2004: Clin Cancer Res 10, 6897-6904).*** A second method for isolating CTCs uses a microfluidic-based CTC capture device which involves flowing whole blood through a chamber embedded with 80,000 microposts that have been rendered functional by coating with antibody to EpCAM. CTCs are then stained with secondary antibodies against either cytokeratin or tissue specific markers, such as PSA in prostate cancer or HER2 in breast cancer and are visualized by automated scanning of microposts in multiple planes along three dimensional coordinates. CTC-chips are able to identifying cytokerating-positive circulating tumor cells in patients with a median yield of 50 cells/ml and purity ranging from 1-80% ***{Nagrath et ai, 2007: Nature 450, 1235-1239).*** Another possibility for isolating CTCs is using the CellSearch™ Circulating Tumor Cell (CTC) Test from Veridex, LLC (Raritan, NJ) which captures, identifies, and counts CTCs in a tube of blood. The CellSearch™ system is a U.S. Food and Drug Administration (FDA) approved methodology for enumeration of CTC in whole blood which is based on a combination of immunomagnetic labeling and automated digital microscopy. There are other methods for isolating CTCs described in the literature all of which can be used in conjunction with the present invention.

A relapse or recurrence occurs when a person is affected again by a condition that affected them in the past. For example, if a patient has suffered from a tumor disease, has received a successful treatment of said disease and again develops said disease said newly developed disease may be considered as relapse or recurrence. However, according to the invention, a relapse or recurrence of a tumor disease may but does not necessarily occur at the site of the original tumor disease. Thus, for example, if a patient has suffered from ovarian tumor and has received a successful treatment a relapse or recurrence may be the occurrence of an ovarian tumor or the occurrence of a tumor at a site different to ovary. A relapse or recurrence of a tumor also includes situations wherein a tumor occurs at a site different to the site of the original tumor as well as at the site of the original tumor. Preferably, the original tumor for which the patient has received a treatment is a primary tumor and the tumor at a site different to the site of the original tumor is a secondary or metastatic tumor.

By "treat" is meant to administer a compound or composition as described herein to a subject in order to prevent or eliminate a disease, including reducing the size of a tumor or the number of tumors in a subject; arrest or slow a disease in a subject; inhibit or slow the development of a new disease in a subject; decrease the frequency or severity of symptoms and/or recurrences in a subject who currently has or who previously has had a disease; and/or prolong, i.e. increase the lifespan of the subject. In particular, the term "treatment of a disease" includes curing, shortening the duration, ameliorating, preventing, slowing down or inhibiting progression or worsening, or preventing or delaying the onset of a disease or the symptoms thereof.

By "being at risk" is meant a subject, i.e. a patient, that is identified as having a higher than normal chance of developing a disease, in particular cancer, compared to the general population. In addition, a subject who has had, or who currently has, a disease, in particular cancer, is a subject who has an increased risk for developing a disease, as such a subject may continue to develop a disease. Subjects who currently have, or who have had, a cancer also have an increased risk for cancer metastases.

The term "immunotherapy" relates to a treatment involving activation of a specific immune reaction. In the context of the present invention, terms such as "protect", "prevent", "prophylactic", "preventive", or "protective" relate to the prevention or treatment or both of the occurrence and/or the propagation of a disease in a subject and, in particular, to minimizing the chance that a subject will develop a disease or to delaying the development of a disease. For example, a person at risk for a tumor, as described above, would be a candidate for therapy to prevent a tumor.

A prophylactic administration of an immunotherapy, for example, a prophylactic administration of a vaccine of the invention, preferably protects the recipient from the development of a disease. A therapeutic administration of an immunotherapy, for example, a therapeutic administration of a vaccine of the invention, may lead to the inhibition of the progress/growth of the disease. This comprises the deceleration of the progress/growth of the disease, in particular a disruption of the progression of the disease, which preferably leads to elimination of the disease.

Immunotherapy may be performed using any of a variety of techniques, in which agents provided herein function to remove diseased cells from a patient. Such removal may take place as a result of enhancing or inducing an immune response in a patient specific for an antigen or a cell expressing an antigen.

Within certain embodiments, immunotherapy may be active immunotherapy, in which treatment relies on the *in vivo* stimulation of the endogenous host immune system to react against diseased cells with the administration of immune response-modifying agents (such as polypeptides and nucleic acids as provided herein).

The agents and compositions provided herein may be used alone or in combination with conventional therapeutic regimens such as surgery, irradiation, chemotherapy and/or bone marrow transplantation (autologous, syngeneic, allogeneic or unrelated).

The term "immunization" or "vaccination" describes the process of treating a subject with the purpose of inducing an immune response for therapeutic or prophylactic reasons.

The term "in vivo" relates to the situation in a subject.

The terms "subject", "individual", "organism" or "patient" are used interchangeably and relate to vertebrates, preferably mammals. For example, mammals in the context of the present invention are humans, non-human primates, domesticated animals such as dogs, cats, sheep, cattle, goats, pigs, horses etc., laboratory animals such as mice, rats, rabbits, guinea pigs, etc. as well as animals in captivity such as animals of zoos. The term "animal" as used herein also includes humans. The term "subject" may also include a patient, i.e., an animal, preferably a human having a disease, preferably a disease as described herein.

The term "autologous" is used to describe anything that is derived from the same subject. For example, "autologous transplant" refers to a transplant of tissue or organs derived from the same subject. Such procedures are advantageous because they overcome the immunological barrier which otherwise results in rejection.

The term "heterologous" is used to describe something consisting of multiple different elements. As an example, the transfer of one individual's bone marrow into a different individual constitutes a heterologous transplant. A heterologous gene is a gene derived from a source other than the subject.

As part of the composition for an immunization or a vaccination, preferably one or more agents as described herein are administered together with one or more adjuvants for inducing an immune response or for increasing an immune response. The term "adjuvant" relates to compounds which prolongs or enhances or accelerates an immune response. The composition of the present invention preferably exerts its effect without addition of adjuvants. Still, the composition of the present application may contain any known adjuvant. Adjuvants comprise a heterogeneous group of compounds such as oil emulsions (e.g., Freund's adjuvants), mineral compounds (such as alum), bacterial products (such as *Bordetella pertussis* toxin), liposomes, and immune-stimulating complexes. Examples for adjuvants are monophosphoryl-lipid-A (MPL SmithKline Beecham). Saponins such as QS21 (SmithKline Beecham), DQS21 (SmithKline

Beecham; WO 96/33739), QS7, QS17, QS18, and QS-L1 (So et al., 1997, Mol. Cells 7: 178-186), incomplete Freund's adjuvants, complete Freund's adjuvants, vitamin E, montanid, alum, CpG oligonucleotides (Krieg et al., 1995, Nature 374: 546-549), and various water-in-oil emulsions which are prepared from biologically degradable oils such as squalene and/or tocopherol.

Other substances which stimulate an immune response of the patient may also be administered. It is possible, for example, to use cytokines in a vaccination, owing to their regulatory properties on lymphocytes. Such cytokines comprise, for example, interleukin-12 (IL-12) which was shown to increase the protective actions of vaccines (cf. *Science* 268:1432-1434, 1995), GM-CSF and IL-18.

There are a number of compounds which enhance an immune response and which therefore may be used in a vaccination. Said compounds comprise co-stimulating molecules provided in the form of proteins or nucleic acids such as B7-1 and B7-2 (CD80 and CD86, respectively).

According to the invention, a bodily sample may be a tissue sample, including body fluids, and/or a cellular sample. Such bodily samples may be obtained in the conventional manner such as by tissue biopsy, including punch biopsy, and by taking blood, bronchial aspirate, sputum, urine, feces or other body fluids. According to the invention, the term "sample" also includes processed samples such as fractions or isolates of biological samples, e.g. nucleic acid or cell isolates.

The agents such as vaccines and compositions described herein may be administered via any conventional route, including by injection or infusion. The administration may be carried out, for example, orally, intravenously, intraperitoneally, intramuscularly, subcutaneously or transdermally. In one embodiment, administration is carried out intranodally such as by injection into a lymph node. Other forms of administration envision the *in vitro* transfection of antigen presenting cells such as dendritic cells with nucleic acids described herein followed by administration of the antigen presenting cells.

The agents described herein are administered in effective amounts. An "effective amount" refers to the amount which achieves a desired reaction or a desired effect alone or together with further doses. In the case of treatment of a particular disease or of a particular condition, the desired reaction preferably relates to inhibition of the course of the disease. This comprises slowing down the progress of the disease and, in particular, interrupting or reversing the progress of the disease. The desired reaction in a treatment of a disease or of a condition may also be delay of the onset or a prevention of the onset of said disease or said condition.

An effective amount of an agent described herein will depend on the condition to be treated, the severeness of the disease, the individual parameters of the patient, including age, physiological condition, size and weight, the duration of treatment, the type of an accompanying therapy (if present), the specific route of administration and similar factors. Accordingly, the doses administered of the agents described herein may depend on various of such parameters. In the case that a reaction in a patient is insufficient with an initial dose, higher doses (or effectively higher doses achieved by a different, more localized route of administration) may be used.

The pharmaceutical compositions described herein are preferably sterile and contain an effective amount of the therapeutically active substance to generate the desired reaction or the desired effect.

The pharmaceutical compositions described herein are generally administered in pharmaceutically compatible amounts and in pharmaceutically compatible preparation. The term "pharmaceutically compatible" refers to a nontoxic material which does not interact with the action of the active component of the pharmaceutical composition. Preparations of this kind may usually contain salts, buffer substances, preservatives, carriers, supplementing immunity-enhancing substances such as adjuvants, e.g. CpG oligonucleotides, cytokines, chemokines, saponin, GM-CSF and/or RNA and, where appropriate, other therapeutically active compounds. When used in medicine, the salts should be pharmaceutically compatible. However, salts which are not pharmaceutically compatible may used for preparing pharmaceutically compatible salts and are included in the invention. Pharmacologically and pharmaceutically compatible salts of this kind comprise in a non-limiting way those prepared from the following acids: hydrochloric,

hydrobromic, sulfuric, nitric, phosphoric, maleic, acetic, salicylic, citric, formic, malonic, succinic acids, and the like. Pharmaceutically compatible salts may also be prepared as alkali metal salts or alkaline earth metal salts, such as sodium salts, potassium salts or calcium salts.

A pharmaceutical composition described herein may comprise a pharmaceutically compatible carrier. The term "carrier" refers to an organic or inorganic component, of a natural or synthetic nature, in which the active component is combined in order to facilitate application. According to the invention, the term "pharmaceutically compatible carrier" includes one or more compatible solid or liquid fillers, diluents or encapsulating substances, which are suitable for administration to a patient. The components of the pharmaceutical composition of the invention are usually such that no interaction occurs which substantially impairs the desired pharmaceutical efficacy.

The pharmaceutical compositions described herein may contain suitable buffer substances such as acetic acid in a salt, citric acid in a salt, boric acid in a salt and phosphoric acid in a salt.

The pharmaceutical compositions may, where appropriate, also contain suitable preservatives such as benzalkonium chloride, chlorobutanol, paraben and thimerosal.

The pharmaceutical compositions are usually provided in a uniform dosage form and may be prepared in a manner known per se. Pharmaceutical compositions of the invention may be in the form of capsules, tablets, lozenges, solutions, suspensions, syrups, elixirs or in the form of an emulsion, for example.

Compositions suitable for parenteral administration usually comprise a sterile aqueous or nonaqueous preparation of the active compound, which is preferably isotonic to the blood of the recipient. Examples of compatible carriers and solvents are Ringer solution and isotonic sodium chloride solution. In addition, usually sterile, fixed oils are used as solution or suspension medium.

The present invention is described in detail by the figures and examples below, which are used only for illustration purposes and are not meant to be limiting. Owing to the description and the

examples, further embodiments which are likewise included in the invention are accessible to the skilled worker.

FIGURES

Figure 1. MHC binding prediction overview

Figure 2. Analysis of immunogenicity as a function of the M_{mut} score for 50 prioritized B16F10 mutations and 82 prioritized CT26.WT mutations (132 mutations in total), of which 30 were immunogenic. All vaccinations were performed with RNA. For B16F10 immunogenicity was assayed by challenging BMDCs with RNA and measuring the immune response of splenocytes with ELISPOT and FACS. For CT26.WT immunogenicity was assayed by challenging BMDCs with RNA and peptides separately and measuring the immune response of splenocytes with ELISPOT; a mutation was considered immunogenic if either peptide or RNA registered an immune response. **A** Cumulative distribution of immunogenic mutations as a function of the M_{mttt} score. The graph shows the total number of mutations below a given M_{mut} score (red), of these, the number of mutations that were immunogenic (blue), and the percent of immunogenic mutations from the total (black). **B** Histogram of percent of immunogenic mutations per M_{mut} bin for the following ranges: <0.3, (0.3, 1], >1. Errors shown are standard errors.

Figure 3. Analysis of B16F10 and CT26.WT immunogenicity as a function of M_{mut}. Cumulative distribution of immunogenic mutations as a function of the M_{mut} score for B16 **(A)** and CT26 **(C).** Histogram of percent of immunogenic mutations per M_{mut} bin for the following ranges: [0.1, 0.3], (0.3, 1], (**l,oo)** for B16 **(B)** and CT26 **(D).** Figures A and B are based on analysis of 50 B16F10 prioritized mutations, of which 12 were immunogenic. Figures C and D are based on analysis of 82 B16F10 prioritized mutations, of which 30 were immunogenic. For more details see legend of Fig. 2. Errors are standard errors.

Figure 4. Models of immunogenicity and control hypotheses. Class I immunogenicity, denoted by H_A, makes the assumption that both the WT and MUT epitopes are presented by

cells, and that the mutation sufficiency altered the physico-chemical properties of the amino acid so that the immune system registers this change and generates an immune response (denoted by the lightning bolt). The H_n hypothesis, serving as a control for H_A, is simply the inverted H_A hypothesis, namely, that the mutation did not significantly alter the physico-chemical properties of the amino acid and therefore has a lower likelihood of being "detected" by the immune system and generating an immune response. In class II immunogenicity ($H_B \cup H_C$) the WT epitope is not presented but the MUT epitope is presented. H_B and H_C are distinguished by high ($T > \tau$) versus low ($T < \tau$) T scores, respectively. Note that for $\alpha^* = \alpha$, the ***Hgci*** model for immunogenicity ($M_{mut} < \beta$) is a composite of all four groups: $H_{BCI} = \cup[H_A, H_B, H_C, H_n]$.

Figure 5. Hypothesized relation of the T score to immunogenincity. According to the class I immunogenicity model, during T cell development TCRs that bound strongly to the wild type epitope were deleted. Extant TCRs should exhibit only weak or no binding affinity to the wild type epitope (A). Epitopes that contain an amino acid substituion that has a high T score have similar physico-chemical properties to the wild type amino acid and therefore will likely have little impact on the binding affinity to extant TCRs (B). Epitopes that contain an amino acid substituion with a T score have a greater chance to increase the binding affinity to exact TCRs and therefore a greater likelihood to be immunogenic **(C).** In this schematic illustration, color coding is used to pair T cells with a matching peptide. Orange/yellow mutations represent mutations with high T scores (similar to the WT), where as blue/purple mutiations represent mutations with low T scores (significant physico-chemical difference compared to the WT).

Figure 6. Cumulative distribution of immunogenic mutations as a function of M_{mut}. A Comparison of the percent of immunogenic mutations that satisfy the baseline control hypothesis H_{BC1}: $\{M_{mut} \leq \beta\}$, with the percent of immunogenic mutations that satisfy the partial hypothesis H_A: $H_{BC1} \cap \{T \leq \tau\}$, the partial hypothesis H_{BC2}: $H_{BCI} \cap \{M_{mut} \leq a\}$ and the full hypothesis H_A: $H_{BC1} \cap \{M_{mM} \leq \alpha\} \cap \{T \leq \tau\}$, for $a = l$, r = l . B Comparison of the percent of immunogenic mutations that satisfy the baseline control hypotheses H_{BC1}: $\{M_{mut} \leq \beta\}$ with the percent of immunogenic mutations that satisfy the inverse partial hypotheses: $H_{BCI} \cap \{T > \tau\}$ and

$H_{BCL} \cap \{M_{mut} > a)$. The analysis in **A** and **B** are based on the pooled B16F10 and CT26.WT datasets, comprising of 132 mutations, of which 30 were immunogenic. Each data point in the graphs is based on ≥4 mutations.

Figure 7. **Cumulative distribution of immunogenic mutations as a function of the M_{mut} score.** Comparison of the percent of immunogenic mutations that satisfy the baseline control hypothesis H_{BCI}: $\{M_{mut} \leq \beta)$, with the percent of immunogenic mutations that satisfy the partial hypothesis H_A: $H_{BCI} \cap \{T \leq r\}$, the partial hypothesis H_{BC2}: $H_{BCI} \cap \{M_{mut} \leq a)$ *and* the full hypothesis H_A: $H_{BCI} \cap \{M_{mut} \leq \alpha\} \cap \{T \leq \tau\}$, given $\alpha = 1$, $\tau = 1$ for B16 **(A)** and CT26 **(C).** Comparison of the percent of immunogenic mutations that satisfy the baseline control hypotheses H_{BCI}: $\{M_{mut} \leq \beta\}$ with the percent of immunogenic mutations that satisfy the inverse partial hypotheses: H_{BCI} **n** $\{T > r\}$ and H_{BCL} **n** $\{M_{mut} > a)$ for B16 **(B)** and CT26 **(D).**

Figures A and B are based on analysis of the 50 B16F10 prioritized mutations, of which 12 were immunogenic. Figures C and D are based on the 82 B16F10 prioritized mutations, of which 30 were immunogenic. Each data point in the graphs is based on ≥4 mutations.

Figure 8. Controlling for WT immunogenicity. To check whether omitting MUT+/WT+ solutions had an impact on these findings we excluded from the dataset 9 MUT+/WT+ mutations and 2 mutations for which the WT has not been measured, leaving in total 121 mutations (43 B16 and 78 CT26) of which 19 were MUT+/WT- (5 for B16 and 14 for CT26). We again found the same trends as in the complete dataset, namely, highly non-linear response as a function of the M_{MUT} score, superiority of the H_A hypothesis over partial hypothesis, and inferiority of inverted hypotheses compared to the baseline control H_{BCI}. **A** Cumulative distribution of immunogenicity as a function of the M_{mut} score. **B** Histogram of percent of immunogenic mutations per M_{mut} bin. **C** Comparison of the percent of immunogenic mutations that satisfy the baseline control hypotheses H_{BCI} with H_A, H_{BC2} and H_A. **D** Comparison of the percent of immunogenic mutations that satisfy the baseline control hypotheses H_{BCI} with the inverse hypotheses. See Fig. 5 legend for additional details.

Figure 9. Fraction of immunogenic mutations as a function of RPKM. Red: all 50 B16 mutations and 82 CT26 mutations with no filtering (132 mutations in total). Blue: mutations passing the H_A hypothesis with $\alpha = 1, \beta = 0.5, r = 1$. **B.** Percent of immunogenic mutations for different RPKM ranges with no filtering. RPKM bins are: 1=(0,1],2=(1,5],3=(5,50],4=(50, ∞). **C** Percent of immunogenic mutations for different RPKM ranges under the H_A hypothesis with $\alpha = 1, \beta = 0.5, r = 1$. RPKM bins are: 1=(0,1], 2=(1, ∞). Errors are S.E.

Figure 10. Anchor and non-anchor position mutated class II immunogenic epitopes. Anchor position motifs were analyzed using SYFPEITHI.

Figure 11. Proposed models for immunogenic tumor-associated epitopes.

Figure 12. Example of a method for weighing rank position of mutations. For each mutation the rank position in the list of ranked mutations can be further weighed by the number of solutions for which the combination of HLA types for the patient, possible window lengths for the HLA type and mutation position within the epitope resulted in a solution with low M_{MUT} or resulted in a H_A and/or $H_B \cup H_C$ classification. Since all solutions per mutation potentially can be presented in parallel, this weighing factor may be an important contributor to the rank position of the mutation.

Figure 13. Example of scatter plot of all epitope solutions for mutation chrl4_52837882 from CT26 against M_{mut} and $\Delta M = M_{mut} - M_{wt}$.

EXAMPLES

The techniques and methods used herein are described herein or carried out in a manner known per se and as described, for example, in Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd Edition (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. All methods including the use of kits and reagents are carried out according to the manufacturers' information unless specifically indicated.

Example 1: Establishing a model for predicting immunogenicity of T cell epitopes

Previously we explored the immunogenicity of 50 somatic mutations identified in the B16F10 murine melanoma cell line (J. C. Castle et al., Exploiting the mutanome for tumor vaccination. Cancer Research 72, 1081 (2012)). These 50 mutations were selected from a pool of 563 expressed nonsynonymous somatic mutations primarily to maximize MHC class I expression (J. C. Castle et al., Exploiting the mutanome for tumor vaccination. Cancer Research 72, 1081 (2012)) (see also Example 2). For each mutation we predicted the minimal epitope, i.e., the epitope scoring the lowest MHC class I consensus score (Y. Kim et al., Nucleic Acids Research 40, W525 (2012)) (defined here as M_{mut}) when searching the space of all possible MHC class I alleles, potential epitope lengths and sequence windows (where to position the mutation) (J. C. Castle et al., Exploiting the mutanome for tumor vaccination. Cancer Research 72, 1081 (2012)). Measuring the immunogenicity of these mutations using RNA vaccination followed by peptide readout (see Example 2) confirmed earlier findings using peptide vaccination (J. C. Castle et al., Exploiting the mutanome for tumor vaccination. Cancer Research 72, 1081 (2012)), and showed that only 12 out of 50 mutations (24%) were immunogenic (Table 1), with MUT+/WT- sequences comprising only 10% for of all mutations tested.

Table 1. Number of immunogenic mutations after RNA vaccination of B16F10 and CT26.WT murine strains.

		MUT/WT immunogenicity					% immunogenic
		+/+	+/-	-/+	-/-	*Total*	*MUT+/total*
Strain	B16F10	7	5	0	38	50	24%
	CT26	2*	14*	1	63	80	20%
	Total	**9**	**19**	**1**	**101**	**130**	**21.50%**

Two CT26 MUT+ mutations were excluded from this table because their WT reactivity has not been measured yet. In total there were 18 MUT+ mutations out of 82 CT26 mutations measured thusfar, resulting in 22% success rate.

The results of the B16F10 murine test case demonstrate that naively selecting expressed nonsynonymous mutations with low M_{mut} scores (≤3.9) yields rather low success rates for predicting immunogenicity. Hence a better understanding of the mechanisms driving immunogenicity is required if personalized vaccines targeting tumor-specific neoantigens are to become effective therapies. In an effort to uncover additional variables that contribute to

immunogenicity we explored the immunogenicity of expressed nonsynonymous somatic mutations identified in a colorectal murine cell line CT26.WT. In total, 96 mutations were selected based on their M_{mut} scores (low vs. high), mean RPKM (low vs. high), and cellular localization (intra- vs. extra- cellular), and tested for immunogenicity using RNA vaccination with both peptide and RNA readout (see Example 2 for further details). Together with the B16F10 cell line, our dataset comprised of 132 epitopes, whose immunogenicity was measured *ex vivo* on murine splenocytes.

The MHC consensus score. To investigate the dependence of immunogenicity on M_{mut}, we plotted the cumulative percent of immunogenic mutations as a function of M_{mut}, that is, the percent of mutations with an M_{mut} score smaller than a given threshold (denoted by β) that were immunogenic. An analysis of the combined B16 and CT26 datasets spanning a total of 132 mutations reveals a highly nonlinear dependence of the immunogenicity success rate on M_{mut} (Fig. 2A). Fig. 2A shows that immunogenic mutations are enriched for extremely low M_{mut} scores (<~0.2). For $M_{mut} \leq 0.4$ the percent of immunogenic mutations peaks at ~60%, and quickly decays as M_{mut} increases, dropping below ~25% for $M_{mut} \geq 2$. The percent of immunogenic mutations with $M_{mut} \leq 0.3$ versus >0.3 was 44.4% compared with 17.1%, a statistically significant difference (P value = 0.004, Fisher's exact test, one tailed). A histogram of the percent of immunogenic mutations for three M_{mut} bins: <0.3, (0.3, 1] and >1 shows that the percent of immunogenic mutations drops as M_{mut} increases (Fig. 2B). The differences between the success rate of the lowest bin (M_{mut} <0.3), 44.4%, and both the central bin, 20.7%, and the highest bin ($M_{mut} > 1$), 15.8%, in Fig. 2B were statistically significant (P values = 0.05 and 0.004, respectively, Fisher's exact test, one tailed), indicating that for M_{mut} >~0.3 the success rate drops in a statistically significant manner. A similar trend in the success rate is also observed when analyzing B16 and CT26 mutanomes separately (Fig. 3).

Thus far our criteria for selecting mutations focused on presentation, and we have seen that restricting the MHC binding score of the mutated epitope allows prediction of immunogenic epitopes with up to 60% precision. Presentation, however, is a necessary but not sufficient condition to induce immunogenicity. By identifying additional criteria for TCR recognition we may be able to further improve the precision of our prediction. We hypothesized two mutually

exclusive mechanisms for driving immunogenicity, which we refer to as the class I and class II immunogenicity models.

Class I immunogenicity. In order for the TCR repertoire to recognize a mutated epitope and generate an immune response we hypothesized that three conditions must be satisfied (¾ Fig. **4):** (i) the wild type epitope, at some point during the development of the organism, was presented to the immune system leading to deletion of matching TCRs via strong TCR/pMHC binding, (ii) the mutated epitope is presented, and (iii) the physico-chemical properties of the mutated amino acid are sufficiently "different" from the wild type amino acid (by some metric that we shall define below) so that the TCR repertoire is able to "detect" or "register" this substitution. Conditions (i) and (ii) ensure that the immune system is actually exposed to the change, i.e., the mutation. Condition (iii) requires that the mutation significantly change the physico-chemical character of the wild type amino acid so that the binding affinity of the mutated epitope to extant (undeleted) TCR potentially increases, thereby turning on the signaling cascade that leads to an immune response (Fig. **5).**

The TCR recognition score. Class I immunogenicity models requires a metric to estimate the physico-chemical difference between two amino acids. It is well known in molecular evolution that amino acids that interchange frequently are likely to have chemical and physical similarities whereas amino acids that interchange rarely are likely to have different physico-chemical properties. The likelihood for a given substitution to occur in nature compared with the likelihood for this substitution to occur by chance is measured by log-odds matrices. The patterns observed in log-odd matrices imposed by natural selection "reflect the similarity of the functions of the amino acid residues in their weak interactions with one another in the three dimensional conformation of proteins" (M. O. Dayhoff, R. M. Schwartz, B. C. Orcutt, A model for evolutionary change. MO Dayhoff, ed. Atlas of protein sequence and structure Vol.5, 345 (1978)). We therefore used evolutionary based log-odds matrices, which we refer to here as "T scores" to reflect TCR recognition, as effective scoring matrices for cancer associated amino acid substitutions. Substitutions with positive T scores (i.e., log-odds) are likely to occur in nature, and hence correspond to two amino acids that have similar physico-chemical properties. The class I model predicts that substitutions with positive T scores would have a lower likelihood of

being immunogenic. Conversely, substitutions with negative T scores reflect substitutions that are unlikely to occur in nature and hence correspond to two amino acids that have significantly different physico-chemical properties. According to our model, such substitutions would have a greater chance of being immunogenic. We compared different methods of estimating log-odds matrices and found results to be largely robust to the exact method chosen. The maximum likelihood (ML) based estimation approach known as WAG (S. Whelan, N. Goldman, Molecular biology and evolution 18, 691 (2001)), using a PAM (point accepted mutation) distance of 250 appeared to separate predicted immunogenic from non-immunogenic mutations best, and therefore we present results with this matrix (see Example 2 for further details).

Class II immunogenicity. In the class II model for immunogenicity we hypothesize that a mutation is likely to be immunogenic if the immune system has never before seen the wild type epitope, and is therefore challenged by the mutated epitope. Therefore in order for a mutation to be immunogenic in this model we hypothesized that two conditions must be satisfied: (i) the wild-type epitope was never presented to the immune system, (ii) the mutated peptide is presented. These conditions can be co-satisfied if, for example, the mutation hits an anchor position thereby changing a "nonbinder" epitope into a "binder". Formally, class II immunogenicity can be separated into two sub-hypotheses: high T scores (H_B in Fig. 4) and low T scores (H_C in Fig. 4). However, since the assumption is that the wild type epitope is not presented, the nature of the amino acid substitution is not expected to have an impact on TCR recognition and we shall therefore equate class II immunogenicity with the united hypothesis: $H_B \cup H_C$.

Testing class I immunogenicity. The assumptions of class I immunogenicity (H_A in Fig. 4) can be restated mathematically as follows: we require that the wild type epitope is presented ($M_{wt} \leq \alpha$), the mutated epitope is presented ($M_{mut} \leq \beta$), and the amino acid substitution is non-trivial ($T \leq \tau$), where M_{wt} is defined as the MHC consensus score of the mutated epitope (same HLA allele and window length) replacing the mutated amino acid with the wild type amino acid, and T denotes the T score. Since all three conditions are necessary, we expect that the precision of the H_A classifier will be higher compared to a classifier based on M_{MUT} alone (H_{CI} in Fig. 4) or compared to the partial hypotheses: $H_{BCl} \cap \{M_{wt} \leq \alpha\}$ and $H_{CI} \cap \{T \leq \tau\}$. We therefore

calculated the percent of immunogenic mutations (number of true positives divided by the sum of true positives and false positives) as a function of β for $H_B ci$, for the partial hypothesis H_A: $H_{BCl} \cap \{T \leq \tau\}$ and for the partial hypothesis H_{BC2}: $H_{BC1} \cap \{M_{wt} \leq a)$. We found that a conservative threshold for r in the range of ≈0.5 to 1 performed best (the range of the WAG250 matrix is from -5.1 (F<→G substitution) to +5.4 (F«→Y substitution). We also found that a can be restricted conservatively compared to β, setting $a \approx 1$. Fig. 6A indeed shows that, when considering the pooled mutanome of B16 and CT26, classifiers based on H_{BC2}, and H_A attained greater precision than the baseline control hypothesis H_{BCI}- Moreover, a classifier based on the complete hypothesis H_A attained greater precision than the partial hypotheses H_{BCI}, and H_{BC_2}, thereby demonstrating an additive effect. The same conclusions hold when analyzing the B16 and CT26 datasets separately (Fig. 7).

Since the conditions $M_{wt} \leq a$ and $T \leq \tau$ are postulated to be necessary conditions for immunogenicity, one would except that a classifier based on either the condition $H_{BCl} \cap \{T > r\}$ or the condition H_{BCL} n $\{M_{WT} > a)$ (i.e., negating the secondary condition) would perform worse than $H_{BC}I$- Indeed we found that this is the case for B16 and CT26 when analyzed together (Fig. 6B) or separately (Fig. 7). Therefore we conclude that the B16 and CT26 datasets support both together and separately the H_A hypothesis. Omitting mutations where the WT RNA also showed reactivity did not affect these conclusions (Fig. 8).

Controlling for the H_A hypothesis. Although mutations with high T scores may still be immunogenic, a hypothesis that enriches for such mutations should statistically enrich for non-immunogenic mutations. Therefore if we compare the H_A hypothesis ($H_{gC2} \cap \{T < r\}$) with its inverse, $H_{gC2} \cap \{T > \tau\}$ ($H_{„}$ in Fig. 4), we should observe a statistically significant depletion of immunogenic mutations. Table 2 indeed shows that for $M_{mut} \leq \beta = 0.5$, $M_{wt} \leq \alpha = 1$, and T<τ=1, H_A outperforms $H_{„}$, with a success rate of 52.5% (n=21) compared to 21.4% (n=14; P=0.068, one tailed Fisher's exact test).

Table 2. Percent of immunogenic mutations under various hypotheses based on the B16 and CT26 pooled datasets comprising 133 mutations.

		Hypothesis parameters			
		M_{mut} threshold (β)	M_{wt} threshold (α)	T score threshold (τ)	% of immunogenic mutations
Hypothesis	H_{BC1}	≤ 2			22/83 (26.5%)
		≤ 1			18/56 (32.1%)
		≤ 0.1			6/10 (60%)
	$H_{A'}$	≤ 0.1		≤ 1	6/9 (66.7%)
	H_{BC2}		≤ 1		6/9 (66.7%)
	H_A			≤ 1	6/8 (75%)
	$H_{A'}$	≤ 0.2		≤ 1	8/14 (57.1%)
	H_{BC2}		≤ 1		9/17 (52.9%)
	H_n			>1	1/6 (16.7%)
	H_A			≤ 1	8/12 (66.7%)
	H_{BC2}	≤ 0.5	≤ 1		14/34 (41.2%)
	H_n			>1	3/14 (21.4%)
	H_A			≤ 1	11/21 (52.4%)
	$H_B \cup H_C$	≤ 0.5	>1		1/6 (16.7%)

H_A also performs better than the baseline control H_{BC2}, which achieves 41.2% (n=35). As we decrease β the difference between the success rates of H_A and $H_{„}$ become larger since the more stringent the condition on β, the more false positives are removed from the H_A group. For example, for β =0.25 the success rate of the H_A group was 67% (n=14) compared to a success rate of 17% (n=6) for group H_N (P=0.066, one tailed Fisher's exact test) - see Table 3.

Table 3. Ranked list of 133 measured B16F10/CT26.WT mutations that satisfy the basic control hypothesis HB_{CI} $(M_{MUT}<0.25)$ broken down into the three disjoint hypothesis classes: HA hypothesis for immunogenic mutations $(M_{wt}<0.8,\ T<0.5)$, $H_{„}$/inverse H_A hypothesis enriching for non-immunogenic mutations $(M_{WT}<0.S,\ T>0.5)$, and $H_B \cup H_C$ hypothesis for immunogenic mutations (M_{wt}>0.8). H_A and $H_B \cup H$c candidates are proposed to be ranked based on the relative importance of distinguishing variables. For H_A the proposed order is: M_{MUt} (descending) → T score (descending) → M_{WT} descending. For $H_B \cup H_C$ the proposed order is: M_{MM} (descending) → M_{WT} (ascending). Errors are standard errors.

Sample	Mut	Response	RNA response (MUT)	RNA response (WT)	Symbol (Ingenuity)	MHC1 allele	epitope (MUT)	epitope (WT)	Mean Expression	M_{mut}	M_{wt}	T score (WAG 250)
Class 1imtnunogenicity (H_A): 67±14% success rate												
B16	33	CD8	yes	no	PBK	H-2-Db	AAVILRDALHM	AAVILRVALHM	19.6	0.1	0.1	-2.7
CT26	14		no	no	Nphp3	H-2-Dd	GGPGSEKSL	GGPGSGKSL	0.6	0.1	0.1	-0.5
B16	37		no	no	DPF2	H-2-Db	LALPNNYCDV	LALPNNYCDF	20.0	0.1	0.1	-0.1
B16	25	CD4	yes	yes	PLOD2	H-2-Db	SHLNNDVWQI	SHLNNDFWQI	21.7	0.1	0.1	-0.1
CT26	37	CD4	yes	no	Dhx35	H-2-Kd	YYMRDVIAI	YYMRDVTAI	5.5	0.1	0.1	-0.1
CT26	26	CD8	yes	no	E2f8	H-2-Kd	TYLQPAQAQM	IYLQPAQAQM	29.5	0.1	0.1	-0.1
B16	21	CD4	yes	no	ATP11A	H-2-Db	QSLGFTYL	QRLGFTYL	42.7	0.1	0.8	-0.1
CT26	13		yes	no	Nphp3	H-2-Kd	EYWASRALDS	EYWASRALGS	0.5	0.1	0.1	0.1
CT26	16		no	no	H2-Q8	H-2-Kd	GYLQFAYEGC	GYLQFAYEGR	5.1	0.2	0.2	-1.7
B16	46	CD4	yes	yes	ACTN4	H-2-Kb	VTFQAFIDVMS	VTFQAFIDFMS	125.6	0.2	0.1	-0.1
CT26	15		no	no	Slc41a2	H-2-Kd	PYLTALDDLL	PYLTALGDLL	2.8	0.2	0.2	0.1
CT26	27		yes		Agxt2l2	H-2-Dd	AGGLFVADAI	AGGLFVADEI	29.7	0.2	0.3	0.1
Class II immunogencity (H_BUH_C): 0% success rate												
CT26	25		no	no	Pcdhga8	H-2-Dd	VGINFLQSYQ	VGINSLQSYQ	26.2	0.1	2.7	-2.1
CT26	1		no	no	Gm8909	H-2-Dd	TRPARDGTF	TRPAGDGTF	11.0	0.2	4 6"	-1.0
CT26	40		no	no	Zfp449	H-2-Ld	EPQIAMDDM	EPQIDMDDM	0.7	0.2	5	-0.2
$H_{„}$: 17±15% success rate												
B16	49		no	no	FAT1	H-2-Db	IAMQNTTQL	IAIQNTTQL	18.8	0.1	0.1	2.6
B16	36		no	no	TM9SF3	H-2-Kb	AIYHHASRAI	AIYYHASRAI	51.1	0.2	0.3	2.7
CT26	39	CD8	yes	no	Als2	H-2-Kd	SYIALVDKNI	SYLALVDKNI	4.7	0.2	0.1	2.7
CT26	2		no	no	Snap47	H-2-Dd	VIPILEMQF	VIPILEVQF	15.5	0.2	0.2	1.8
CT26	17		no	no	H2-Q8	H-2-Kd	GYLQFAYDGR	GYLQFAYEGR	5.1	0.2	0.2	2.9
CT26	38		no		Cspp1	H-2-Kd	VYLNLLLKFT	VYLNLFLKFT	0.7	0.2	0.1	2.0

An Example of additional weighing factors that may further improve immunogenicity ranking is given in Example 3.

More generally the list of mutations that satisfy H_BCI ($M_{mut}<\beta$) can be classified into the three categories: H_A, H_N, and H_BUH_C (Table 3), where H_A enriches for immunogenic mutations, $H_{„}$ enriches for non-immunogenic mutations. In the case of B16 and CT26, all three candidates in the H_BUH_C group were non-immunogenic, contrary to our expectation. However, if a more realistic threshold a^* for M_{wt} is chosen such that $\alpha^* \gg \alpha$, then there would be no predictions that could be tested for H_BUH_C.

Maximal precision of immunogenicity classifiers. According to Table 1 the average success rate for prediction immunogenicity in the combined B16 and CT26 datasets was 22.7% (=30/132). By applying the most stringent threshold on the M_{mut} score ($\beta = 0.1$), the precision of an immunogenicity classifier increases to 60% (=6/10; HB_{CI} in Table 2). By combining *HBCI* with either the $M_{wt} \le a$ condition or the $T \le \tau$ condition ($\alpha = 1, r = 1$) precision is increased to

66.7% (=6/9). The H_A based classifier, which combines both criteria, results in an additive response, which increases the precision to 75% (=6/8) (Table 2).

B16 epitope MUT33. The H_1-class epitope that was ranked the highest by all evolutionary models (except the PAM matrix) in the pooled B16/CT26 dataset was B16's MUT33 (see Table 3). Further analysis revealed that MUT33 indeed invoked an MHC class I restricted CD8+ response and exhibited *ex vivo* immunogenicity against the minimal predicted epitope (data not shown).

Role of gene expression. Plotting the fraction of immunogenic mutations (no. of immunogenic mutations with RPKM values below a given threshold over the total no. of immunogenic mutations) as a function of RPKM values for B16 and CT26 indicates that this ratio somewhat stagnates at very low RPKM values (Fig. 9A). This effect is observed whether the H_A criterion is applied or not. Plotting the percent immunogenic mutations for different RPKM bins (Fig. 9B and C) suggests that RPKM values $\leq \sim 1$ have a somewhat lower success rate (both with or without applying the H_A filtering hypothesis), although suggestive, it should be noted that these results are within the range of error.

Survey of published CD8+ epitopes. We were next interested to see if published T cell-defined tumor antigens with single amino acid substitutions eliciting CD8+ restricted response fulfilled our models for immunogenicity. Of the 17 epitopes that were published (P. Van der Bruggen, V. Stroobant, N. Vigneron, B. Van den Eynde. (Cancer Immun, http://www.cancerimmunity.org/peptide/, 2013)) (Table 4), five satisfied the criteria for H_A (α=0.7, β=0.2, x=0.5), four satisfied the criteria for $H_C \cup H_B$ (α=2.2, β=0.4), and two satisfied the H_n criterion (α=0.6, β=0.3, τ= 1.7).

Table 4. Published epitopes with single amino acid substitution generating CD8+ responses. See Example 2 for list of references. Anchor position mutations in the $H_B \cup H_C$ group are highlighted in red.

Hypothesis	M_{mut} threshold	M_{wt} threshold	T score threshold	M_{mut}	M_{wt}	T score*	Gene	MUT epitope	WT epitope	Source	
H_A	≤0.2	≤0.7	≤0.5	0.10	0.10	-2.1	SIRT2	KIFSEVTLK	KIFSEVTPK	1	←MZ7-MEL
				0.10	0.10	-0.1	SNRPD	SHETVIIEL	SHETVTIEL	1	←MZ7-MEL
				0.14	0.22	0.5	ME1	FLDEFMEGV	FLDEFMEAV	2	
				0.20	0.70	-1.0	RBAF600	RPHVPESAF	GPHVPESAF	1	←MZ7-MEL
				0.15	0.20	-2.1	PRDX5	LLLDDLLVSI	LLLDDSLVSI	3	
H_n	≤0.3	≤0.6	≥1.7	0.25	0.20	1.7	N-ras	ILDTAGREEY	ILDTAGQEEY	4	
				0.30	0.60	1.9	EEF2	ETVSEQSNV	ETVSEESNV	10	
H_B	≤0.25	≥2.25	>0.5	0.15	10.20	1.2	EFTUD2	KILDAVVAQK	KILDAVVAQE	1	←MZ7-MEL
				0.25	2.25	1.2	MYO1B	KINKNPKYK	EINKNPKYK	5	
H_C	<0.4	>10	≤0.5	0.20	10.95	-2.3	FLT3	YVDFREYEYY	YVDFREYEYD	6	
				0.38	10.58	-2.1	CTNNB1	SYLDSGIHF	SYLDSGIHS	7	
other	≥0.65	≥1		0.65	2.70	-2.1	OS-9	KELEGILLL	KELEGILLP	8	
				1.20	1.00	0.1	GPNMB	TLDWLLQTPK	TLGWLLQTPK	1	←MZ7-MEL
				2.68	3.58	1.0	ACTN4	FIASNGVKLV	FIASKGVKLV	9	
				4.20	4.60	-1.2	P53	VVPCEPPEV	VVPYEPPEV	11	
				26.40	29.75	-1.7	CDK4	ACDPHSGHFV	ARDPHSGHFV	12	
				33.30	29.80	-2.9	K-ras	VVVGAVGVG	VVVGAGGVG	13	

* Based on WAG250 log-odds matrix, color legend: T ≤ 0.5, 0.5< T≤1, T>1

Thus, the H_A and $H_C \cup H_B$ hypotheses together accounted for roughly 50% of the published epitopes. Interestingly, 3 out of the **4** published epitopes that satisfied the $H_C \cup H_B$ condition (red boxes in Table **4)** had an M_{wt} score that was larger than 10 due to anchor position mutations (Fig. 10). Since the requirement for the $H_C \cup H_B$ hypothesis is that the probability that any cell present the wild type epitope during the development of the organism is kept negligibly small it is expected that the threshold for M_{w_i} should be kept high, i.e., $\alpha^* \gg \alpha$. Indeed when increasing a from 0.8 to >3 the false positives for B16/CT26 in Table 3 disappear. Therefore a more realistic threshold for M_{wt} under the $H_C \cup H_B$ hypothesis may be somewhere between 3 and 10.

The MZ7-MEL cell line. To test the ability of our immunogenicity models to predict immunogenic epitopes in a human tumor model setting, we explored the MZ7-MEL cell line, established in 1988 from a splenic metastasis of a patient with malignant melanoma (V. Lennerz et al., Proceedings of the National Academy of Sciences of the United States of America 102, 16013 (2005)). Screening of a cDNA library from MZ7-MEL cells with autologous tumor-reactive T cells revealed at least five neoantigens capable of generating CD8+ responses (V. Lennerz et al., Proceedings of the National Academy of Sciences of the United States of America 102, 16013 (2005)). This constitutes the largest set of CD8+ neoantigens derived from a patient to date. Applying our immunogenicity models to these epitopes we found that three neoantigens were classified as H_A epitopes, and one neoantigen, an anchor position mutation, was classified

as an $H_B U H_C$ epitope (arrows in Table 4, and Fig. 10). Thus, four of the five epitopes could be explained by our immunogenicity models.

To test our ability to predict these epitopes *de novo* in the MZ7-MEL cell line we sequenced the exome of the MZ7-MEL cell line (see Methods). In total 743 expressed nonsynonymous mutations were identified. All five mutations previously identified by Lennerz et al. (V. Lennerz et al., Proceedings of the National Academy of Sciences of the United States of America 102, 16013 (2005)) were found. We then calculated for each mutation the T score, M_{mut} and M_{wt}, reporting also the HLA allele and epitope that resulted in the minimal MHC consensus score for the given mutation. Mutations were classified into one of three groups: H_A, $H_B U H_C$, and H_n using the thresholds $\alpha = 0.3\frac{1}{4}$ $\beta = 0.2, \tau = 0.5$ (pi_{us} the condition RPKM >0.2), and then ranked based on their potential to be immunogenic, as explained in Table 3. We found that out of 743 mutations, 32 mutations satisfied the H_A criteria (Table 5), 12 satisfied the $H_B U Hc$ criterion (Table 6) and 15 satisfied the H_n criterion.

Table 5. //^-classified MZ7-MEL cell mutations. 32 of the 743 expressed nonsynonymous mutations in MZ7-MEL were classified as H i-immunogenic using the thresholds: $\alpha = 0.8, \beta = 0.2$ and $\tau = \mathbf{0.5}$. Rank is based on an M_{mut} (descending) $\rightarrow$ T score (descending) sorting scheme. Immunogenic neoantigens identified by Lennerz *et al.* are highlighted in yellow. In addition RPKM was required to exceed 0.2.

Rank*	Gene	M_{mut}	T score**	M_{wt}	Mean exp	
1	DPH2	0.1	-2.9	0.1	10.7	
	ADHFE1	0.1	-2.9	0.1	2.7	
2	DDX41	0.1	-2.1	0.1	24.8	
	SIRT2	0.1	-2.1	0.1	15.7	←MZ7-MEL
3	PRIC285	0.1	-1.0	0.4	5.4	
4	CSTF3	0.1	-0.5	0.1	11.2	
	ETFDH	0.1	-0.5	0.1	10.8	
5	MED12	0.1	-0.1	0.1	21.9	
	SNRPD1	0.1	-0.1	0.1	18.0	←MZ7-MEL
	MLLT6	0.1	-0.1	0.2	14.7	
	AFAP1	0.1	-0.1	0.2	5.1	
6	MAP3K11	0.1	0.1	0.1	41.9	
7	DHX30	0.1	0.3	0.1	34.4	
	ALK	0.1	0.3	0.1	0.4	
	CHMP4B	0.1	0.3	0.7	52.8	
8	HADHB	0.1	0.5	0.1	60.6	
	SUPT6H	0.1	0.5	0.1	25.4	
	C12orf35	0.1	0.5	0.1	3.1	
	ZDHHC5	0.1	0.5	0.4	27.6	
9	WIPF1	0.15	-2.1	0.15	37.2	
	ZNF740	0.15	-2.1	0.5	9.7	
10	MLL	0.15	0.5	0.3	3.7	
11	KIAA1715	0.2	-2.1	0.1	6.3	
	CHD8	0.2	-2.1	0.2	10.3	
12	DDX28	0.2	-1.7	0.2	7.2	
13	MAPK1IP1L	0.2	-1.0	0.2	12.1	
	UBR4 (RBAF600)	0.2	-1.0	0.7	11.8	←MZ7-MEL
14	TRAK2	0.2	-0.1	0.5	21.8	
15	MAEA	0.2	0.0	0.25	25.3	
16	KLHL13	0.2	0.1	0.2	3.4	
17	FOSL2	0.2	0.3	0.2	9.4	
18	UTRN	0.2	0.5	0.15	6.3	

*Rank is based on M_{mut} and the T score

** T score is based on the WAG250 log-odds matrix

Table 6. ***HBUH*c-classified MZ7-MEL cell mutations.** 12 of the 743 expressed nonsynonymous mutations in MZ7-MEL were classified as ***HB*U*H*^immunogenic using the thresholds: $\alpha^* = 0.8$, $\beta \approx 0.2$ and RPKM > 2. Rank is based on a M_{mut} (descending) → M_{wt} (ascending) sorting scheme. Immunogenic neoantigens identified by Lennerz et al. are highlighted in yellow.

Rank*	Gene	M_{mut}	M_{wt}	Mean exp	
1	NF1	0.1	1.4	6.8	
2	MESP2	0.1	1.3	0.3	
3	EFTUD2 (SNRP116)	0.15	10.2	22.0	←MZ7-MEL
4	SEC31A	0.15	2.55	33.3	
5	*ZNF335*	*0.2*	*18.35*	*8.3*	
6	CPEB1	0.2	4.2	6.0	
7	UBAC2	0.2	2.8	7.5	
8	FANCF	0.2	2.4	3.9	
	RAPH1	0.2	2.4	3.0	
9	ZNF557	0.2	1.4	1.5	
10	TLK2	0.2	1.35	2.9	
11	ST7	0.2	1.15	10.9	

*Rank is based on M_{mut} and M_{wt}

Of the 32 mutations classified as H_A, the three $H^{\wedge}$-class mutations identified by Lennerz *et al.* (SIRT2, SNRPD1 and RBAF600) were ranked in 2nd, 4th and 13th positions out of 18 rank-classes, using a M_{mu_i}->T score ranking scheme (see Table 3). Of the 12 mutations classified as $H_B U H_C$, the forth Lennerz *et al.* mutation (SNRP116) ranked in the 3rd position. Moreover, if a higher (more realistic) threshold for M_{wl} was employed (e.g., $\alpha^{*} \sim 5$) then the forth Lennerz *et al.* mutation is ranked in the 1st position (together with just one additional anchor position mutation - Table 7). Finally, the four Lennerz *et al.* mutations were predicted to have the correct HLA allele, epitope length and mutation position as reported by the authors.

Table 7. $H_B U H$c-classified MZ7-MEL cell mutations. 2 of the 743 expressed nonsynonymous mutations in MZ7-MEL were classified as $H_B U H$c-immunogenic using the thresholds: $\alpha^{*} = 5$, $\beta = 0.2$ and RPKM > 2. Rank is based on a M_{mut} (descending) $\rightarrow$ M_{wl} (ascending) sorting scheme. Immunogenic neoantigens identified by Lennerz et al. are highlighted in yellow.

Rank*	Gene	M_{mut}	M_{wt}	Mean exp	
1	EFTUD2 (SNRP116)	0.15	10.2	22.0	←MZ7-MEL
2	ZNF335	0.2	18.35	8.3	

*Rank is based on M_{miJt} and M_{wt}

Conclusions

The analysis of the B16 and CT26 datasets support a model where immunogenicity is conferred if three conditions are satisfied: the wild type peptide is presented, the mutated peptide is presented, and the amino acid substitution has a sufficiently low log-odds score (Fig. 11A). This model for immunogenicity, which we refer to as class I immunogenicity, is further supported in the human melanoma cell line model, MZ7-MEL. The MZ7-MEL model and published CD8+ restricted neoantigens support a second model, which we refer to as class II immunogenicity, in which the wild type epitope is not presented, but a substitution (e.g., in an anchor position) leads to a significant increase in the MHC consensus score (>5 to 10), resulting in a novel, never-before-seen epitope (Fig. 1IB). This framework for defining immunogenicity is captured with a

three-variable classification scheme (M_{mut}, M_{wt}, T score). Using this classification scheme we were able to reduce the MZ7-MEL 743 mutations to a list of 34 mutations, with 3 of the 5 Lennerz ***et al.*** epitopes ranking in the top 5 classes.

Table 7 demonstrates that class II immunogenic mutations are rare. Out of 743 mutations only 2 were classified as class II immunogenic (using a realistic threshold for M_{wt}) compared with roughly 30 class I immunogenic mutations. A paucity of $H_B \cup H_C$-class mutations was also observed in the mouse melanoma models (Table 8). This observation underscores the importance of class I immunogenic of mutations for personalize vaccines, which are expected to be the dominate type mutations found in patient samples that can be used for vaccination. At the same time, the fact that one of the five epitopes found by Lennerz ***et al.*** was class II immunogenic may indicate that class II immunogenic mutations are more potent or somehow selected by the immune system.

Table 8. Number of candidate H_A and $H_B \cup H_C$ mutations in different tumor models. ($\alpha = 0.8, \alpha^* = 5. \beta = 0.2. \tau = 0.5$).

		Hypothesis	
		H_A	$H_B \cup H_C$
Strain	B16	27	0
	CT26	13	0
	MZ7-MEL	35	2

Example 2: Materials and methods

The materials and methods used in Example 1 are described below:

Animals

C57BL/6J and Balb/cJ mice (CRL) were kept in accordance with federal and state policies on animal research at the University of Mainz.

Cells for melanoma and colorectal murine tumor model

B16F10 melanoma cell line (Product: ATCC CRL-6475, Lot Number: 58078645) and CT26.WT colon carcinoma cell line (Product: ATCC CRL-2638, Lot Number: 58494154) were purchased in 2010 from the American Type Culture Collection. Early (3rd, 4th) passages of cells were used for sequencing experiments. Cells were routinely tested for Mycoplasma. Re-authentification of cells has not been performed since receipt. MZ7-MEL cell line (established January 1988) and an autologous Epstein-Barr virus-transformed B cell line were obtained from Dr. Thomas Wolfel (Department of Medicine, Hematology Oncology, Johannes Gutenberg University).

Synthetic peptides

Peptides were purchased from Jerini Peptide Technologies (Berlin, Germany) or synthesized from the TRON peptide facility. Synthetic peptides were 27 amino acids long with the mutated (MUT) or wild-type (WT) amino acid on position 14.

Immunization of mice

Age-matched female C57BL/6 or Balb/c mice were injected intravenously with 20 µg in vitro transcribed mRNA formulated with 20 µï Lipofectamine™ RNAiMAX (Invitrogen) in PBS in a total injection volume of 200 µl (3 mice per group). The mice were immunized on day 0, 3, 7, 14 and 18. Twenty-three days after the initial injection mice were sacrificed and splenocytes were isolated for immunological testing (see ELISPOT assay). DNA-sequences representing one (Monoepitope) or two mutations (Biepitope) were constructed using the sequence of 27 amino acids (aa) with the mutation on position 14 and cloned into the pSTl-2BgUTR-A120 backbone (S. Holtkamp et al., Blood 108, 4009 (2006)). *In vitro* transcription from this template and

purification were previously described (S. Kreiter et al., Cancer Immunology, Immunotherapy 56, 1577 (2007)).

Enzyme-linked immunospot assay

Enzyme-linked immunospot (ELISPOT) assay (S. Kreiter et al., Cancer Research 70, 9031 (2010)) and generation of syngeneic bone marrow derived dendritic cells (BMDCs) as stimulators were previously described (L. MB et al., J. Immunol. Methods 223, 77 (1999)). For the B16F10 model BMDCs were peptide pulsed (6 μg/ml), with the indicated mutation, the corresponding wild-type or with control peptide (VSV-NP). For the CT26 model in addition to the restimulation with peptides BMDCs were transfected with the corresponding *in vitro* transcribed mRNA and used for restimulation, as well. For the assay, 5×10^4 BMDCs were coincubated with 5×10^5 freshly isolated splenocytes in a microtiter plate coated with anti-IFN-γ antibody (10 μg/mL, clone AN 18; Mabtech). After 18 hours at 37°C, cytokine secretion was detected with an anti-IFN-γ antibody (clone R4-6A2; Mabtech). Spot numbers were counted and analyzed with the ImmunoSpot® S5 Versa ELISPOT Analyzer, the ImmunoCaptureTM Image Acquisition software and the ImmunoSpot® Analysis software Version 5. Statistical analysis was done by student's t-test and Mann-Whitney test (non-parametric test). Responses were considered significant with a p-value < 0.05.

Intracellular cytokine assay

Aliquots of the splenocytes prepared for the ELISPOT assay were subjected to analysis of cytokine production by intracellular flow cytometry. To this end 2 x 10^6 splenocytes per sample were plated in culture medium (RPMI + 10% FCS) supplemented with the Golgi inhibitor Brefeldin A (10μg/mL) in a 96-well plate. Cells from each animal were restimulated for 5h at 37°C with 2 x 10^5 peptide pulsed or RNA-transfected BMDCs. After incubation the cells were washed with PBS, resuspended in 50μ1PBS and extracellularly stained with the following anti-mouse antibodies for 20 min at 4°C: anti-CD4 FITC, anti-CD8 APC-Cy7 (BD Pharmingen). After incubation the cells were washed with PBS and subsequently resuspended in IOOμL Cytofix/Cytoperm (BD Bioscience) solution for 20 min at 4°C for permeabilization of the outer membrane. After permeabilization the cells were washed with Perm/Wash-Buffer (BD Bioscience), resuspended in **50**μI-/saηιρ ῖe in Perm/Wash-Buffer and intracellularly stained with

the following anti-mouse antibodies for 30 min at 4°C: anti-IFN- γ PE, anti-TNF-a PE-Cy7, anti-IL2 APC (BD Pharmingen). After washing with Perm/Wash-Buffer the cells were resuspended in PBS containing 1% paraformyldehyde for flow cytometry analysis. The samples were analyzed using a BD FACSCanto™ II cytometer and FlowJo (Version 7.6.3).

Next generation sequencing

Nucleic acid extraction: DNA and RNA from bulk cells and DNA from mouse tissues were extracted using Qiagen DNeasy Blood and Tissue kit (for DNA) and Qiagen RNeasy Micro kit (for RNA).

DNA exome sequencing: Exome capture for B16F10, C57BL/6J and CT26.WT and DNA re-sequencing for Balb/cJ were performed in triplicates as previously described (J. C. Castle et al., Exploiting the mutanome for tumor vaccination. Cancer Research 72, 1081 (2012)). Exome capture for MZ7-MEL/EBV-B DNA re-sequencing was performed in duplicates using Agilent XT Human all Exon 50 Mb solution-based capture assay, designed to capture all protein coding regions. 3 μg purified genomic DNA (gDNA) was fragmented to 150-200 bp using a Covaris S2 ultrasound device. Fragments were end repaired, 5' phosphorylated and 3' adenylated according to the maufacturer's instructions. Agilent indexing specific paired-end adapters were ligated to the gDNA fragments using a 10:1 molar ratio of adapter to gDNA. 4 cycle pre-capture amplification was done using Agilent's InPE 1.0 and SureSelect indexing pre-capture PCR primers and Herculasell polymerase. 500 ng of adapter ligated, PCR enriched gDNA fragments were hybridized to Agilent's exome capture baits for 24 hrs at 65 °C. Hybridized gDNA/RNA bait complexes where removed using streptavidin coated magnetic beads, washed and the RNA baits cleaved off during elution in SureSelect elution buffer. The eluted gDNA fragments were PCR amplified post-capture for 10 cycles using SureSelect Indexing Post-Capture PCR and index PCR primers and Herculasell polymerase. All cleanups were done with l,8x volume of Agencourt AMPure XP magnetic beads. All quality controls were done using Invitrogen's Qubit HS assay and fragment size was determined using Agilent's 2100 Bioanalyzer HS DNA assay. Exome enriched gDNA libraries were clustered on the cBot using Truseq SR cluster kit v2.5 using 7 pM library and 1x100 bps were sequenced on the Illumina HiSeq2000 using Truseq SBS kits.

RNA gene expression profiling (RNA-Seq): Barcoded mRNA-seq cDNA libraries were prepared in duplicate, from 5 µg of total RNA (modified Illumina mRNA-seq protocol using NEB reagents) mRNA was isolated using Seramag Oligo(dT) magnetic beads (Thermo Scientific) and fragmented using divalent cations and heat. Resulting fragments (160-220 bp) were converted to cDNA using random primers and SuperScriptll (Invitrogen) followed by second strand synthesis using DNA polymerase I and RNaseH. cDNA was end repaired, 5' phosphorylated and 3' adenylated according to NEB RNA library kit instructions. 3' single T-overhang Illumina multiplex specific adapters were ligated with T4 DNA ligase using a 10:1 molar ratio of adapter to cDNA insert. cDNA libraries were purified and size selected at 300 bp (E-Gel 2% SizeSelect gel, Invitrogen). Enrichment, adding of Illumina six base index and flow cell specific sequences was done by PCR using Phusion DNA polymerase and Illumina specific PCR primers. All cleanups up to this step were done with l,8x volume of Agencourt AMPure XP magnetic beads. All quality controls were done using Invitrogen's Qubit HS assay and fragment size was determined using Agilent's 2100 Bioanalyzer HS DNA assay. Barcoded RNA-Seq libraries were clustered and 50 bps were sequenced as described above.

NGS data analysis, gene expression: The output sequence reads from RNA samples were preprocessed according to the Illumina standard protocol, including filtering for low quality reads. Sequence reads were aligned to the mm9 (A. T. Chinwalla et al., Nature 420, 520 (2002)) or hgl8 (F. Collins, E. Lander, J. Rogers, R. Waterston, I. Conso, Nature 431, 931 (2004)) reference genomic sequence with bowtie (version 0.12.5) (B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Genome Biol 10, R25 (2009)). For genome alignments two mismatches were allowed and only the best alignment ("-v2 -best") reported, for transcript alignments default parameters were used. Reads not alignable to the genomic sequence were aligned to a database of all possible exon-exon junction sequences of the UCSC known genes (F. Hsu et al., Bioinformatics 22, 1036 (2006)). Expression values were determined by intersecting read coordinates with those of RefSeq transcripts, counting overlapping exon and junction reads, and normalizing to RPKM expression units (Reads which map per Kilobase of exon model per million mapped reads) (A. Mortazavi, B. A. Williams, K. McCue, L. Schaeffer, B. Wold, Nature methods 5, 621 (2008)).

.

NGS data analysis, somatic mutation discovery: Somatic mutations were identified as previously described (J. C. Castle et al., Exploiting the mutanome for tumor vaccination. Cancer Research 72, 1081 (2012)). Sequence reads aligned to the mm9 or hgl8 reference genome using bwa (default options, version 0.5.8c) (H. Li, R. Durbin, Bioinformatics 25, 1754 (2009)). Ambiguous reads mapping to multiple locations of the genome were removed. Mutations were identified using a consensus of two software programs: samtools (version 0.1.8) (H. Li, Bioinformatics 27, 1157 (201 1)) and SomaticSniper (A. McKenna et al., Genome Research 20, 1297 (2010)). For B16F10 and C57BL/6J, also GATK was included (A. McKenna et al., Genome Research 20, 1297 (2010)). Potential somatic variations identified in all respective replicates were assigned a "false discovery rate" (FDR) confidence value (M. Lower et al., PLoS computational biology 8, e10027 14 (2012)) (CT26 and MZ7-MEL only).

Mutation selection and validation

The criteria for selecting the 50 B16F10 mutations for immunogenicity testing were previously described (J. C. Castle et al., Exploiting the mutanome for tumor vaccination. Cancer Research 72, 1081 (2012)). These criteria for the mutations included: (i) presence in all three B16F10 replicates and absence from all C57BL/6 triplicates, (ii) occur in a RefSeq transcript, (iii) cause nonsynonymous change, (iv) occurrence in B16F10-expressed genes (median RPKM across replicates >10, exon expression > 0) and (v) for each mutation the M_{mut} score (see below) was required to be < 5. Of the 59 remaining mutations, the product of the quantile ranks of MHC class **I** score, MHC class **II** score and transcript expression was formed, and the first 50 mutations ($0.1 \leq M_{mut} \leq 3.9$) were selected for confirmation by PCR (see (J. C. Castle et al., Exploiting the mutanome for tumor vaccination. Cancer Research 72, 1081 (2012)) for further details). The criteria for the 96 CT26.WT mutations selected for immunogenicity testing were further refined and included the following: (i) presence in all CT26.WT three replicates and absence from all Balb/cJ three replicates, (ii) FDR ≤ 0.05, (iii) occur in a UCSC known gene transcript, (iv) cause nonsynonymous change, (v) not present in dbSNP database (vi) not in a genomic repeat region. From the remaining 493 mutations, eight 12-member groups were defined according to three features: M_{mut} score (lowest - [0.1,1.9] versus highest - [3.9-20.3]), compartment of the protein (extra-cellular, intra-cellular), and gene expression (below versus

above the median of 7.1 RPKM), selecting mutations according to a greedy algorithm, and adjusting thresholds accordingly. 94 of the resultant 96 mutations were confirmed by PCR followed by Sanger sequencing.

The criteria for selecting MZ7-ML mutations for analysis included: (i) presence in two MZ7-MEL replicates and absence from two autologous EBV-B replicates, followed by steps (ii) to (vi) describe above for CT26.WT. Applying steps (i)-(vi) reduced the initial list of ~8000 mutations to 743.

MHC binding prediction and calculation of the M_{mut} score

MHC binding predictions are performed using the IEDB analysis resource Consensus tool (http://tools.immuneepitope.org/analyze/html/mhc_binding.html) (Y. Kim et al., Nucleic Acids Research 40, W525 (2012)), which combines the best performing prediction methods based on benchmarking studies (H. H. Lin, S. Ray, S. Tongchusak, E. L. Reinherz, V. Brusic, BMC immunology 9, 8 (2008); B. Peters et al., PLoS computational biology 2, e65 (2006)) from ANN (C. Lundegaard et al., Nucleic Acids Research 36, W509 (2008); M. Nielsen et al., Protein Science 12, 1007 (2009)), SMM (B. Peters, A. Sette, BMC bioinformatics 6, 132 (2005)) and for some allele models also comblib (J. Sidney et al., Immunome Research 4, 2 (2008)). The consensus approach combines the prediction scores of all tools by generating a percentile rank, which reflects the binding prediction scores of the given peptide against peptide scores of five million random peptides from SWISSPROT.

For each mutation we calculated the predicted MHC consensus scores for all possible (i) sequence windows (where to position the mutation), (ii) epitope lengths and (iii) possible murine MHC class I alleles. The minimum of all MHC consensus scores was defined to be the M_{mut} score.

Calculation of log-odds matrices and the T score

Log-odds matrices can be estimated from sequence alignment comparisons of large protein databases. The early log-odds matrices were based on pairwise comparison of sequences (BLOSUM62 (S. Kreiter et al., Cancer Immunology, Immunotherapy 56, 1577 (2007))) and the

maximum parsimony (MP) estimation method (e.g., PAM250 (M. O. Dayhoff, R. M. Schwartz, B. C. Orcutt, A model for evolutionary change. MO Dayhoff, ed. Atlas of protein sequence and structure Vol.5, 345 (1978)), JTT250 (S. Q. Le, O. Gascuel, Molecular biology and evolution 25, 1307 (2008)), and the Gonnet matrix (C. C. Dang, V. Lefort, V. S. Le, Q. S. Le, O. Gascuel, Bioinformatics 27, 2758 (201 1))). More recently, maximum likelihood (ML) based methods were developed (e.g., VT160 (P. G. Higgs, T. K. Attwood, Bioinformatics and molecular evolution. (Wiley-Blackwell, 2009)), WAG (S. Whelan, N. Goldman, Molecular biology and evolution 18, 691 (2001)) and LG (V. Lennerz et al., Proceedings of the National Academy of Sciences of the United States of America 102, 16013 (2005))). Since ML is not limited to comparison of only closely related sequences, as is the case with MP based approaches, this estimation approach is expected to be the most accurate.

Calculation of log-odds matrix has been described in detail elsewhere (C. C. Dang, V. Lefort, V. S. Le, Q. S. Le, O. Gascuel, Bioinformatics 27, 2758 (201 1)). Briefly, the standard model for amino acid substitution assumes a Markovian, time-continuous, time-reversible model represented by a 20x20 rate matrix Q_{ij}, where q_{ij} $(i \neq j)$ is the number of substitutions from amino acid i to j per unit of time, and where diagonal elements are chosen to satisfy $Q_{ii} = -\sum_{j \neq i} Q_{ij}$. $\mathbf{Q}$ can be decomposed such that $Q_{ij} = S_{ij} \cdot \pi_j$ for $i \neq j$, where s_{ij} is a symmetric exchangeability matrix, and π_i is the probability to observe amino acid i (C. C. Dang, V. Lefort, V. S. Le, Q. S. Le, O. Gascuel, Bioinformatics 27, 2758 (201 1)). Finally, $\mathbf{Q}$ is normalized such that $1 = -\sum_i \pi_i Q_{ii}$, so that a time unit t=1.0 corresponds to 1.0 expected substitution per site, or one "accepted point mutation" per site, denoted by a PAM distance of 100 (M. O. Dayhoff, R. M. Schwartz, B. C. Orcutt, A model for evolutionary change. MO Dayhoff, ed. Atlas of protein sequence and structure Vol.5, 345 (1978); S. Q. Le, O. Gascuel, Molecular biology and evolution 25, 1307 (2008); C. C. Dang, V. Lefort, V. S. Le, Q. S. Le, O. Gascuel, Bioinformatics 27, 2758 (201 1)). The probability for amino acid i to be replaced by amino acid j after time t, $?r(i \rightarrow j \backslash t) = P_{ij}(t)$, is given by the 20x20 probability matrix $\mathbf{P}(t) = e^{t\mathbf{Q}}$ (with notation denoting matrix exponentiation). The log-odds matrix calculated for time t is given by the log-

odds 20x20 matrix $T_{i,j}=10\log_{10}\left(\frac{\pi_i P_{ij}(t)}{\pi_i \pi_j}\right)$ (M. O. Dayhoff, R. M. Schwartz, B. C. Orcutt, A model for evolutionary change. MO Dayhoff, ed. Atlas of protein sequence and structure Vol.5, 345 (1978, 1978)). A time-reversible mean that $\pi,P_{ij}(i) = \pi_j P_{ji}(t)$, and therefore $T_{t.j}$ is symmetric (P. G. Higgs, T. K. Attwood, Bioinformatics and molecular evolution. (Wiley-Blackwell, 2009)).

The T score for the substitution $i \leftrightarrow j$ is defined here as $T_{i.j}$, and depends on the evolutionary model and the time *t*. We explored various models and PAM distances for the T score, including PAM, BLOSUM62, JTT, VT160, Gonnet, WAG, WAG*, and LG (see references above). The figures in this report were generated using a T score based on the WAG model and a PAM distance of 250. Such a large PAM distance means that there is substantial chance for the amino acid to change (P. G. Higgs, T. K. Attwood, Bioinformatics and molecular evolution. (Wiley-Blackwell, 2009)), and is useful in detecting distant relationships between sequences where residues may not be identical but the physico-chemical properties of the amino acids are conserved (M. O. Dayhoff, R. M. Schwartz, B. C. Orcutt, A model for evolutionary change. MO Dayhoff, ed. Atlas of protein sequence and structure Vol.5, 345 (1978); P. G. Higgs, T. K. Attwood, Bioinformatics and molecular evolution. (Wiley-Blackwell, 2009)).

Using a *t*-distribution test statistic we compared the mean T scores of immunogenic versus non-immunogenic epitopes from Table 3 for the WAG matrix using various PAM scores (1, 10, 25, 50, 100, 150, 200, and 250). Analysis of the test statistic showed that the P value decreased monotonically with PAM distance, implying that a PAM distance of 250 was the optimal solution, as would be anticipated (data not shown). The classification into H_A and $H_„$ was the same for all matrices except for the PAM matrix, which is the least accurate of all evolutionary models. Of all evolutionary models, the WAG250 model resulted in the maximum separation between H_A and H_N epitopes in Table 3, measuring separation with the test statistic: [max T score(*H* i)-min T score($H_„$)]/a(T score (H_A), T **score**($H_„$)) (data not shown). The same test statistic was also maximal for a PAM distance 250 compared to smaller distance.

Published CD8+ epitopes

CD8+ epitopes with single mutated amino acids were collected from the list of tumor antigens resulting from mutations published by the Cancer Immunity Journal (P. Van der Bruggen, V. Stroobant, N. Vigneron, B. Van den Eynde. (Cancer Immun, http://www.cancerimmunity.org/peptide/, 2013) (http://cancerimtnunity.org/peptide/mutations/). HLA alleles were taken either from the published table or from the original paper if the latter was more precise. References listed in Table 4 are the following: (1) Lennerz ***et al. PNAS*** 102 (44), pp. 16013-16018 (2005); (2) Karanikas ***et al. Cancer Res*** 61 (9), pp. 3718-3724 (2001); (3) Sensi ***et al. Cancer Res*** 65 (2), pp. 632-640 (2005); (4) Linard ***et al. J. Immunol*** 168 (9), pp. 4802^1808 (2002); (5) Zorn ***et al. Eur. J. Immunol*** 29 (2), pp. 592-601 (1999); (6) Graf ***et al. Blood*** 109 (7), pp. 2985-2988 (2007); (7) Robbins ***et al. J. Exp. Med*** 183 (3), pp. 1185-1192 (1996); (8) Vigneron ***et al. Cancer Immun*** 2, pp. 9 (2002); (9) Echchakir ***et al. Cancer Res*** 61 (10), pp. 4078-4083 (2001); (10) Hogan ***et al. Cancer Res*** 58 (22), pp. 5144-5150 (1998); (11) Ito ***et al. Int. J. Cancer*** 120 (12), pp. 2618-2624 (2007); (12) Wolfel ***et al. Science*** 269 (5228), pp. 1281-1284 (1995); (13) Gjertsen ***et al. Int. J. Cancer*** 72 (5), pp. 784-790 (1997).

Example 3: Example of a scheme for weighing mutation scores to improve prioritization of immunogenic mutations

RNA that is injected into the cell, once translated and cleaved into short peptides, can be presented on different HLA types within the cell. Therefore it stands to reason that the more HLA types that are predicted to have a low MHC consensus (or similar) score, the more likely a given mutation will be immunogenic since it can potentially be displayed on more than one HLA type in parallel. Thus, weighing mutations by the number of HLA types for which the mutation is classified as $\boldsymbol{H_A}$ and/or $\boldsymbol{H_B \cup H_C}$ or even weighing each mutation simply by the number of HLA types that have a low $\boldsymbol{M_{mut}}$ score may improve immunogenicity ranking. In the most general solution, when we inject a 27mer RNA or peptide into the cell, there is not just the freedom to select the HLA type, but also the length of the peptide and the position of the mutation within this peptide. Therefore, one can scan all possible HLA types, all possible window lengths and all possible positions for the mutation within the window and calculate the number of solutions (per

given mutation) that are classified as H_A and/or H_BUHc (Fig. 12). This may be an important weighing factor for mutation prioritization to select the most efficacious epitopes for vaccination. An example of a scatter plot of all these solutions as a function of M_{mM} and $AM = M_{mut} - M_{wt}$ is shown in Fig. 13.

CLAIMS

1. A method for predicting immunogenic amino acid modifications, the method comprising the steps:

a) ascertaining a score for binding of a modified peptide to one or more MHC molecules, and

b) ascertaining a score for binding of the non-modified peptide to one or more MHC molecules,

and/or

c) ascertaining a score for binding of the modified peptide when present in a MHC-peptide complex to one or more T cell receptors.

2. The method of claim 1 wherein the modified peptide comprises a fragment of a modified protein, said fragment comprising the modification(s) present in the protein.

3. The method of claim 1 or 2 wherein the non-modified peptide has the germline amino acid at the position(s) corresponding to the position(s) of the modification(s) in the modified peptide.

4. The method of any one of claims 1 to 3 wherein the non-modified peptide and modified peptide are identical but for the modification(s).

5. The method of any one of claims 1 to 4 wherein the non-modified peptide and modified peptide are 8 to 15, preferably 8 to 12 amino acids in length.

6. The method of any one of claims 1 to 5 wherein the one or more MHC molecules comprise different MHC molecule types, in particular different MHC alleles.

7. The method of any one of claims 1 to 6 wherein the one or more MHC molecules are MHC class I molecules and/or MHC class II molecules.

8. The method of any one of claims 1 to 7 wherein the score for binding to one or more MHC molecules is ascertained by a process comprising a sequence comparison with a database of MHC-binding motifs.

9. The method of any one of claims 1 to 8 wherein step a) comprises ascertaining whether said score satisfies a pre-determined threshold for binding to one or more MHC molecules.

10. The method of any one of claims 1 to 9 wherein step b) comprises ascertaining whether said score satisfies a pre-determined threshold for binding to one or more MHC molecules.

11. The method of any one of claims 1 to 10 wherein the threshold applied in step a) is different to the threshold applied in step b).

12. The method of any one of claims 1 to 11 wherein the pre-determined threshold for binding to one or more MHC molecules reflects a probability for binding to one or more MHC molecules.

13. The method of any one of claims 1 to 12 wherein step c) comprises ascertaining a score for the chemical and physical similarities between the non-modified and modified amino acids.

14. The method of any one of claims 1 to 13 wherein step c) comprises ascertaining whether said score satisfies a pre-determined threshold for the chemical and physical similarities between amino acids.

15. The method of any one of claims 1 to 14 wherein the score for the chemical and physical similarities is ascertained on the basis of the probability of amino acids being interchanged in nature.

16. The method of any one of claims 1 to 15 wherein the more frequently amino acids are interchanged in nature the more similar the amino acids are considered.

17. The method of any one of claims 1 to 16 wherein the chemical and physical similarities are determined using evolutionary based log-odds matrices.

18. The method of any one of claims 1 to 17 wherein, if the non-modified peptide has a score for binding to one or more MHC molecules satisfying a threshold indicating binding to one or more MHC molecules and the modified peptide has a score for binding to one or more MHC molecules satisfying a threshold indicating binding to one or more MHC molecules, the modification or modified peptide is predicted as immunogenic if the non-modified and modified amino acids have a score for the chemical and physical similarities satisfying a threshold indicating chemical and physical dissimilarity.

19. The method of any one of claims 1 to 18 wherein if the non-modified peptide binds to one or more MHC molecules or has a probability for binding to one or more MHC molecules and the modified peptide binds to one or more MHC molecules or has a probability for binding to one or more MHC molecules, the modification or modified peptide is predicted as immunogenic if the non-modified and modified amino acids are chemically and physically dissimilar or have a probability of being chemically and physically dissimilar.

20. The method of claim 18 or 19 wherein the modification is not in an anchor position for binding to one or more MHC molecules.

21. The method of any one of claims 1 to 17 wherein, if the non-modified peptide has a score for binding to one or more MHC molecules satisfying a threshold indicating no binding to one or more MHC molecules and the modified peptide has a score for binding to one or more MHC molecules satisfying a threshold indicating binding to one or more MHC molecules, the modification or modified peptide is predicted as immunogenic.

22. The method of any one of claims 1 to 17 and 21 wherein, if the non-modified peptide does not bind to one or more MHC molecules or has a probability for not binding to one or more MHC molecules and the modified peptide binds to one or more MHC molecules or has a probability for binding to one or more MHC molecules, the modification or modified peptide is predicted as immunogenic.

23. The method of claim 21 or 22 wherein the modification is in an anchor position for binding to one or more MHC molecules.

24. The method of any one of claims 1 to 23 which comprises performing step a) on two or more different modified peptides, said two or more different modified peptides comprising the same modification(s).

25. The method of claim 24 wherein the two or more different modified peptides comprising the same modification(s) comprise different fragments of a modified protein, said different fragments comprising the same modification(s) present in the protein.

26. The method of claim 24 or 25 wherein the two or more different modified peptides comprising the same modification(s) comprise all potential MHC binding fragments of a modified protein, said fragments comprising the same modification(s) present in the protein.

27. The method of any one of claims 24 to 26 further comprising selecting (the) modified peptide(s) from the two or more different modified peptides comprising the same modification(s) having a probability or having the highest probability for binding to one or more MHC molecules.

28. The method of any one of claims 24 to 27 wherein the two or more different modified peptides comprising the same modification(s) differ in length and/or position of the modification(s).

29. The method of any one of claims 1 to 28 which comprises performing step a) and optionally one or both of steps b) and c) on two or more different modified peptides.

30. The method of claim 29 wherein said two or more different modified peptides comprise the same modification(s) and/or comprise different modifications.

31. The method of claim 30 wherein the different modifications are present in the same and/or in different proteins.

32. The method of any one of claims 24 to 31 which comprises comparing the scores of two or more of said different modified peptides.

33. The method of claim 32 wherein a score for binding of the modified peptide to one or more MHC molecules is weighted higher than a score for binding of the modified peptide when present in a MHC-peptide complex to one or more T cell receptors, preferably a score for the chemical and physical similarities between the non-modified and modified amino acids and a score for binding of the modified peptide when present in a MHC-peptide complex to one or more T cell receptors, preferably a score for the chemical and physical similarities between the non-modified and modified amino acids is weighted higher than a score for binding of the non-modified peptide to one or more MHC molecules.

34. The method of any one of claims 1 to 33 further comprising identifying non-synonymous mutations in one or more protein-coding regions.

35. The method of any one of claims 1 to 34 wherein modifications are identified by partially or completely sequencing the genome or transcriptome of one or more cells such as one or more cancer cells and optionally one or more non-cancerous cells and identifying mutations in one or more protein-coding regions.

36. The method of claim 34 or 35 wherein said mutations are somatic mutations.

37. The method of claim 34 to 36 wherein said mutations are cancer mutations.

38. The method of any one of claims 1 to 37 which is used in the manufacture of a vaccine.

39. The method of claim 38 wherein the vaccine is derived from (a) modification(s) or (a) modified peptide(s) predicted as immunogenic by said method.

40. A method for providing a vaccine comprising the step:
identifying (a) modification(s) or (a) modified peptide(s) predicted as immunogenic by the method of any one of claims 1 to 37.

41. The method of claim 40 further comprising the step:

providing a vaccine comprising a peptide or polypeptide comprising the modification(s) or modified peptide(s) predicted as immunogenic, or a nucleic acid encoding the peptide or polypeptide.

42. A vaccine produced according to the method of any one of claims 38 to 41.

Fig. 1

Sequence-based: vs. **Structure-based:**

- Position specific scoring matrices
  - SYFPEITHI[1]
  - SMM-align[2]
  - BIMAS-HLA[3]
  - ARB matrix[4]
  - ProPred-I[5]
  - MAPPP[6]
- Artificial intelligence
  - Artificial neural networks
    - ANNPred[7]
    - ANN_nielsen[8]
  - Hidden Markov Models
    - Gibbs-based[9]
    - S-HMM[10]
  - Support Vector Machines
    - SVMHC[11]
    - SVR-based[12]
- Hybrids
  - NetMHC[13] (ANN + weighted PSSM)
  - ComPred[14] (ANN+PSSM)
- Integrative approaches
  - NetCTL[15] (modelling proteasome, TAP, MHC binding)

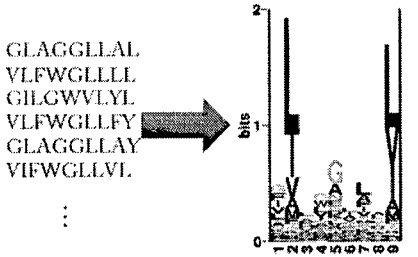


Fig. Sequence logo for aligned peptides (PSSM)[16]

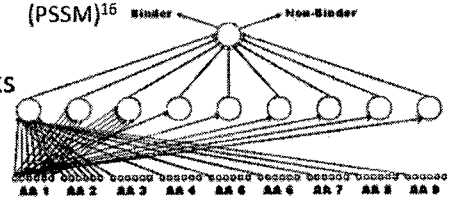


Fig. A three-layer ANN for predicting class I binding peptides[17]

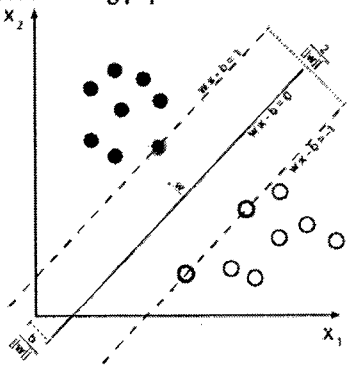


Fig. Maximum-margin hyperplane and margins for an SVM trained with samples from 2 classes[18]

Requires MHC crystal structures

- Predep[19] ("Threading")
- pDock[20] ("docking")

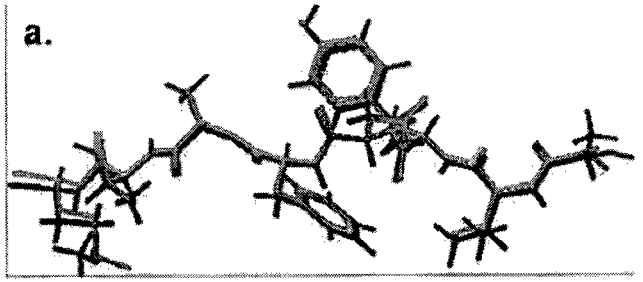


Fig. Comparison of the lowest energy predicted conformation (green)and the experimental structure of a peptide in MHC I[20]

Do

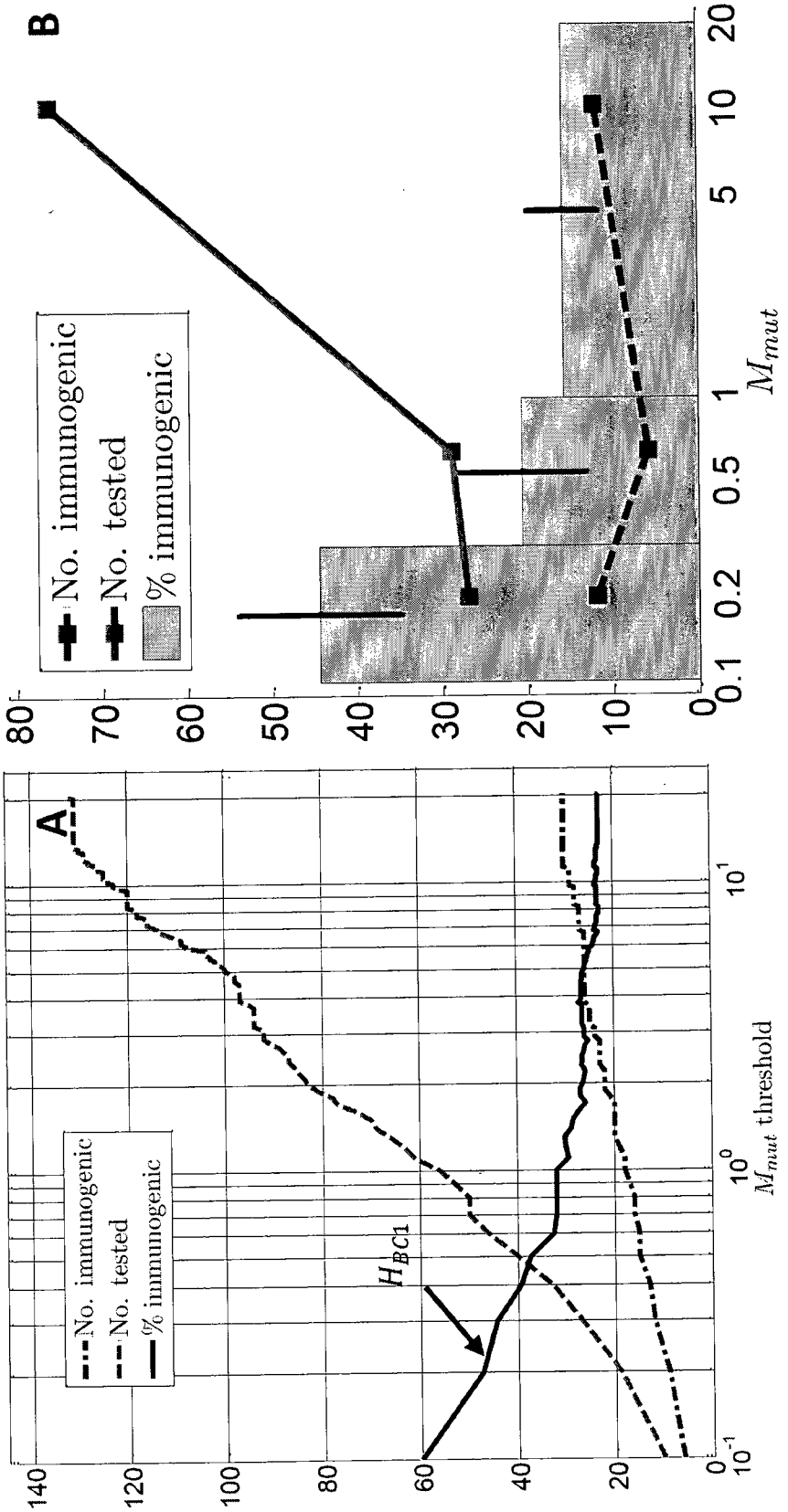
Do not

rely on binding data for training

[1]Rammensee HJ (1999) Immunogenetics, [2]Sette A (2005) Bioinformatics, [3]Parker KC (1994) J Immunol, [4]Sette A (2005) Immunogenetics, [5]Sing H (2003) Bioinformatics, [6]Hakenberg J (2003) Appl. Bioinformatics, [7]Bhatia V (2004) Infect. Immun., [8]Nielsen M (2003) Protein Sci, [9]Nielsen M. (2004) Bioinformatics, [10]Brusic V (2002) J Biosci Bioeng., [11]Dönnes P (2006) NAR, [12]Flower et al. (2006) BMC Bioinform, [13]Lundegaard C (2008) Bioinformatics, [14]Brusic V (2004) Methods, [15]Larsen MV (2005) Eur J Immunol., [16]Lund, Nielsen: Immunological Bioinformatics, 2005. MIT Press. Page 117, [17]Brusic V (1994) Complex Systems, [18]http://en.wikipedia.org/wiki/Support_vector_machine, [19]Altuvia Y (2004) Methods, [20]Khan JM (2010), Imm Research

WO 2014/180490 PCT/EP2013/001400

Fig. 2


A
No. immunogenic
No. tested
% immunogenic
H_{BC1}
M_{mut} threshold
B
No. immunogenic
No. tested
% immunogenic
M_{mut}

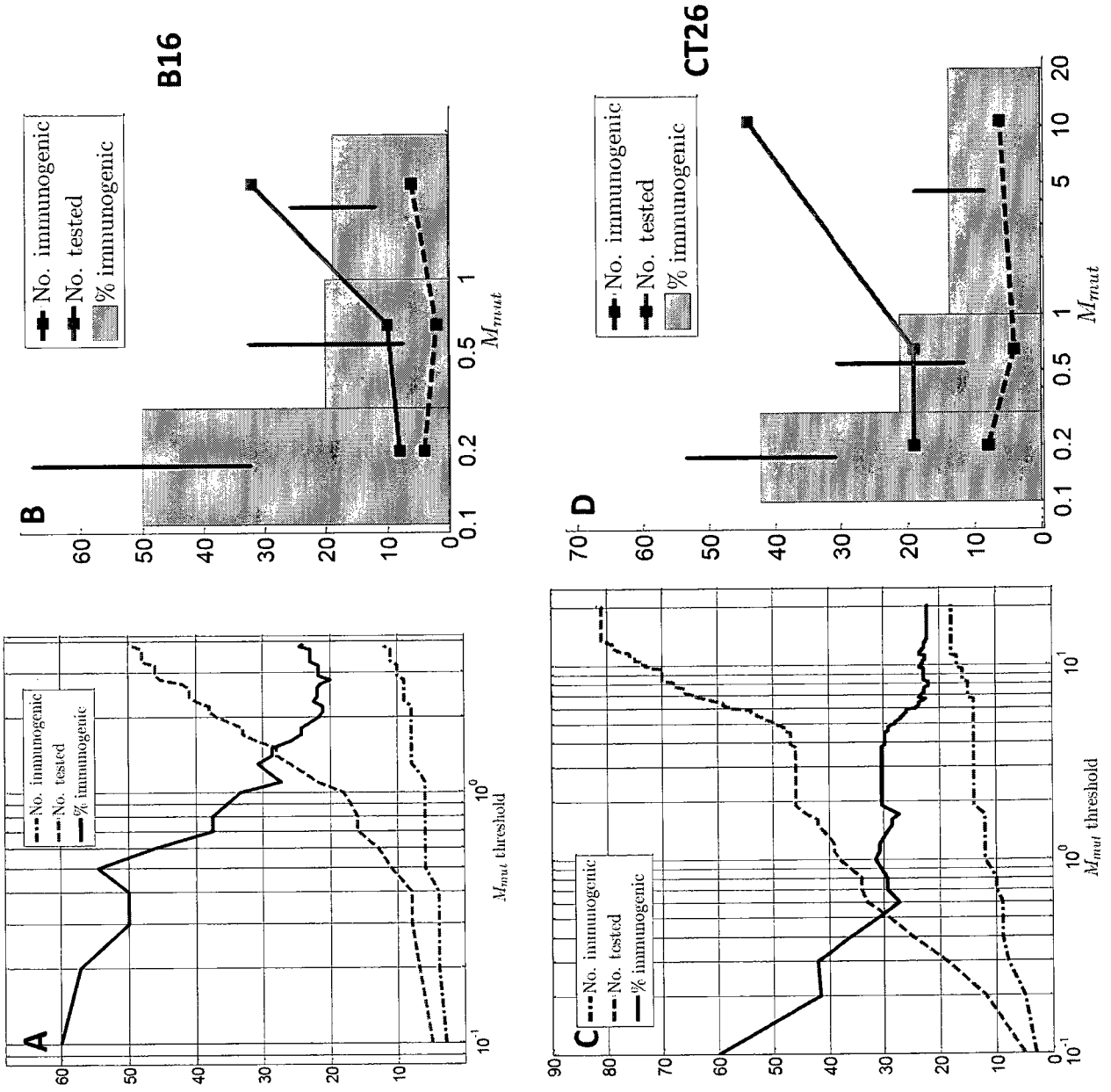

A
B
B16
C
D
CT26
No. immunogenic
No. tested
% immunogenic
M_{mut} threshold
M_{mut}

Fig. 3

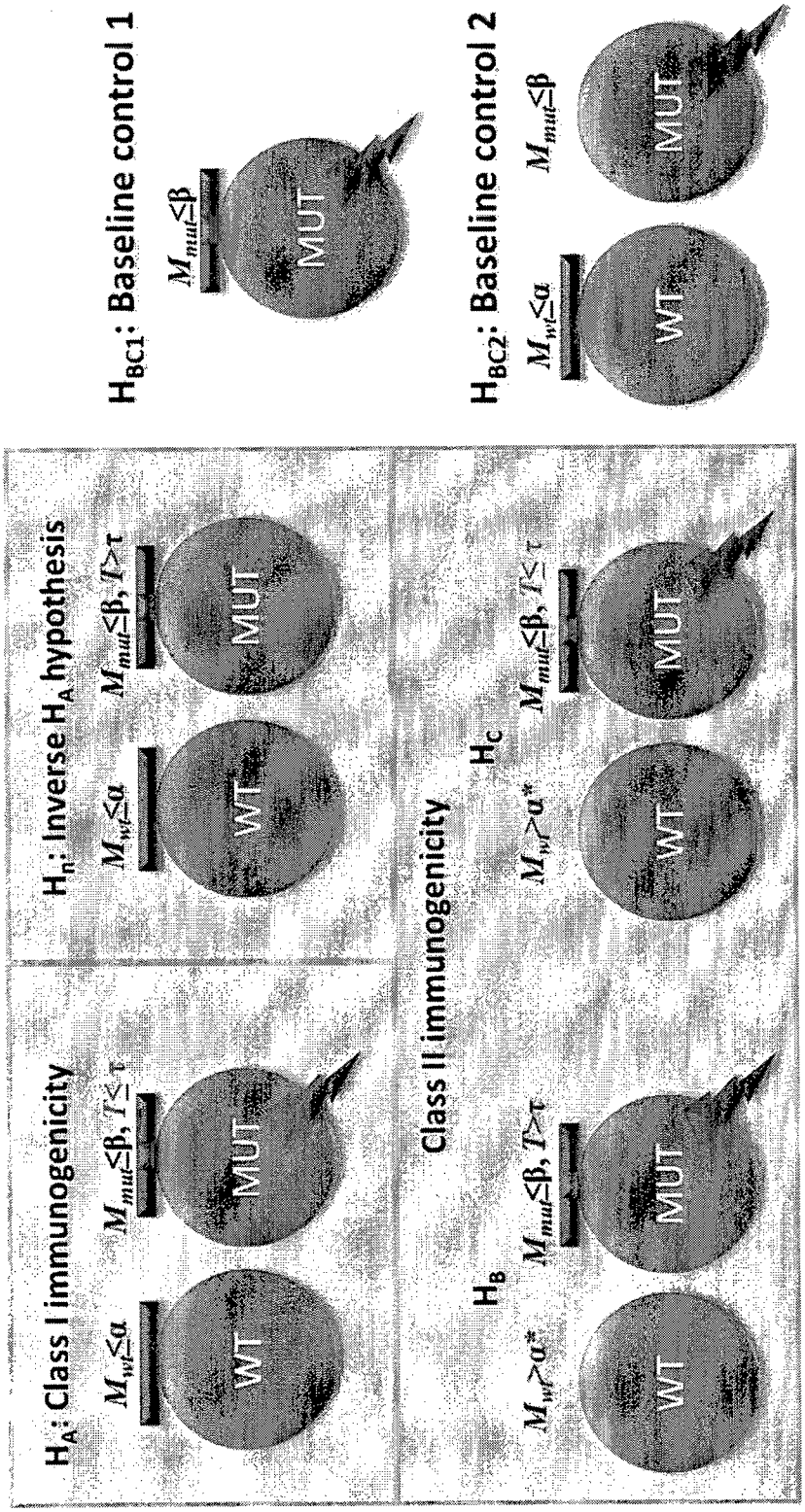

H_BC1: Baseline control 1
M_mut≤β
MUT
H_BC2: Baseline control 2
M_wt≤α
WT
M_mut≤β
MUT
H_n: Inverse H_A hypothesis
M_wt≤α
WT
M_mut≤β, T>τ
MUT
H_A: Class I immunogenicity
M_wt≤α
WT
M_mut≤β, T≤τ
MUT
Class II immunogenicity
H_C
M_wt>α*
WT
M_mut≤β, T≤τ
MUT
H_B
M_wt>α*
WT
M_mut≤β, T>τ
MUT

Fig. 4

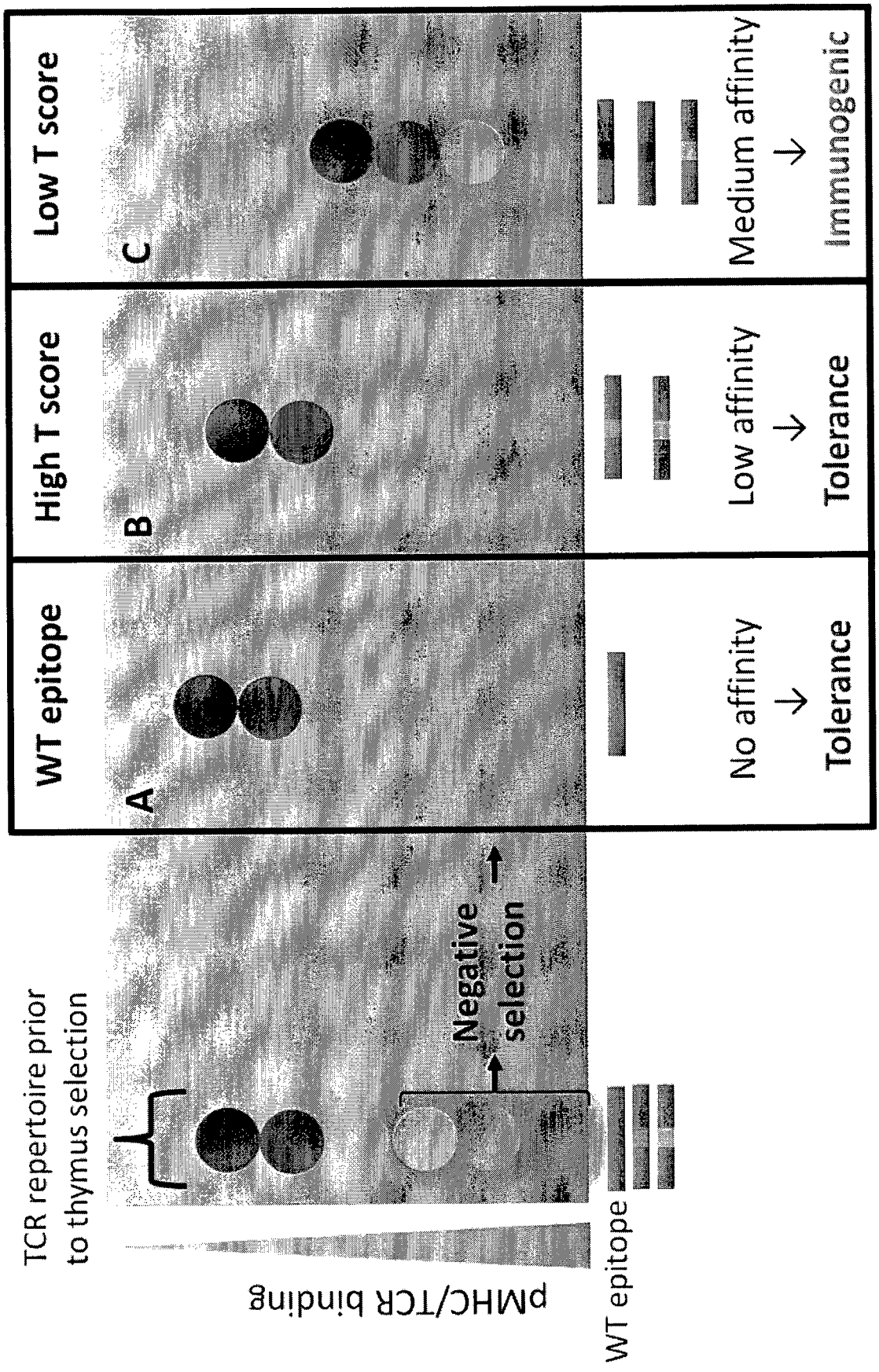

WT epitope
High T score
Low T score
A
B
C
No affinity
Low affinity
Medium affinity
Tolerance
Tolerance
Immunogenic
TCR repertoire prior to thymus selection
Negative selection
pMHC/TCR binding
WT epitope

Fig. 5

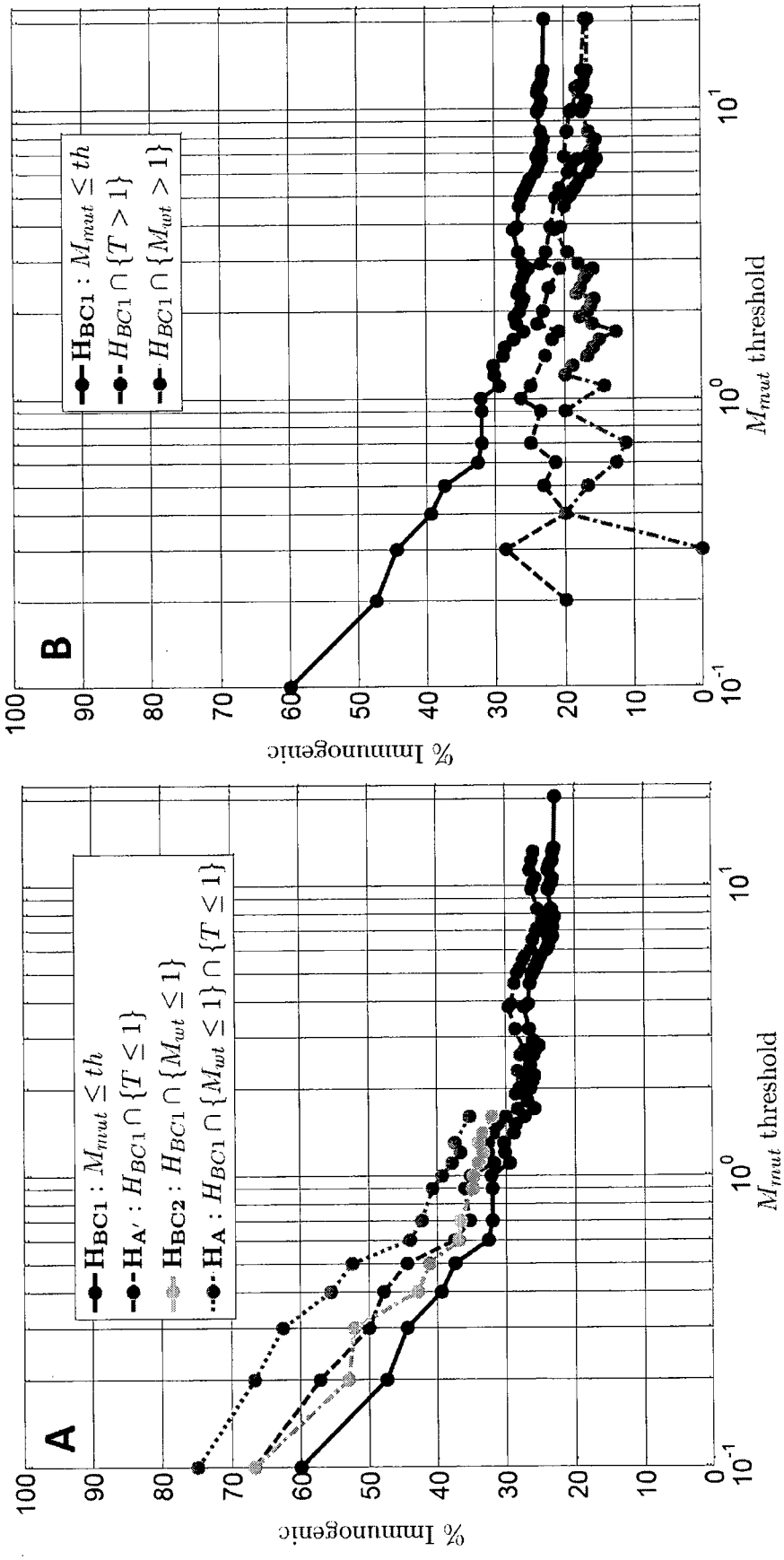

A
B
% Immunogenic
M_{mut} threshold
$\mathbf{H_{BC1}}$: $M_{mut} \leq th$
$\mathbf{H_{A'}}$: $H_{BC1} \cap \{T \leq 1\}$
$\mathbf{H_{BC2}}$: $H_{BC1} \cap \{M_{wt} \leq 1\}$
$\mathbf{H_A}$: $H_{BC1} \cap \{M_{wt} \leq 1\} \cap \{T \leq 1\}$
$\mathbf{H_{BC1}}$: $M_{mut} \leq th$
$H_{BC1} \cap \{T > 1\}$
$H_{BC1} \cap \{M_{wt} > 1\}$

Fig. 6

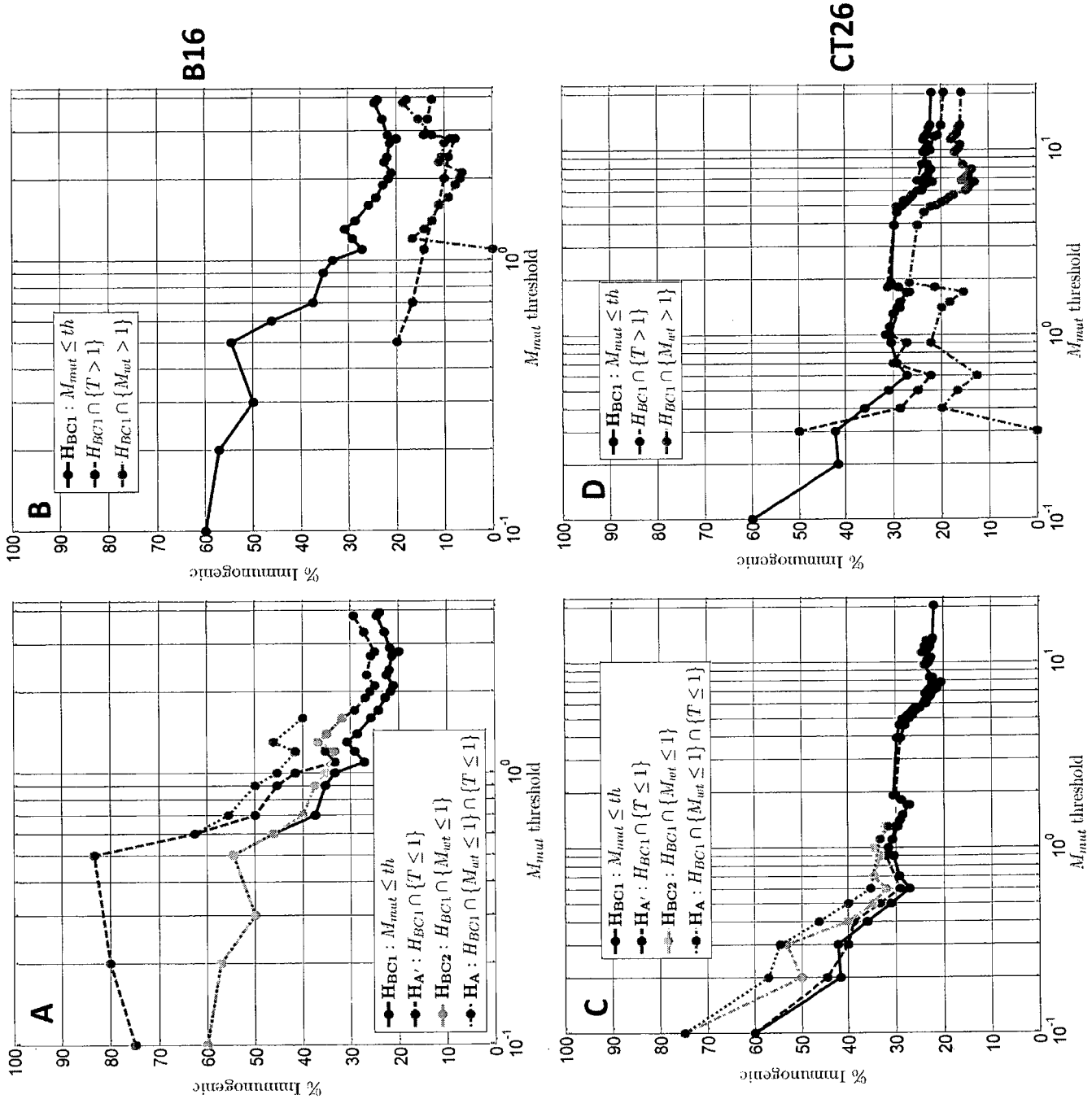

B16
CT26
A
B
C
D
% Immunogenic
M_{mut} threshold
$H_{BC1} : M_{mut} \leq th$
$H_{A'} : H_{BC1} \cap \{T \leq 1\}$
$H_{BC2} : H_{BC1} \cap \{M_{wt} \leq 1\}$
$H_A : H_{BC1} \cap \{M_{wt} \leq 1\} \cap \{T \leq 1\}$
$H_{BC1} \cap \{T > 1\}$
$H_{BC1} \cap \{M_{wt} > 1\}$

Fig. 7

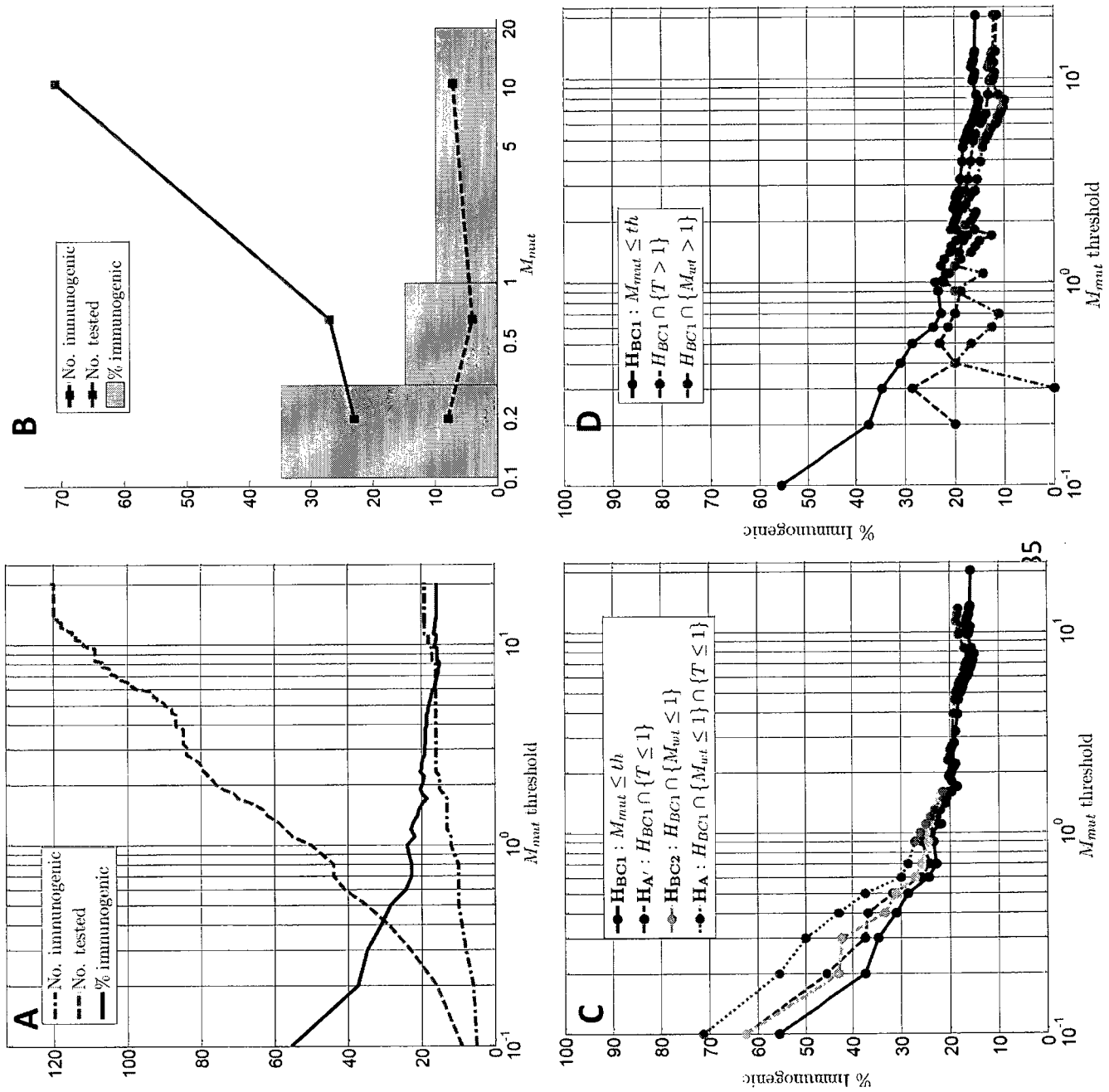
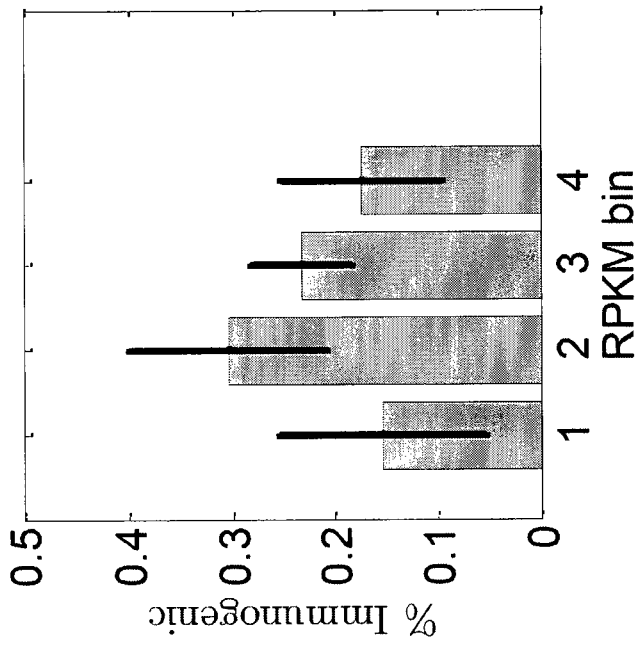
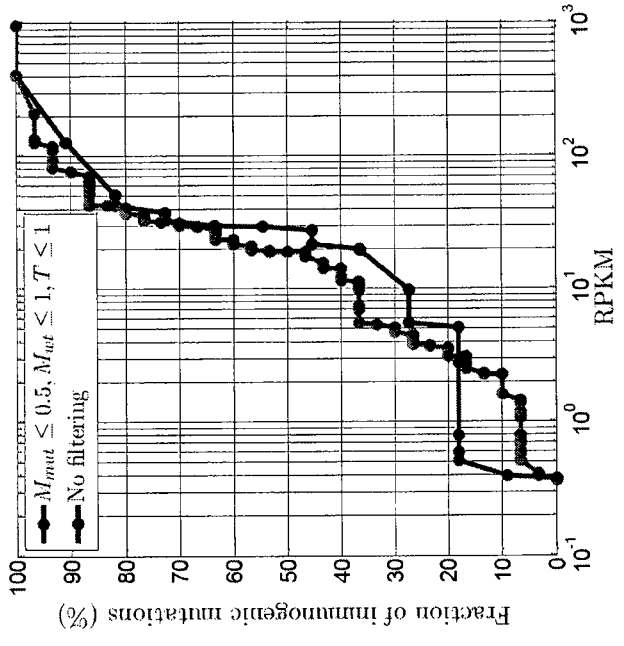

A
No. immunogenic
No. tested
% immunogenic
M_{mut} threshold
B
No. immunogenic
No. tested
% immunogenic
M_{mut}
C
$H_{BC1} : M_{mut} \leq th$
$H_{A'} : H_{BC1} \cap \{T \leq 1\}$
$H_{BC2} : H_{BC1} \cap \{M_{wt} \leq 1\}$
$H_A : H_{BC1} \cap \{M_{wt} \leq 1\} \cap \{T \leq 1\}$
% Immunogenic
M_{mut} threshold
D
$H_{BC1} : M_{mut} \leq th$
$H_{BC1} \cap \{T > 1\}$
$H_{BC1} \cap \{M_{wt} > 1\}$
% Immunogenic
M_{mut} threshold

Fig. 8


% Immunogenic
0.5
0.4
0.3
0.2
0.1
0
1
2
3
4
RPKM bin


Fraction of immunogenic mutations (%)
100
90
80
70
60
50
40
30
20
10
0
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
RPKM
$M_{mut} \leq 0.5, M_{wt} \leq 1, T \leq 1$
No filtering

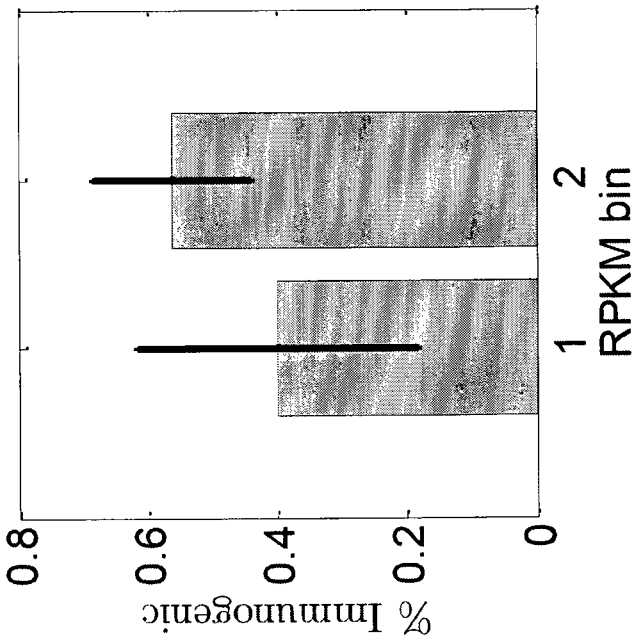

% Immunogenic
0.8
0.6
0.4
0.2
0
1
2
RPKM bin

Fig. 9

Fig. 10

Sample[1]	Class	Gene	HLA allele	Anchor positions[2]	Epitope (WT)[3]	Epitope (MUT)[3]	M_{mut}	M_{wt}	Achnor position mutation	Immunogenic
MZ7-MEL	class II (H_BUH_C)	SNRP116	HLA-A*03	1 **2** 3 4 5 6 7 8 **9**	KI**L**DAV**V**AQE	KI**L**DAV**V**AQ**K**	0.15	10.20	yes	yes
MZ7-MEL	class II (H_BUH_C)	ZNF335	HLA-B*38:01	1 2 3 4 5 6 7 8 **9**	Q**H**SAAPGPP	Q**H**SAAPGP**L**	0.20	18.35	yes	unknown
Zorn *et al.*	class II (H_BUH_C)	MYO1B	HLA-A*03	1 **2** 3 4 5 6 7 8 **9**	EINKNPKY**K**	KINKNPKY**K**	0.25	2.25	no	yes
Graf *et al.*	class II (H_BUH_C)	FLT3	HLA-A*01	1 2 **3** 4 5 6 7 8 **9**	YVDFREYEY**D**	YVDFREYEY**Y**	0.20	10.95	yes	yes
Robbins *et al.*	class II (H_BUH_C)	CTNNB1	HLA-A24	1 **2** 3 4 5 6 7 8 **9**	S**Y**LDSGIHS	S**Y**LDSGIH**F**	0.38	10.58	yes	yes

[1] See Table 3

[2] Anchors (bold) and auxiliary anchors (underline) positions are indicated

[3] Conserved anchor (bold) and auxiliary anchor (underline) residues for the given HLA type and epitope are printed in large fonts. Mutated residue is highlighted in red.

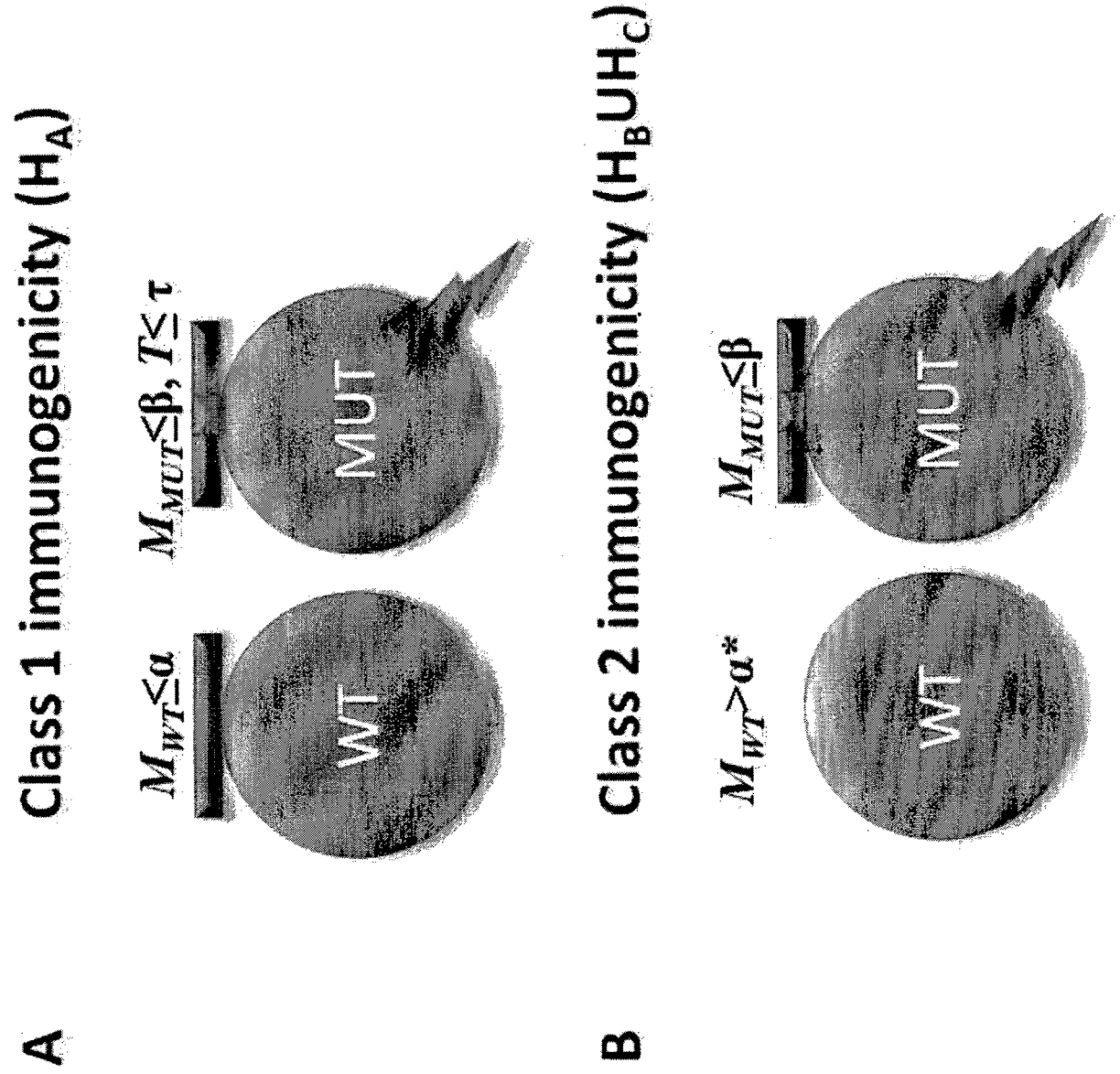

A
Class 1 immunogenicity (H_A)
$M_{WT} \leq \alpha$
WT
$M_{MUT} \leq \beta$, $T \leq \tau$
MUT
B
Class 2 immunogenicity ($H_B UH_C$)
$M_{WT} > \alpha^*$
WT
$M_{MUT} \leq \beta$
MUT

Fig. 11

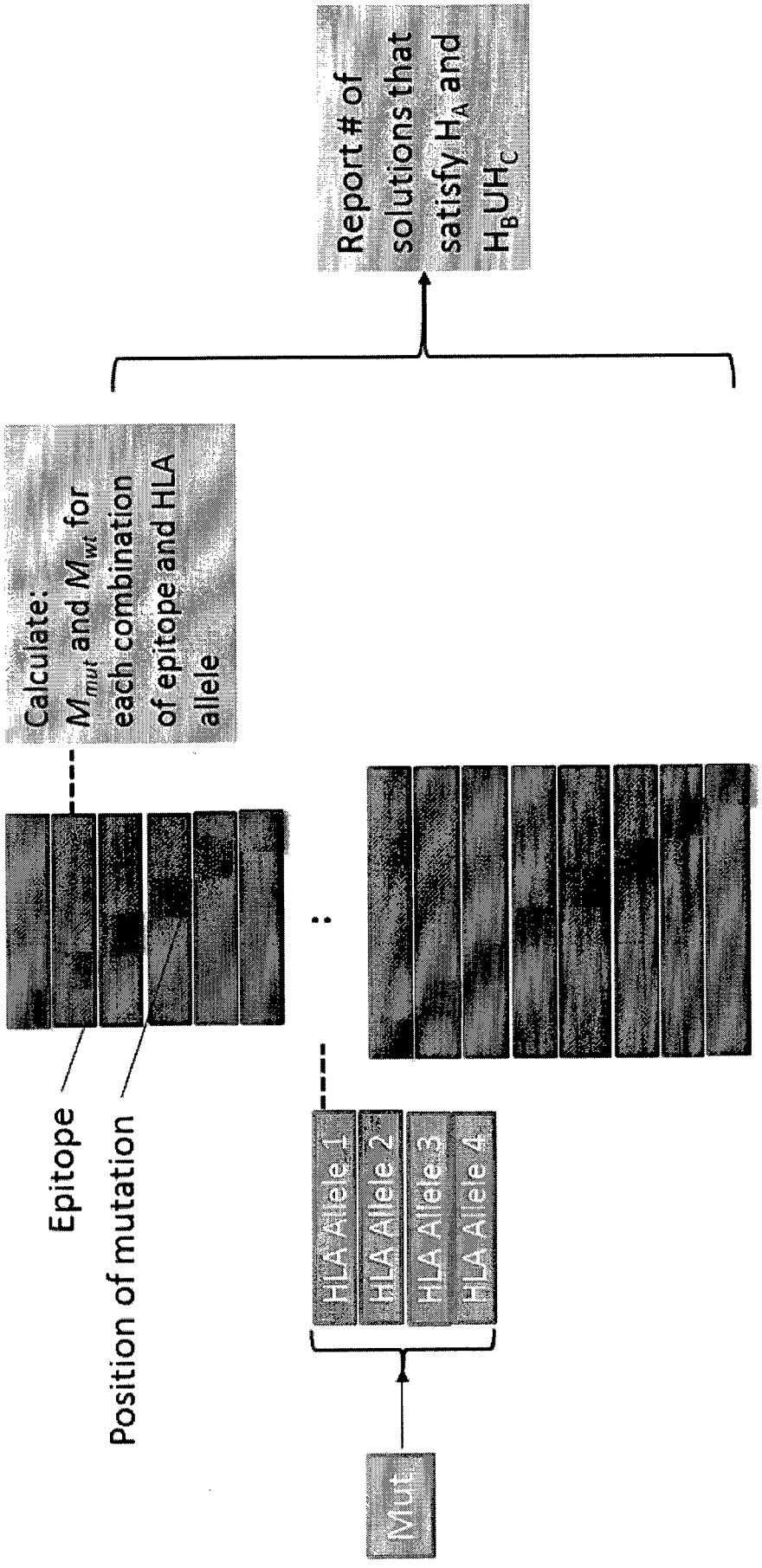

Report # of solutions that satisfy H_A and $H_B UH_C$
Calculate: M_{mut} and M_{wt} for each combination of epitope and HLA allele
Epitope
Position of mutation
HLA Allele 1
HLA Allele 2
HLA Allele 3
HLA Allele 4
Mut.

Fig. 12

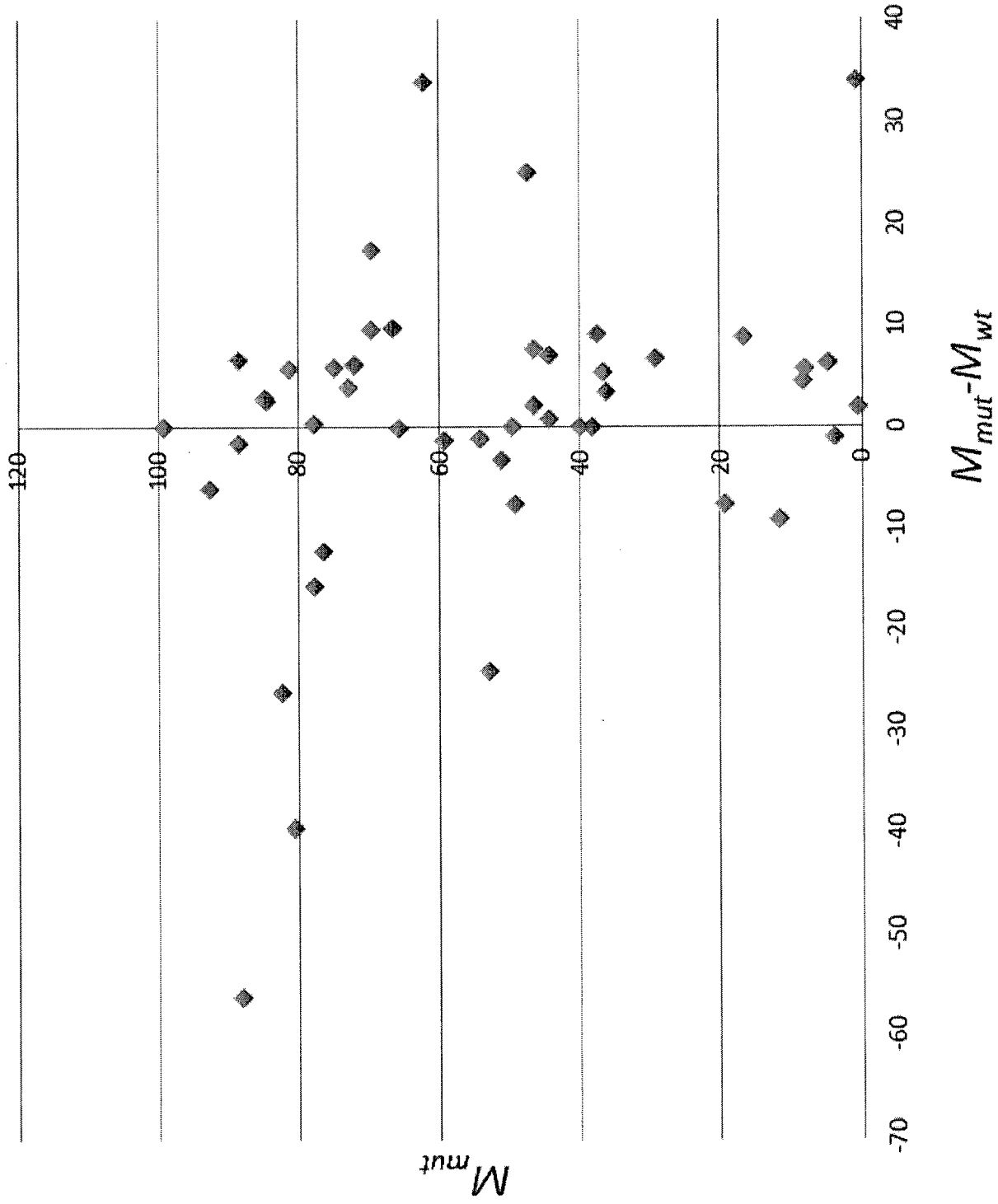


Fig. 13

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2013/001400

A. CLASSIFICATION OF SUBJECT MATTER
INV. A61K39/00 G06F19/18
ADD.

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
A61K G06F

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

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	MICHIELIN O ET AL: "Binding Free Energy Differences in a TCR-Peptide-MHC Complex Induced by a Peptide Mutation: A Simulation Analysis", JOURNAL OF MOLECULAR BIOLOGY, ACADEMIC PRESS, UNITED KINGDOM, vol. 324, no. 3, 29 November 2002 (2002-11-29), pages 547-569, XP027251246, ISSN: 0022-2836, DOI: 10.1016/S0022-2836(02)00880-X [retrieved on 2002-11-29] whole doc, in particular scheme 1, eq. (2), p. 550, left col., fig. 5	1-41

-/- .

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" documentwhich may throw doubts on priority claim(s) orwhich is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
3 December 2013	02/01/2014

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Ludemann , Susanna

2

Form PCT/ISA/210 (second sheet) (April 2005)

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WAN SHUNZHOU ET AL: "Molecular basis of peptide recognition by the TCR: Affinity differences calculated using large scale computing", JOURNAL OF IMMUNOLOGY, vol. 175, no. 3, August 2005 (2005-08), pages 1715-1723, XP002717288, ISSN: 0022-1767 whole doc, in particular fig.2 table II	1-41
A	ESTHER M. LAFUENTE ET AL: "Prediction of MHC-Peptide Binding: A Systematic and Comprehensive Overview", CURRENT PHARMACEUTICAL DESIGN, vol. 15, no. 28, 1 October 2009 (2009-10-01), pages 3209-3220, XP55006121, ISSN: 1381-6128, DOI: 10.2174/138161209789105162 the whole document	1-41
X	WO 2012/159754 A2 (BIONTECH AG [DE]; TRON TRANSLATIONALE ONKOLOGIE AN DER UNIVERSITAETSME) 29 November 2012 (2012-11-29) whole doc, in particular p. 103	42

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2013/001400

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2012159754 A2	29-11-2012	AU 2012261237 Al	14-11-2013
		WO 2012159754 A2	29-11-2012