Innovation, Sciences et
Développement économique Canada
Office de la Propriété Intellectuelle du Canada

Innovation, Science and
Economic Development Canada
Canadian Intellectual Property Office

CA 2836494 C 2023/01/03

(11)(21) **2 836 494**

(12) **BREVET CANADIEN**
CANADIAN PATENT

(13) **C**

(86) **Date de dépôt PCT/PCT Filing Date:** 2012/05/23

(87) **Date publication PCT/PCT Publication Date:** 2012/11/29

(45) **Date de délivrance/Issue Date:** 2023/01/03

(85) **Entrée phase nationale/National Entry:** 2013/11/18

(86) **N° demande PCT/PCT Application No.:** EP 2012/002209

(87) **N° publication PCT/PCT Publication No.:** 2012/159754

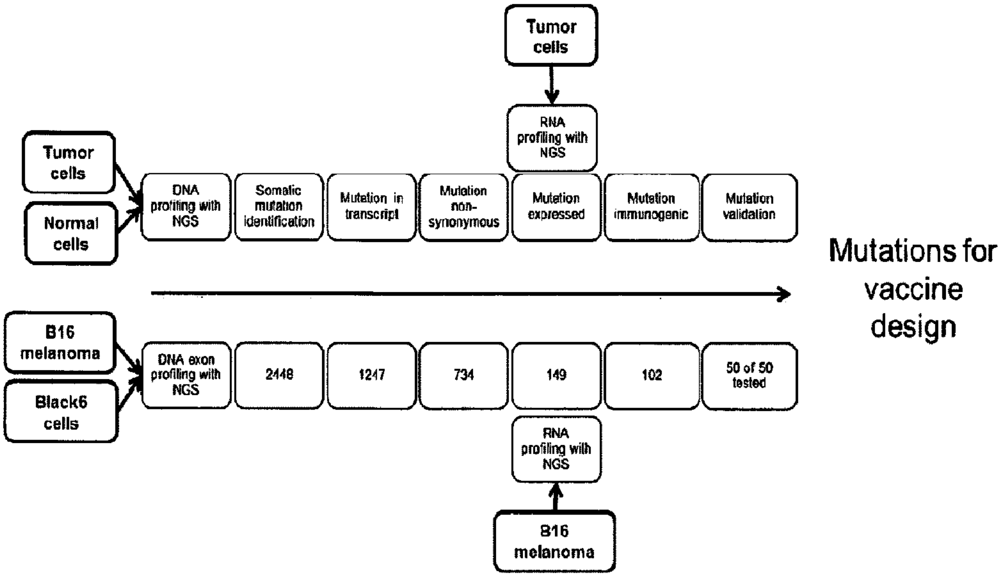
(30) **Priorités/Priorities:** 2011/05/24 (EP PCT/EP2011/002576);
2012/01/02 (EP PCT/EP2012/000006)

(51) **Cl.Int./Int.Cl.** ***A61K 39/00*** (2006.01),
A61P 35/00 (2006.01), ***G01N 33/50*** (2006.01)

(72) **Inventeurs/Inventors:**
SAHIN, UGUR, DE;
KREITER, SEBASTIAN, DE;
DIKEN, MUSTAFA, DE;
DIEKMANN, JAN, DE;
KOSLOWSKI, MICHAEL, DE;
...

(73) **Propriétaires/Owners:**

(54) **Titre : VACCINS INDIVIDUALISES POUR LE CANCER**
(54) **Title: INDIVIDUALIZED VACCINES FOR CANCER**



(57) **Abrégé/Abstract:**
The present invention relates to the provision of vaccines which are specific for a patient's tumor and are potentially useful for immunotherapy of the primary tumor as well as tumor metastases. In one aspect, the present invention relates to a method for providing an individualized cancer vaccine comprising the steps: (a) identifying cancer specific somatic mutations in a tumor specimen of a cancer patient to provide a cancer mutation signature of the patient; and (b) providing a vaccine featuring the cancer mutation signature obtained in step (a). In a further aspect, the present invention relates to vaccines which are obtainable by said method.

(11)(21) **2 836 494**

(13) **C**

(72) **Inventeurs(suite)/Inventors(continued):** BRITTEN, CEDRIK, DE; CASTLE, JOHN, DE; LOWER, MARTIN, DE; RENARD, BERNHARD, DE; OMOKOKO, TANA, DE; DE GRAAF, JOHANNES HENDRIKUS, DE

(73) **Propriétaires(suite)/Owners(continued):**
TRON-TRANSLATIONALE ONKOLOGIE AN DER UNIVERSITATSMEDIZIN DER JOHANNES GUTENBERG-UNIVERSITAT MAINZ GEMEINNUTZIGE GMBH, DE; BIONTECH SE, DE

(74) **Agent:** BERESKIN & PARR LLP/S.E.N.C.R.L.,S.R.L.

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
29 November 2012 (29.11.2012)

WIPO | PCT

(10) International Publication Number
WO 2012/159754 A3

(51) International Patent Classification:
A61K 39/00 (2006.01) *G01N 33/50* (2006.01)
C12Q 1/68 (2006.01) *G06F 19/18* (2011.01)
A61P 35/00 (2006.01) *G06F 19/24* (2011.01)

(21) International Application Number:
PCT/EP2012/002209

(22) International Filing Date:
23 May 2012 (23.05.2012)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
PCT/EP2011/002576 24 May 2011 (24.05.2011) EP
PCT/EP2012/000006
2 January 2012 (02.01.2012) EP

(71) Applicants *(for all designated States except US)*: **BIONTECH AG** [DE/DE]; Hölderlinstrasse 8, 55131 Mainz (DE). **TRON -Translationale Onkologie an der Universitätsmedizin der Johannes Gutenberg-Universität Mainz gemeinnützige GmbH** [DE/DE]; Langenbeckstrasse 1, 55131 Mainz (DE).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **SAHIN, Ugur** [DE/DE]; Philipp-von-Zabern-Platz 1, 55116 Mainz (DE). **KREITER, Sebastian** [DE/DE]; Külbstrasse 16, 55131 Mainz (DE). **DIKEN, Mustafa** [TR/DE]; Bettelpfad 5, 55130 Mainz (DE). **DIEKMANN, Jan** [DE/DE]; Holunderweg 45, 55128 Mainz (DE). **KOSLOWSKI, Michael** [DE/DE]; Autogenstrasse 50, 65933 Frankfurt (DE). **BRITTEN, Cedrik** [DE/DE]; Faberrebenweg 16, 55129 Mainz (DE). **CASTLE, John** [US/DE]; Rudolf-Diesel-Strasse 40, 55131 Mainz (DE). **LÖWER, Martin** [DE/DE]; Freseniusstrasse 19, 65193 Wiesbaden (DE). **RENARD, Bernhard** [DE/DE]; Josefsstrasse 55, 55118 Mainz (DE). **OMOKOKO, Tana** [DE/DE]; Klarastrasse 1, 55116 Mainz (DE). **DE GRAAF, Johannes Hendrikus** [NL/DE]; Nelkenstrasse 5, 55296 Lörzweiler (DE).

(74) Agents: **VOSSIUS, Volker** et al.; Geibelstrasse 6, 81679 Munich (DE).

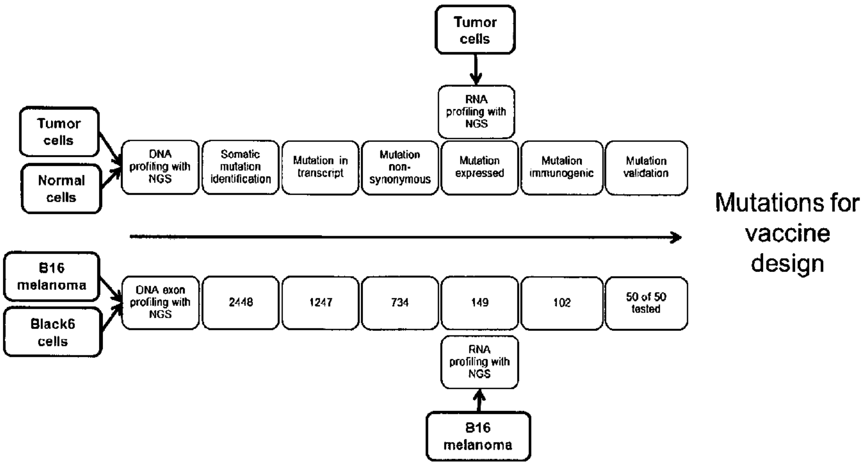
(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, 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, 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).

[Continued on next page]

(54) Title: INDIVIDUALIZED VACCINES FOR CANCER

Fig. 1



(57) Abstract: The present invention relates to the provision of vaccines which are specific for a patient's tumor and are potentially useful for immunotherapy of the primary tumor as well as tumor metastases. In one aspect, the present invention relates to a method for providing an individualized cancer vaccine comprising the steps: (a) identifying cancer specific somatic mutations in a tumor specimen of a cancer patient to provide a cancer mutation signature of the patient; and (b) providing a vaccine featuring the cancer mutation signature obtained in step (a). In a further aspect, the present invention relates to vaccines which are obtainable by said method.

WO 2012/159754 A3

WO 2012/159754 A3

Published:

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

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

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

(88) Date of publication of the international search report:
27 June 2013

INDIVIDUALIZED VACCINES FOR CANCER

TECHNICAL FIELD OF THE INVENTION

The present invention relates to the provision of vaccines which are specific for a patient's tumor and are potentially useful for immunotherapy of the primary tumor as well as tumor metastases.

BACKGROUND OF THE INVENTION

Cancer is a primary cause of mortality, accounting for 1 in 4 of all deaths. The treatment of cancer has traditionally been based on the law of averages - what works best for the largest number of patients. However, owing to the molecular heterogeneity in cancer, often less than 25% of treated individuals profit from the approved therapies. Individualized medicine based on tailored treatment of patients is regarded as a potential solution to low efficacies and high costs for innovation in drug development.

Antigen specific immunotherapy aims to enhance or induce specific immune responses in patients and has been successfully used to control cancer diseases. T cells play a central role in cell-mediated immunity in humans and animals. The recognition and binding of a particular antigen is mediated by the T cell receptors (TCRs) expressed on the surface of T cells. The T cell receptor (TCR) of a T cell is able to interact with immunogenic peptides (epitopes) bound to major histocompatibility complex (MHC) molecules and presented on the surface of target cells. Specific binding of the TCR triggers a signal cascade inside the T cell leading to proliferation and differentiation into a maturated effector T cell.

The identification of a growing number of pathogen- and tumor-associated antigens (TAA) led to a broad collection of suitable targets for immunotherapy. Cells presenting immunogenic peptides (epitopes) derived from these antigens can be specifically targeted by either active or passive immunization strategies. Active immunization may tend to induce and expand antigen specific T cells in the patient, which are able to specifically recognize and kill diseased cells. Different antigen formats can be used for tumor vaccination including whole cancer cells, proteins, peptides or immunizing vectors such as RNA, DNA or viral vectors that can be applied either directly *in vivo* or *in vitro* by pulsing of DCs following transfer into the patient.

Cancers may arise from the accumulation of genomic mutations and epigenetic changes, of which a fraction may have a causative role. In addition to tumor associated antigens, human cancers carry on average 100-120 non-synonymous mutations, of which many are targetable by vaccines. More than 95% of mutations in a tumor are unique and patient specific *(Weide et al. 2008: J. Immunother. 31, 180-188)*. The number of protein changing somatic mutations, which may result in tumor specific T cell epitopes, is in the range of 30 to 400. It has been predicted in silico that there are 40 to 60 HLA class I restricted epitopes per patient derived from tumor specific somatic mutations *(Azuma et al. 1993: Nature 366, 76-79)*. Moreover, *de novo* immunogenic HLA class II restricted epitopes likely also result from tumor-associated mutations, however their number is still unknown.

Notably, some non-synonymous mutations are causally involved in neoplastic transformation, crucial for maintaining the oncogenic phenotype (driver mutations) and may represent a potential "Achilles' heel" of cancer cells. As such non-synonymous mutations are not subject to central immune tolerance, they can be ideal candidates for individual cancer vaccine development. Mutations found in the primary tumor may also be present in metastases. However, several studies demonstrated that metastatic tumors of a patient acquire additional genetic mutations during individual tumor evolution which are often clinically relevant *(Suzuki et al. 2007: Mol. Oncol. 1 (2), 172- 180; Campbell et al. 2010: Nature 467 (7319), 1109-1113)*. Furthermore, also the molecular characteristics of many metastases deviate significantly from those of primary tumors.

The technical problem underlying the present invention is to provide a highly effective individualized cancer vaccine.

The present invention is based on the identification of patient specific cancer mutations and targeting a patient's individual cancer mutation "signature". Specifically, the present invention which involves a genome, preferably exome, or transcriptome sequencing based individualized immunotherapy approach aims at immunotherapeutically targeting multiple individual mutations in cancer. Sequencing using Next Generation Sequencing (NGS) allows a fast and cost effective identification of patient specific cancer mutations.

The identification of non-synonymous point mutations resulting in amino acid changes that will be presented the patient's major histocompatibility complex (MHC) molecules provides novel epitopes (neo-epitopes) which are specific for the patient's cancer but are not found in normal cells of the patient. Collecting a set of mutations from cancer cells such as circulating tumor cells (CTC) allows the provision of a vaccine which induces an immune response potentially targeting the primary tumor even if containing genetically distinct subpopulations as well as tumor metastases. For vaccination, such neo-epitopes identified according to the present application are provided in a patient in the form of a polypeptide comprising said neo-epitopes and following appropriate processing and presentation by MHC molecules the neo-epitopes are displayed to the patient's immune system for stimulation of appropriate T cells.

Preferably, such polypeptide is provided in the patient by administering RNA encoding the polypeptide. A strategy wherein *in vitro* transcribed RNA (IVT-RNA) is directly injected into a patient by different immunization routes has been successfully tested in various animal models. RNA may be translated in transfected cells and the expressed protein following processing presented on the MHC molecules on the surface of the cells to elicit an immune response.

The advantages of using RNA as a kind of reversible gene therapy include transient expression and a non-transforming character. RNA does not need to enter the nucleus in order to be expressed and moreover cannot integrate into the host genome, thereby eliminating the risk of oncogenesis. Transfection rates attainable with RNA are relatively high. Furthermore, the amounts of protein achieved correspond to those in physiological expression.

The rationale for the immunotherapeutic targeting of multiple individual mutations is that (i) these mutations are exclusively expressed, (ii) mutated epitopes can be expected to be ideal for T cell immunotherapy since T cells recognizing them have not undergone thymic selection, (iii) tumor immune escape can be reduced e.g. by targeting “driver mutations” that are highly relevant for the tumor phenotype, and (iv) a multiepitopic immune response has a higher likelihood to result in improved clinical benefit.

DESCRIPTION OF INVENTION

SUMMARY OF THE INVENTION

The present invention relates to efficient methods for providing individualized recombinant cancer vaccines inducing an efficient and specific immune response in a cancer patient and potentially targeting the primary tumor as well as tumor metastases. The cancer vaccines provided according to the invention when administered to a patent provide a collection of MHC presented epitopes specific for the patient's tumor suitable for stimulating, priming and/or expanding T cells directed against cells expressing antigens from which the MHC presented 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 cancer expressed antigens 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.

In one aspect, the present invention relates to a method for providing an individualized cancer vaccine comprising the steps:
(a) identifying cancer specific somatic mutations in a tumor specimen of a cancer patient to provide a cancer mutation signature of the patient; and
(b) providing a vaccine featuring the cancer mutation signature obtained in step (a).

In one embodiment, the method of the invention comprises the following steps:

i) providing a tumor specimen from a cancer patient and a non-tumorigenous specimen which preferably is derived from the cancer patient;

ii) identifying sequence differences between the genome, exome and/or transcriptome of the tumor specimen and the genome, exome and/or transcriptome of the non-tumorigenous specimen;

iii) designing a polypeptide comprising epitopes incorporating the sequence differences determined in step (ii);

iv) providing the polypeptide designed in step (iii) or a nucleic acid, preferably RNA, encoding said polypeptide; and

v) providing a vaccine comprising the polypeptide or nucleic acid provided in step (iv).

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.

According to the invention, 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 method of the invention provides for the identification of a number of mutations which provides a sufficient number of neo-epitopes to be included into a vaccine. A "cancer mutation" relates to a sequence difference between the nucleic acid contained in a cancer cell and the nucleic acid contained in a normal cell.

Preferably, the mutations identified in the methods 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 method of 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. The exome is 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).* Thus, the method of 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. The transcriptome is 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. Thus, the method of 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 method of 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)

A vaccine provided according to the methods of the present invention relates to a vaccine which when administered to a patent preferably provides a collection of MHC presented 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 MHC presented epitopes, incorporating sequence changes based on the identified mutations or sequence differences. Such MHC presented epitopes incorporating sequence changes based on the identified mutations or sequence differences 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 MHC presented epitopes are derived and presenting the same epitopes on the surface of the tumor cells.

For providing a vaccine, the method of the invention may comprise the arbitrary inclusion of a sufficient number of neo-epitopes (preferably in the form of an encoding nucleic acid) into a vaccine or it may comprise the further step of determining the usability of the identified mutations in epitopes for cancer vaccination. Thus further steps can involve one or more of the following: (i) assessing whether the sequence changes are located in known or predicted MHC presented epitopes, (ii) *in vitro* and/or *in silico* testing whether the sequence changes are located in MHC presented epitopes, e.g. testing whether the sequence changes are part of peptide sequences which are processed into and/or presented as MHC presented epitopes, and (iii) *in vitro* testing whether the envisaged mutated 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 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.

Mutations or sequence differences 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 mutations 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 one or more, preferably all of the following steps:

- identifying expressed, protein modifying mutations, e.g. by analyzing transcripts;
- identifying mutations which are potentially immunogenic, i.e. by comparing the data obtained with available datasets of confirmed immunogenic epitopes, e.g. those contained in public immune epitope databases such as i.e. the IMMUNE EPITOPE DATABASE AND ANALYSIS RESOURCE

The step of identifying mutations which are potentially immunogenic may comprise determining and/or ranking epitopes according to a prediction of their MHC-binding capacity, preferably MHC class-I binding capacity.

In another embodiment of the invention, the epitopes can be selected and/or ranked by using further parameters such as protein impact, associated gene expression, sequence uniqueness, predicted presentation likelihood, and association with oncogenes.

Multiple CTC analyses also allow selection and prioritization of mutations. For example, a mutation which is found in a larger portion of CTC may be prioritized higher than a mutation found in a lower portion of CTC.

The collection of mutation based neo-epitopes identified according to the invention and provided by a vaccine of the invention is preferably present in the form of a polypeptide comprising said neo-epitopes (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 $(GGS)_a(GSS)_b(GGG)_c(SSG)_d(GSG)_e$ 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 another embodiment of the present invention the collection of mutation based neo-epitopes identified according to the invention and provided by a vaccine of the invention is preferably present in the form of a collection of polypeptides comprising said neo-epitopes on different polypeptides, wherein said polypeptides each comprise one or more neo-epitopes, which can also be overlapping, or a collection of nucleic acids, in particular RNAs, encoding said polypeptides.

In one particularly preferred embodiment, 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 II-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 II-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.

In a further aspect, the present invention provides a vaccine which is obtainable by the method according to the invention. Accordingly, the present invention relates to a vaccine comprising a recombinant polypeptide comprising mutation based neo-epitopes, said neo-epitopes resulting from cancer specific somatic mutations in a tumor specimen of a cancer patient, or a nucleic acid encoding said polypeptide. Such recombinant polypeptide may also include epitopes not including cancer specific somatic mutations as discussed above.

Preferred embodiments of such vaccine are as described above in the context of the method of the invention.

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 an individualized cancer vaccine by the method 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.

Another aspect of the invention relates to a method for determining a false discovery rate based on next generation sequencing data, said method including:

taking a first sample of genetic material from an animal or human;

taking a second sample of genetic material from an animal or human;

taking a first sample of genetic material from tumor cells;

taking a second sample of genetic material from said tumor cells;

determining a common coverage tumor comparison by counting all bases of the reference genome which is included in both the tumor and at least one of said first sample of genetic material from an animal or human and said second sample of genetic material from an animal or human;

determining a common coverage same vs. same comparison by counting all bases of the reference genome which are covered by both said first sample of genetic material from an animal or human and said second sample of genetic material from an animal or human;

dividing said common coverage tumor comparison by said common coverage same vs. same comparison to form a normalization;

determining a false discovery rate by dividing 1) the number of single nucleotide variations with a quality score greater than Q in a comparison of said first sample of genetic material from an animal or human and said second sample of genetic material from an animal or human, by 2) the number of single nucleotide variations with a quality score greater than Q in a comparison of said first sample of genetic material from said tumor cells and said second sample of genetic material from said tumor cells and 3) multiplying the result by said normalization.

In one embodiment, said genetic material is a DNA.

In one embodiment, Q is determined by:

establishing a set of quality properties $S=(s_1,\ldots,s_n)$ wherein S is preferable to $T=(t_1,\ldots,t_n)$, denoted by S>T, when $s_i > t_i$ for all $i=1,\ldots,n$;

defining an intermediate false discovery rate by dividing 1) the number of single nucleotide variations with a quality score S>T in a comparison of said first DNA sample from an animal or human and said second DNA sample from an animal or human, by 2) the number of single nucleotide variations with a quality score S>T in a comparison of said first DNA sample from said tumor cells and said second DNA sample from said tumor cells and 3) multiplying the result by said normalization,

determining the value range for each property for m mutations with n quality properties each;

sampling up to p values out of said value range;

creating each possible combination of sampled quality values which results in p^n data points;

using a random sample of said data points as a predictor for random forest training;

using the corresponding intermediate false discovery rate value as a response for said random forest training,

wherein the resulting regression score of said random forest training is Q.

In one embodiment, said second DNA sample from an animal or human is allogenic to said first DNA sample from an animal or human. In one embodiment, said second DNA sample from an animal or human is autologous to said first DNA sample from an animal or human. In one embodiment, said second DNA sample from an animal or human is xenogenic to said first DNA sample from an animal or human.

In one embodiment, said genetic material is a RNA.

In one embodiment, Q is determined by:

establishing a set of quality properties $S=(s_1,\ldots,s_n)$ wherein S is preferable to $T=(t_1,\ldots,t_n)$, denoted by S>T, when $s_i > t_i$ for all $i=1,\ldots,n$;

defining an intermediate false discovery rate by dividing 1) the number of single nucleotide variations with a quality score S>T in a comparison of said first RNA sample from an animal or human and said second RNA sample from an animal or human, by 2) the number of single nucleotide variations with a quality score S>T in a comparison of said first RNA sample from said tumor cells and said second RNA sample from said tumor cells and 3) multiplying the result by said normalization,

determining the value range for each property for *m* mutations with *n* quality properties each;

sampling up to *p* values out of said value range;

creating each possible combination of sampled quality values which results in p^n data points;

using a random sample of said data points as a predictor for random forest training;

using the corresponding intermediate false discovery rate value as a response for said random forest training,

wherein the resulting regression score of said random forest training is Q.

In one embodiment, said second RNA sample from an animal or human is allogenic to said first RNA sample from an animal or human. In one embodiment, said second RNA sample from an animal or human is autologous to said first RNA sample from an animal or human. In one embodiment, said second RNA sample from an animal or human is xenogenic to said first RNA sample from an animal or human.

In one embodiment, said false discovery rate is used to make a vaccine formulation. In one embodiment, said vaccine is deliverable intravenously. In one embodiment, said vaccine is deliverable dermally. In one embodiment, said vaccine is deliverable muscularly. In one embodiment, said vaccine is deliverable subcutaneously. In one embodiment, said vaccine is tailored for a specific patient.

In one embodiment, one of said first sample of genetic material from an animal or human and said second sample of genetic material from an animal or human is from said specific patient.

In one embodiment, said step of determining a common coverage tumor comparison by counting all bases of the reference genome which is included in both the tumor and at least one of said first sample of genetic material from an animal or human and said second sample of genetic material from an animal or human uses an automated system to count all bases.

In one embodiment, said step of determining a common coverage same vs. same comparison by counting all bases of the reference genome which are covered by both said first sample of genetic material from an animal or human and said second sample of genetic material from an animal or human uses said automated system.

In one embodiment, said step of dividing said common coverage tumor comparison by said common coverage same vs. same comparison to form a normalization uses said automated system.

In one embodiment, said step of determining a false discovery rate by dividing 1) the number of single nucleotide variations with a quality score greater than Q in a comparison of said first sample of genetic material from an animal or human and said second sample of genetic material from an animal or human, by 2) the number of single nucleotide variations with a quality score greater than Q in a comparison of said first sample of genetic material from said tumor cells and said second sample of genetic material from said tumor cells and 3) multiplying the result by said normalization uses said automated system.

Another aspect of the invention relates to a method for determining an estimated receiver operating curve (ROC), said method including:

receiving a dataset of mutations, each mutation associated with a false discovery rate (FDR); and

for each mutation:

determining a true positive rate (TPR) by subtracting said FDR from one; and

determining a false positive rate (FPR) by setting said FPR equal to said FDR; and

forming an estimated ROC by plotting, for each mutation, a point at the cumulative TPR and FPR values up to said mutation, divided by the sum of all TPR and FPR values.

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. For example, if in a preferred embodiment RNA comprises a poly(A)-tail consisting of 120 nucleotides and in another preferred embodiment the RNA molecule comprises a 5'-cap analog, then in a preferred embodiment, the RNA comprises the poly(A)-tail consisting of 120 nucleotides and the 5'-cap analog.

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. Kölbl, 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.

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

The vaccine provided according to the invention is 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.

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 "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.

The term "immune response" refers to an integrated bodily response to 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 against a particular antigen before induction, but it may also mean that there was a certain level of immune response against a particular antigen 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.

A "cellular immune response", a "cellular response", a "cellular response against an antigen" or a similar term is meant to include a cellular response directed to cells characterized by presentation of an antigen with class I or class II MHC. The cellular response relates to cells called 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 that will elicit an immune response. In particular, an "antigen" relates to any substance, preferably a peptide or protein, that reacts specifically with antibodies or T-lymphocytes (T cells). According to the present invention, the term "antigen" comprises any molecule which comprises at least one 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). According to the present invention, any suitable antigen may be used, which is a candidate for an immune reaction, wherein the immune reaction is preferably a cellular immune reaction. In the context of the embodiments of the present invention, the antigen 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 reaction against the antigen. An antigen is preferably a product which corresponds to or is derived from a naturally occurring antigen. Such naturally occurring antigens include tumor antigens.

In a preferred embodiment, the antigen is a tumor antigen, 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, α1-fetoprotein, isoferritin, and fetal sulphoglycoprotein, α2-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. In one embodiment, the term "tumor antigen" or "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 antigen 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 of an antigen to induce an immune reaction.

An "antigen peptide" 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, an antigen peptide 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). Preferably, the antigen peptides according to the invention are MHC class I and/or class II presented peptides or can be processed to produce MHC class I and/or class II presented peptides. Preferably, the antigen peptides comprise 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. Preferably, an antigen peptide according to the invention comprises an amino acid sequence substantially corresponding to the amino acid sequence of such fragment and is processed to produce such fragment, i.e., an MHC class I and/or class II presented peptide derived from an antigen.

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 7-20 amino acids in length, more preferably 7-12 amino acids in length, more preferably 8-11 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 7-20 amino acids in length, more preferably 7-12 amino acids in length, more preferably 8-11 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, i.e., its sequence substantially corresponds and is preferably completely identical to a fragment of an antigen. Thus, an antigen peptide or vaccine sequence according to the invention in one embodiment comprises a sequence of 7-20 amino acids in length, more preferably 7-12 amino acids in length, more preferably 8-11 amino acids in length, in particular 9 or 10 amino acids in length which substantially corresponds and is preferably completely identical to a fragment of an antigen and following processing of the antigen peptide or vaccine sequence makes up the presented peptide. According to the invention, such peptide produced by processing comprises the identified sequence change.

According to the invention, an antigen peptide or 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 antigen peptide or epitope. The presented antigen peptide or epitope is produced following suitable processing.

Peptides having amino acid sequences substantially corresponding to a sequence of a peptide which is presented by the class I MHC may differ at one or more residues that are not essential for TCR recognition of the peptide as presented by the class I MHC, or for peptide binding to MHC. Such substantially corresponding peptides are also capable of stimulating an antigen-responsive CTL and may be considered immunologically equivalent. Peptides having amino acid sequences differing from a presented peptide at residues that do not affect TCR recognition but improve the stability of binding to MHC may improve the immunogenicity of the antigen peptide, and may be referred to herein as "optimized peptide". Using existing knowledge about which of these residues may be more likely to affect binding either to the MHC or to the TCR, a rational approach to the design of substantially corresponding peptides may be employed. Resulting peptides that are functional are contemplated as antigen peptides.

An antigen peptide when presented by MHC should be recognizable by a T cell receptor. Preferably, the antigen peptide if 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 antigen peptide. Preferably, antigen peptides, in particular if presented in the context of MHC molecules, are capable of stimulating an immune response, preferably a cellular response against the antigen from which they are derived or cells characterized by expression of the antigen and preferably characterized by presentation of the antigen. Preferably, an antigen peptide is capable of stimulating a cellular response against a cell characterized by presentation of the antigen with class I MHC and preferably is capable of stimulating an antigen-responsive CTL. Such cell preferably is a target cell.

"Antigen processing" or "processing" refers to the degradation of a polypeptide or antigen into procession products, which are fragments of said polypeptide or antigen (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 Fcγ 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 CD11) 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 IFNγ.

"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 the antigen.

In some embodiments, a pharmaceutical composition of the invention comprising a gene 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 or an antigen epitope, 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 "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" or "antigen peptide". The term "MHC binding peptide" relates to a peptide which binds to an MHC class I and/or an MHC class II molecule. In the case of class I MHC/peptide complexes, the binding peptides are typically 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 10-25 amino acids long and are in particular 13-18 amino acids long, whereas longer and shorter peptides may be effective.

The terms "epitope", "antigen peptide", "antigen epitope", "immunogenic peptide" and "MHC binding peptide" are used interchangeably herein and preferably relate to an incomplete representation of an antigen which is preferably capable of eliciting an immune response against the antigen or a cell expressing or comprising and preferably presenting the antigen.

Preferably, the terms relate to an immunogenic portion of an antigen. Preferably, it is a portion of an antigen that is recognized (i.e., specifically bound) by a T cell receptor, in particular if presented in the context of MHC molecules. Preferred such immunogenic portions bind to an MHC class I or class II molecule. As used herein, an immunogenic portion is said to "bind to" an MHC class I or class II molecule if such binding is detectable using any assay known in the art.

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.

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 an antigen peptide derived from an antigen 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 an antigen peptide derived from an antigen 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 an antigen peptide derived from said antigen to be responsive or reactive. If the cell is a helper T cell ($CD4^+$ T cell) bearing receptors that recognize an antigen or an antigen peptide derived from an antigen 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-α, 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 recognizes an antigen or an antigen peptide derived from an antigen 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 (TCRα and TCRβ) genes and are called α- 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 major histocompatibility complex (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 a professional antigen presenting cell (APC), usually a dendritic cell in the case of naïve responses, although B cells and macrophages can be important APCs. The peptides presented to CD8+ T cells by MHC class I molecules are typically 8-10 amino acids in length; the peptides presented to CD4+ T cells by MHC class II molecules are typically longer, as the ends of the binding cleft of the MHC class II molecule are open.

According to the present invention, a T cell receptor is capable of binding to a predetermined 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 T cell receptor 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.

A T cell receptor is preferably capable of binding specifically to a predetermined target. A T cell receptor is specific for a predetermined target if it is capable of binding to said predetermined target while it is not (substantially) capable of binding to other targets, i.e. has no significant affinity for other targets and does not significantly bind to other targets in standard assays.

Cytotoxic T lymphocytes may be generated *in vivo* by incorporation of an antigen or an antigen peptide into antigen-presenting cells *in vivo*. The antigen or antigen peptide 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 *98*: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.

The term "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 α-chain and β2-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 α- 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.

In humans, genes in the MHC region that encode antigen-presenting proteins on the cell surface are referred to as human leukocyte antigen (HLA) genes. However the abbreviation MHC is often used to refer to HLA gene products. HLA genes include the nine so-called classical MHC genes: HLA-A, HLA-B, HLA-C, HLA-DPA1, HLA-DPB1, HLA-DQA1, HLA-DQB1, HLA-DRA, and HLA-DRB1.

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

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 peptide derived from an antigen 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 peptide derived from an antigen 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-α, and specific cytolytic killing of antigen expressing target cells.

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 the coding regions of a genome. The term "transcriptome" relates to the set of all RNA molecules.
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, 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 TM 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 TM 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 is 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^{TM} DNA Sample Preparation Kit and the Exome Enrichment Kit TruSeq^{TM} 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.

According to the invention, a high-throughput genome-wide single cell genotyping method can be applied.

In one embodiment of the high-throughput genome-wide single cell genotyping the Fluidigm platform may be used. Such approach may comprise the following steps:

1. Sample tumor tissue/cells and healthy tissue from a given patient.
2. The genetic material is extracted from cancerous and healthy cells and then its exome (DNA) is sequenced using standard next generation sequencing (NGS) protocols. The coverage of the NGS is such that heterozygote alleles with at least 5% frequency can be detected. The transcriptome (RNA) is also extracted from the cancer cells, converted into cDNA and sequenced to determine which genes are expressed by the cancer cells.
3. Non-synonymous expressed single nucleotide variations (SNVs) are identified as described herein. Sites that are SNPs in the healthy tissue are filtered out.
4. N=96 mutations from (3) are selected spanning different frequencies. SNP genotyping assays based on florescence detection are designed and synthesized for these mutations (examples of such assays include: TaqMan based SNP assays by Life Technologies or

SNPtype assays by Fluidigm). Assays will include specific target amplification (STA) primers to amplify amplicons containing the given SNV (this is standard in TaqMan and SNPtype assays).

5. Individual cells will be isolated from the tumor and from healthy tissue either by laser microdissection (LMD) or by disaggregation into single-cell suspensions followed by sorting as previously described (*Dalerba P. et al. (2011) Nature Biotechnology 29: 1120-1127*). Cells can either be chosen without pre-selection (i.e., unbiased), or alternatively, cancerous cells can be enriched. Enriching methods include: specific staining, sorting by cell size, histological inspection during LMD, and so on.

6. Individual cells will be isolated in PCR tubes containing a master mix with the STA primers and the amplicons containing the SNVs will be amplified. Alternatively the genome of the single cell will be amplified via whole genome amplification (WGA) as previously described (*Frumkin D. et al. (2008) Cancer Research 68: 5924*). Cell lysis will be achieved either via the 95°C heating step or via a dedicated lysis buffer.

7. STA amplified samples are diluted and loaded onto the Fluidigm genotyping array.

8. Samples from healthy tissue will be used as positive controls to determine homozygote allele clusters (no mutation). Since NGS data indicates that homozygote mutations are extremely rare, typically only two clusters are expected: XX and XY, with X=healthy.

9. The number of arrays that can be executed is not limited, allowing, in practice up to ~1000 single cells to be assayed (~10 arrays). If performed in 384 plates sample prep can be reduced to a few days.

10. SNVs for each cell are then determined.

In another embodiment of the high-throughput genome-wide single cell genotyping the NGS platform may be used. Such approach may comprise the following steps:

1. Steps 1 through 6 above are identical, except that N (number of SNVs assayed) can be much larger than 96. In case of WGA, several cycles of STA will be performed after. STA primers will contain two universal tag sequences on each primer.

2. After the STA, barcode primers will be PCR amplified into the amplicons. Barcode primers contain unique barcode sequences and the above universal tag sequences. Each cell will thus contain a unique barcode.

3. Amplicons from all cells will be mixed and sequenced via NGS. The practical limitation on the number of cells that can be multiplexed is the number of plates that can be prepared. Since samples can be prepared in 384 plates, a practical limit would be ~5000 cells.

4. Based on sequence data SNVs (or other structural anomalies) of the individual cells are detected.

For prioritizing antigens, tumor phylogenetic reconstruction based on single cell genotyping ("phylogenetic antigen prioritization") may be used according to the invention. Besides antigen prioritization based on criteria such as expression, the type of mutation (non-synonymous versus other), MHC binding characteristics and so on, a further dimension for prioritization designed to cope with intra and inter-tumor heterogeneity and biopsy bias can be used as described for example below.

1. Identifying the most abundant antigens

The frequency of each SNV can be accurately estimated based on the single cell assay described above in connection with the high-throughput genome-wide single cell genotyping method and the most abundant SNVs present can be selected for providing individualized vaccines for cancer (IVAC).

2. Identifying primary basal antigens based on rooted tree analysis

NGS data from tumors suggest that homozygote mutations (hits in both alleles) are rare events. Therefore there is no need for haplotyping and a phylogenetic tree of the tumor somatic mutations can be created from the singe cell SNV dataset. The germline sequence will be used to root the tree. Using algorithms to reproduce ancestral sequences the sequences of nodes near the root of the tree will be reproduced. These sequences contain the earliest mutations predicted to exist in the primary tumor (defined here as the *primary basal mutations/antigens*). Due to the low probability that two mutations will occur on the same alleles in the same position on the genome, the mutations in the ancestral sequences are predicted to be fixed in the tumor.

Prioritizing primary basal antigens is not equivalent to prioritizing the most frequent mutations in the biopsy (although primary basal mutations are expected to be among the most frequent in the biopsy). The reason is the following: say two SNVs appear to be present in all cells derived from a biopsy (and thus have the same frequency – 100%), but one mutation is basal and the other is not, then the basal mutation should be selected for IVAC. This is because the basal mutation is likely to present in all regions of the tumor, whereas the latter mutation may be a more recent mutation that by chance was fixed in the region where the biopsy was taken. In addition, basal antigens are likely to exist in metastatic tumors derived from the primary tumor. Therefore by prioritizing basal antigens for IVAC one may greatly increase the chance that IVAC will be able to eradicate the entire tumor and not just a part of the tumor.

If secondary tumors exist and these were also sampled, an evolutionary tree of the all tumors can be estimated. This can improve the robustness of the tree and allow the detection of mutations basal to all tumors.

3. Identifying antigens that maximally span the tumor(s)

Another approach to obtaining antigens that maximally cover all tumor sites is to take several biopsies from the tumor. One strategy would be to select antigens identified by the NGS analysis to be present in all biopsies. To improve the odds of identifying basal mutations, a phylogenetic analysis based on single cell mutations from all biopsies can be performed.

In case of metastasis, biopsies from all tumors can be obtained and mutations identified via NGS which are common to all tumors can be selected.

4. Using CTCs to prioritize antigens that inhibit metastasis

It is believed that metastatic tumors are derived from single cells. Therefore by genotyping individual cells extracted from different tumors of a given patient in conjunction with genotyping the patient's circulating tumor cells (CTCs), one can reconstruct the evolutionary history of the cancer. The expectation is to observe the metastatic tumor evolving from the original tumor through a clade of CTCs derived from the primary tumor.

Below (unbiased method to identify, count and genetically probe CTCs) we describe an extension of the above described high-throughput genome-wide single cell genotyping method for an unbiased isolation and genomic analysis CTCs. Using the analysis described above, one can then reconstruct a phylogenetic tree of the primer tumor, CTCs and secondary tumors arising from metastasis (if they exist). Based on this tree one can identify mutations (passenger or driver) that occurred at the time or closely after CTCs first detached from the primary tumor. The expectation is that the genomes of CTCs arising from the primary tumor are evolutionary more similar to the primary tumor genomes than to secondary tumor genomes. Furthermore it is expected that the genomes of CTCs arising from the primary tumor will contain unique mutations that are fixed in the secondary tumors, or that will likely be fixed if secondary tumors will be formed in the future. These unique mutations can be prioritized for IVAC to target (or prevent) metastasis.

The advantage of prioritizing CTC mutations versus primary basal mutations is that antigens derived from CTCs can mobilize T cells specifically to target metastasis, and therefore will be an independent arm from the T cells targeting the primary tumor (using different antiges). In addition, if there are few (or no) secondary tumors, then the chance for immune escape from CTC derived antigens is expected to be lower as the probably for tumor escape should scale with the number of cancer cells carrying the given antigen.

5. Identifying antigens co-occurring on the same cell (the "cocktail" IVAC)

It is believed that the tumor evolves to suppress mutations due to the selection pressure of the immune system and therapy. Cancer vaccines targeting multiple antigens that co-occur on the same cell and that are also frequent in the tumor have a greater chance of overriding tumor escape mechanisms and therefore reduce the chance for relapse. Such "cocktail vaccines" would be analogous to the antiretroviral combination therapy for HIV+ patients. Co-occurring mutations can be identified by phylogenetic analysis or by inspecting the SNV alignment of all cells.

Furthermore, according to the invention, an unbiased method to identify, count and genetically probe CTCs can be used. Such approach may comprise the following steps:

1. Obtain biopsy of tumor(s) and determine atlas of somatic mutations.
2. **Option 1:** Select N≥96 mutations for further investigation based on previously established prioritization schemes.

   Option 2: Perform single cell assay (see above described high-throughput genome-wide single cell genotyping method) followed phylogenetic analysis to select N≥96 primary basal mutations and possibly more recent mutations to maximize diversity. The former mutations are useful for identifying the CTCs (see below), and the latter for generating a phylogenetic analysis (see section "Identifying antigens co-occurring on the same cell (the "cocktail" IVAC")).
3. Obtain whole blood from the cancer patient
4. Lyse red blood cells

5. Remove white blood cells by depleting CD45+ cells (e.g., via sorting, magnetic beads conjugated to anti CD45 antibody, etc.) to enrich for CTCs.

6. Remove free DNA by DNAase digestion. The origin of free DNA can be DNA present in the blood or DNA from dead cells.

7. Sort remaining cells into PCR tubes, perform the STA (based on selected mutations) and screen on Fluidigm (above described high-throughput genome-wide single cell genotyping method). CTCs should generally be positive for multiple SNVs.

8. Cells identified as cancerous (=CTCs) can be then be further analyzed phylogenetically based on the panel of SNVs screened (see section "Identifying antigens co-occurring on the same cell (the "cocktail" IVAC")).

It is also possible to combine this method with previous established methods for isolated CTCs. For example, one can sort for EpCAM+ cells, or cells positive for cytokeratins (Rao CG. et al. (2005) International journal of oncology 27: 49; Allard WJ. et al. (2004) Clinical Cancer Research 10: 6897-6904). These putative CTCs can then be verified/profiled on the Fluidigm/NGS to derive their mutations.

This method can be used to count CTCs. Since the method does not rely on one particular marker, which may nor may not be expressed by the cancer cells, but rather on the mutation profile of cancer somatic mutations unique to the patient, this is an unbiased method to detect and enumerate CTCs.

According to the invention, an approach involving tumor phylogenetic reconstruction based on single cell genotyping to enrich for driver mutations ("phylogenetic filtering") may be used.

In one embodiment of this approach, a pan-tumor phylogenetic analysis to recover driver mutations is performed.

For example, driver mutations from n=1 tumors may be detected.

In the above section "Identifying primary basal antigens based on rooted tree analysis" we describe a method to recover ancestral sequences and/or identify cells that have sequences close to the root of the tree. The number of mutations in these sequences is expected to be significantly less than the number of mutations in the bulk sample of the cancer since by definition these are sequences close to the root of the tree. Therefore, by selecting sequences close to the root of the tree many passenger mutations are expected to be "phylogenetically filtered" out. This procedure has the potential to greatly enrich for driver mutations. Driver mutations can then be used to identify/selects treatment for a patient or can be used as leads for novel therapies.

In another example, driver mutations from n>1 tumors of a given type may be detected.

By reconstructing primary basal mutations from many tumors of a particular type one can greatly increase the chance of detecting driver mutations. Since basal sequences near the root of the tree filter out many passenger mutations, the signal to noise ratio in detecting driver mutations is expected to greatly increase. This method therefore has the potential to detect (1) less frequent driver mutation (2) frequent driver mutations from less samples.

In another embodiment of the approach involving tumor phylogenetic reconstruction based on single cell genotyping to enrich for driver mutations ("phylogenetic filtering"), a phylogenetic analysis to recover metastasis causing driver mutations is performed.

In the above section "Using CTCs to prioritize antigens that inhibit metastasis" we describe a method to detect CTC-associated mutations. This method can also be used to enrich for driver mutations leading to metastasis. For example, by mapping the combined phylogeny of the primer tumor, secondary tumors and CTCs, CTCs derived from the primary tumor should connect between the clades of the primary secondary tumors. Such a phylogenetic analysis can help pinpoint the mutations unique at this transition between primer and secondary tumors. A fraction of these mutations can be driver mutations. Furthermore, by comparing unique CTC mutations from different instances of the same cancer (i.e., n>1 tumors), one can further enrich for the unique driver mutations causing metastasis.

According to the invention, phylogenetic analysis to identify primary versus secondary tumors may be used.

In case of metastasis, if all tumors are sampled, a rooted tree can be used to predict the temporal order that tumors appeared: which tumor is the primary tumor (nodes closest to the root of the tree) and which tumors are the most recent ones. This can be helpful in cases where it is difficult to determine which tumor is the primary.

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.

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.

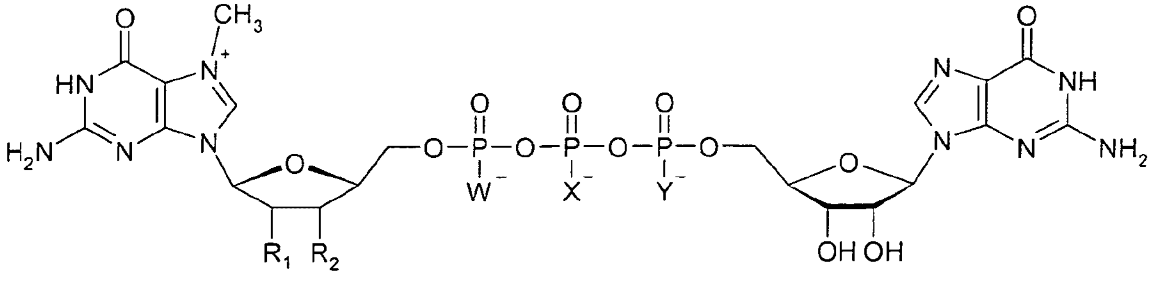
The term "modification" in the context of the RNA used in the present invention includes any modification of an RNA which is not naturally present in said RNA.

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.

Preferably, the 5' end of the RNA includes a Cap structure having the following general formula:



wherein R_1 and R_2 are independently hydroxy or methoxy and W^-, X^- and Y^- are independently oxygen, sulfur, selenium, or BH_3. In a preferred embodiment, R_1 and R_2 are hydroxy and W^-, X^- and Y^- are oxygen. In a further preferred embodiment, one of R_1 and R_2, preferably R_1 is hydroxy and the other is methoxy and W^-, X^- and Y^- are oxygen. In a further preferred embodiment, R_1 and R_2 are hydroxy and one of W^-, X^- and Y^-, preferably X^- is sulfur, selenium, or BH_3, preferably sulfur, while the other are oxygen. In a further preferred embodiment, one of R_1 and R_2, preferably R_2 is hydroxy and the other is methoxy and one of W^-, X^- and Y^-, preferably X^- is sulfur, selenium, or BH_3, preferably sulfur while the other are oxygen.

In the above formula, the nucleotide on the right hand side is connected to the RNA chain through its 3' group.

Those Cap structures wherein at least one of W^-, X^- and Y^- is sulfur, i.e. which have a phosphorothioate moiety, exist in different diastereoisomeric forms all of which are encompassed herein. Furthermore, the present invention encompasses all tautomers and stereoisomers of the above formula.

For example, the Cap structure having the above structure wherein R_1 is methoxy, R_2 is hydroxy, X^- is sulfur and W^- and Y^- are oxygen exists in two diastereoisomeric forms (Rp and Sp). These can be resolved by reverse phase HPLC and are named D1 and D2 according to their elution order from the reverse phase HPLC column. According to the invention, the D1 isomer of $m_2^{7,2'-O}Gpp_SpG$ is particularly preferred.

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, alpha1-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 or polypeptides, e.g. by transcription and/or translation. With respect to RNA, the term "expression" or "translation" relates in particular to the production of peptides or polypeptides. 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 or polypeptide.

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, the 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.

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 "sequence change" with respect to peptides or proteins relates to 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.

The term "derived" means according to the invention that a particular entity, in particular a particular sequence, is present in the object from which it is derived, in particular an organism or molecule. 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.

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 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 an antigen peptide. Particularly preferred diseases are cancer diseases. The 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.

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.

Malignant tumor is essentially synonymous with cancer. Malignancy, malignant neoplasm, and malignant tumor are essentially synonymous with cancer.

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.

A benign tumor is a tumor that lacks all three of the malignant properties of a cancer. Thus, by definition, a benign tumor does not grow in an unlimited, aggressive manner, does not invade surrounding tissues, and does not spread to non-adjacent tissues (metastasize).

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 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.

The main types of lung cancer are small cell lung carcinoma (SCLC) and non-small cell lung carcinoma (NSCLC). There are three main sub-types of the non-small cell lung carcinomas: squamous cell lung carcinoma, adenocarcinoma, and large cell lung carcinoma. Adenocarcinomas account for approximately 10% of lung cancers. This cancer usually is seen peripherally in the lungs, as opposed to small cell lung cancer and squamous cell lung cancer, which both tend to be more centrally located.

Skin cancer is a malignant growth on the skin. The most common skin cancers are basal cell cancer, squamous cell cancer, and melanoma. Malignant melanoma is a serious type of skin cancer. It is due to uncontrolled growth of pigment cells, called melanocytes.

According to the invention, a "carcinoma" is a malignant tumor derived from epithelial cells. This group represents the most common cancers, including the common forms of breast, prostate, lung and colon cancer.

"Bronchiolar carcinoma" is a carcinoma of the lung, thought to be derived from epithelium of terminal bronchioles, in which the neoplastic tissue extends along the alveolar walls and grows in small masses within the alveoli. Mucin may be demonstrated in some of the cells and in the material in the alveoli, which also includes denuded cells.

"Adenocarcinoma" is a cancer that originates in glandular tissue. This tissue is also part of a larger tissue category known as epithelial tissue. Epithelial tissue includes skin, glands and a variety of other tissue that lines the cavities and organs of the body. Epithelium is derived embryologically from ectoderm, endoderm and mesoderm. To be classified as adenocarcinoma, the cells do not necessarily need to be part of a gland, as long as they have secretory properties. This form of carcinoma can occur in some higher mammals, including humans. Well differentiated adenocarcinomas tend to resemble the glandular tissue that they are derived from, while poorly differentiated may not. By staining the cells from a biopsy, a pathologist will determine whether the tumor is an adenocarcinoma or some other type of cancer. Adenocarcinomas can arise in many tissues of the body due to the ubiquitous nature of glands within the body. While each gland may not be secreting the same substance, as long as there is an exocrine function to the cell, it is considered glandular and its malignant form is therefore named adenocarcinoma. Malignant adenocarcinomas invade other tissues and often metastasize given enough time to do so. Ovarian adenocarcinoma is the most common type of ovarian carcinoma. It includes the serous and mucinous adenocarcinomas, the clear cell adenocarcinoma and the endometrioid adenocarcinoma.

Renal cell carcinoma also known as renal cell cancer or renal cell adenocarcinoma is a kidney cancer that originates in the lining of the proximal convoluted tubule, the very small tubes in the kidney that filter the blood and remove waste products. Renal cell carcinoma is by far the most common type of kidney cancer in adults and the most lethal of all the genitorurinary tumors. Distinct subtypes of renal cell carcinoma are clear cell renal cell carcinoma and papillary renal cell carcinoma. Clear cell renal cell carcinoma is the most common form of renal cell carcinoma. When seen under a microscope, the cells that make up clear cell renal cell carcinoma appear very pale or clear. Papillary renal cell carcinoma is the second most common subtype. These cancers form little finger-like projections (called papillae) in some, if not most, of the tumors.

Lymphoma and leukemia are malignancies derived from hematopoietic (blood-forming) cells.

Blastic tumor or blastoma is a tumor (usually malignant) which resembles an immature or embryonic tissue. Many of these tumors are most common in children.

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.

In ovarian cancer, metastasis can occur in the following ways: by direct contact or extension, it can invade nearby tissue or organs located near or around the ovary, such as the fallopian tubes, uterus, bladder, rectum, etc.; by seeding or shedding into the abdominal cavity, which is the most common way ovarian cancer spreads. Cancer cells break off the surface of the ovarian mass and "drop" to other structures in the abdomen such as the liver, stomach, colon or diaphragm; by breaking loose from the ovarian mass, invading the lymphatic vessels and then traveling to other areas of the body or distant organs such as the lung or liver; by breaking loose from the ovarian mass, invading the blood system and traveling to other areas of the body or distant organs.

According to the invention, metastatic ovarian cancer includes cancer in the fallopian tubes, cancer in organs of the abdomen such as cancer in the bowel, cancer in the uterus, cancer in the bladder, cancer in the rectum, cancer in the liver, cancer in the stomach, cancer in the colon, cancer in the diaphragm, cancer in the lungs, cancer in the lining of the abdomen or pelvis (peritoneum), and cancer in the brain. Similarly, metastatic lung cancer refers to cancer that has spread from the lungs to distant and/or several sites in the body and includes cancer in the liver, cancer in the adrenal glands, cancer in the bones, and cancer in the brain.

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 al., 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 al., 2007: Nature 450, 1235-1239*). Another possibility for isolating CTCs is using the CellSearch^{TM} Circulating Tumor Cell (CTC) Test from Veridex, LLC (Raritan, NJ) which captures, identifies, and counts CTCs in a tube of blood. The CellSearch^{TM} 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 the composition 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 the composition 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 "tumor specimen" is a sample such as a bodily sample containing tumor or cancer cells such as circulating tumor cells (CTC), in particular a tissue sample, including body fluids, and/or a cellular sample. According to the invention, a "non-tumorigenous specimen" is a sample such as a bodily sample not containing tumor or cancer cells such as circulating tumor cells (CTC), in particular 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 therapeutically active agents, 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 of the invention 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 of the invention 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 of the invention 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 of the invention 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:

Top: Process to discover and prioritize likely immunogenic somatic mutations in bulk tumor samples. Bottom: Process as applied to the B16 and Black6 system.

Figure 2: Example Validated Mutation in Kif18b

A mutation identified in gene Kif18b by NGS exome-sequencing that was confirmed by Sanger sequencing. .In the wild type cells, the sequence is T/T. In the tumor cells, the sequence is a mix of T/G.

Figure 3: Immunologic reactivity against mutated sequences

Mice (n=5) were immunized twice (d0, d7) with mutated peptide sequences (100 µg + 50 µg PolyI:C; s.c.). At day 12 mice were sacrificed and the spleen cells harvested. IFNγ ELISpot was performed using $5x10^5$ spleen cells /well as effectors and $5x10^4$ bone marrow dendritic cells loaded with peptides (2 µg/ml for 2h at 37°C and 5% CO_2) as target cells. The effector spleen cells were tested against the mutated peptide, the wild type peptide and a control peptide (vesiculostomatitis virus nucleoprotein, VSV-NP, aa 52 - 59). Shown is the mean measured spot number from which the background spots against VSV-NP were subtracted for every mouse (empty circles: mice immunized with wildtype peptide; filled boxes: mice immunized with mutated peptides). Data are shown for each mouse and mean ± SEM is depicted.

Figure 4: Survival benefit for mice vaccinated with newly identified mutated peptide sequence

B16F10 cells (7,5 x 10^4)were inoculated subcutaneously on d0. Mice were vaccinated with peptide 30 (Jerini Peptide Technologies (Berlin); 100 µg peptide + 50 µg PolyI:C s.c. (Invivogen)) on day -4, day +2, day +9. The control group received only Poly I:C (50 µg s.c.). Tumor growth was monitored until day + 16 *, $p < 0,05$ in Log-rank (Mantel-Cox) test.

Figure 5:

(A) Examples of enhanced protein expression (*left* eGFP, *right* Luciferase) with RNA optimized for stability and translational efficiency (B) Example of polyepitopic expansion of antigen-specific CD8$^+$ and CD4$^+$ T cells with RNA optimized for effective antigen routing (s. Reference Kreiter, Konrad, Sester et al, Cancer Immunol. Immunother. 56: 1577-1587, 2007). T (C) Example of a preclinical proof of antitumoral efficacy in B16 melanoma model using an RNA vaccine that codes for a single epitope (OVA-SIINFEKL). Survival data were obtained for mice treated with vaccine alone or vaccine in combination with adjuvant. (D) Individualized, poly-neo-epitopic vaccine design. The vaccine vehicle integrates functional elements for increased expression and optimized immunogenicity. Up to 30 mutated epitopes that are spaced by linkers can be integrated per molecule in their natural sequence context.

Figure 6: Construct design

(A) Schematic diagram of a RNA polyepitope construct. Cap : cap analogon; 5´UTR : 5´untranslated region; L : linker; Seq. 1 : RNA sequence coding for peptide containing mutated aa; 3´UTR : 3´untranslated seuquence; poly-A : poly-A tail. (B) Sequence of the RNA constructs coding for 2 aa sequences including a mutated aa from B16F10. The start- and stop-codon as well as the signal peptide and the MITD sequence are not part of the schematic drawing which is symbolized by "....".

Figure 7: Functionality of RNA poly epitope

(A-C) Data for IFNγ ELISpot using 5 x 10^5 spleen cells per well as effectors and 5 x 10^4 BMDC as target cells. The BMDC were loaded with peptide (2 µg/ml for 2h at 37°C and 5% CO_2) or transfected with RNA (20 µg) by electroporation. The control RNA was eGFP (left panel) or a RNA construct coding for 2 unrelated peptides containing mutated aa separated by a linker. Data are shown as mean ± SEM. (A) Data for mutation peptide 30, wild type peptide 30 and RNA coding for mutation 30 and 31 are shown. (B) Data for mutation peptide 12, wild type peptide 12 and RNA coding for mutation 12 and 39 are shown. (C) Representative ELISpot scan from a single mouse of the read-out shown in (B) is depicted.

Figure 8: Two embodiments of RNA poly-neo-epitopic vaccines showing junction epitopes

The RNA vaccine can be constructed with (top) or without linkers (bottom) between mutation-encoding peptides. Good epitopes include those that include the somatic mutation ("*") and bind to MHC molecules. Bad epitopes include epitopes that bind to MHC molecules but contain either parts of two peptides (bottom) or parts of peptide and linker sequences (top).

Figure 9: Discovery and characterization of the "T-cell druggable mutanome"

(A) Flow chart gives an overview of the experimental procedure starting from B16F10 and C57BL/6 samples to ELISPOT readout. **(B)** The number of hits for each evaluation step and the process for selection of mutations for DNA validation and immunogenicity testing is shown. Mutations selected for validation and immunogenicity testing were those predicted to be immunogenic and in genes expressed at RPKM > 10. **(C)** The T-cell druggable mutanome was mapped to the genome of B16F10. Rings from outside to inside stand for following subsets: (1) present in all triplicates, (2) have an FDR < 0.05, (3) are located in protein coding regions , (4) cause nonsynonymous changes, (5) are locaized in expressed genes , and (6) are in the validated set. Mouse chromosomes (outer circle), gene density (green), gene expression (green(low)/yellow/red(high)), and somatic mutations (orange).

Figure 10: Immune responses elicited *in vivo* by vaccination of mice with mutation representing long synthetic peptides

(A,B) IFN-γ ELISPOT analysis of T-cell effectors from mice vaccinated with mutation coding peptides. Columns represent means (±SEM) of 5 mice per group. Asterisks indicate statistically significant differences of reactivity against mutation and wild-type peptide (student´s t-test; value $p < 0.05$). **(A)** Splenocytes of vaccinated mice were restimulated with BMDCs transfected with the mutation coding peptide used for vaccination, the corresponding wild-type peptide and an irrelevant control peptide (VSV-NP). **(B)** For analysis of T-cell reactivity against endogenously processed mutations splenocytes of vaccinated mice were restimulated with BMDCs transfected with control RNA (eGFP) or a RNA coding for the indicted mutation. **(C)** Mutation 30 (gene Kif18B, protein Q6PFD6, mutation p.K739N). Sanger sequencing trace and sequence of mutation (top). Protein domains and mutation location (bottom).

Figure 11: Antitumoral effects of mutated peptide vaccines in mice with aggressively growing B16F10 tumors

(A) C57BL/6 mice (n = 7) were inoculated with 7.5 x 10^4 B16F10 cells s.c. into the flank of the mice. On day 3 and 10 after tumor inoculation the mice were vaccinated with 100 μg MUT30 or MUT44 peptide + 50 μg poly(I:C) or with adjuvant alone. **(B)** C57BL/6 mice (n = 5) received one immunization of 100 μg MUT30 peptide + 50 μg poly(I:C) on day -4. On day

0 7.5 x 10^4 B16F10 cells were inoculated s.c. into the flank of the mice. Booster immunizations with MUT30 peptid (+ poly(I:C)) were done on days 2 and 9.
Kaplan-Meier survival Blot (left). Tumor growth kinetics (right).

Figure 12: Vaccination with mutation coding RNAs leads to CD4$^+$ and CD8$^+$ T-cell responses
Intracellular cytokine staining analysis data for IFN-γ in CD4$^+$ and CD8$^+$ T-cell effectors from mice vaccinated with mutation coding RNAs. RNAs were coding for 1 (Monoepitope, upper row), 2 (Biepitope, middle row), or 16 (Polyepitope, lower row) different mutations. Dots represent means of 3 mice per group. Asterisks indicate statistically significant differences of reactivity against mutation and control peptide (VSV-NP) (student's t-test; value $p < 0.05$). FACS plots show effectors from the highest IFN-γ secreting animal for each mutation and indicate phenotype of the T-cell response.

Figure 13: Vaccination with mutation coding Polyepitope RNA leads T-cell reponses against several mutations
IFN-γ ELISPOT analysis of T-cell effectors from mice vaccinated with mutation coding Polyepitope including 16 different mutations. Columns represent means (±SEM) of 3 mice per group. Photograph shows triplicate wells of cells from one exemplary animal restimulated with the indicated peptides.

Figure 14: Vaccination with 5 different model epitopes encoded by one RNA leads to immune responses against all encoded epitopes
A) IFN-γ ELISPOT analysis of T-cell effectors from mice vaccinated with mutation coding model Polyepitope including 5 different model epitopes (SIINFEKL, Trp2, VSV-NP, Inf-NP, OVA class II). Splenocytes were restimulated with the indicated peptides. Spots represent means of triplicate wells from 5 mice per group. B) Pentamer staining of blood lymphocytes of one control mouse and one mouse immunized with the model Polyepitope. Inf-NP Pentamer stained CD8$^+$ cells are specific for the Inf-NP peptide.

Figure 15: A CD4$^+$ T-cell inducing mutation can induce a potent anti-tumoral effect B16F10 melanoma in synergy with a weak CD8$^+$ T-cell epitope
C57BL/6 mice (n = 8) were inoculated with 1 x 10^5 B16F10 cells s.c. into the flank of the mice. On day 3, 10 and 17 after tumor inoculation the mice were vaccinated with 100 μg

MUT30, Trp2 or both peptides + 50 µg poly(I:C). A) Shown are the mean tumor growth kinetics of each group. On day 28 the mean values between the single treatment groups and the untreated animals and the combination group are statistically different (Mann-Whitney test, p-value < 0.05). B) Kaplan-Meyer survival plot of the different groups. The survival curves of MUT30 and MUT30 + Trp2 vaccinated mice are statistically different (Log-Rank test, p-value = 0.0029).

Figure 16: Overview of process for finding somatic mutations in B16

Numbers for the individual steps are given as an example for one B16 sample, compared to one black6 sample. "Exons" refers to the exon coordinates defined by all protein coding RefSeq transcripts.

Figure 17: Venn diagramm showing the numbers of somatic variations in protein coding exons, found by the individual, two or all three software tools, respectively

The numbers were calculated after filtering and represent the consensus of all three samples.

Figure 18: A Examples of single nucleotide variations found: A somatic mutation found in all three B16 samples (left), a non-somatic mutation found in all B16 and black6 samples (middle) and a mutation found in only one black6 sample (right). **B** The calculated FDR distribution for the dataset of which the validated mutations were selected; the distribution is visualized as an average estimated ROC curve with the grey bars giving the 95% confidence interval for the mean in both dimensions at uniformly sampled positions. The mean was obtained from the distribution of estimated ROC curves of the FDRs for all possible 18 combinations (see text).

Figure 19: A Estimated ROC curves for the comparison of the three different software tools (duplicates, 38x coverage). **B** Estimated ROC curves for the comparison of different average sequencing depths (samtools, no replication). 38x denotes the coverage obtained by the experiment, while other coverages were downsampled starting with this data. **C** Estimated ROC curves visualizing the effect of experiment replication (38x coverage, samtools). **D** Estimated ROC curves for different sequencing protocols (samtools, no replication). The curves were calculated using the results of the 2x100 nt library.

Figure 20: **A** Ten validated mutations with the lowest FDRs, selected using the optimal set of parameters out of a final set of 2396 variations. None of these mutations is present in dbSNP (version 128; genome assembly mm9). **B** Relative amount of variations found in the same dataset as A for a given FDR cutoff, plotted separately for all variants in the dataset and the validated mutations. For visual clarity only values of 0 to 10% FDR are shown.

Figure 21: Antitumoral activity of a mutation-encoding polyepitope RNA vaccine C57BL/6 mice (n = 10) were inoculated with 1 x 10^5 B16F10 cells s.c. into the flank of the mice. On day 3, 6, 10, 17 and 21 after tumor inoculation the mice were vaccinated with a polytope RNA formulated a liposomal RNA transfection reagent. The control group received liposomes without RNA. The figure shows the Kaplan-Meyer survival plot of the different groups. The survival curves statistically different (Log-Rank test, p-value = 0.0008).

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: Mutation detection and prioritization

We first demonstrate sequence profiling of tumor and normal samples to identify somatic mutations in an unbiased manner. We demonstrate this not only for bulk tumor samples but also, for the first time, demonstrate the ability to identify mutations from individual circulating tumor cells. Next, we prioritize the mutations for inclusion in a poly-neo-epitopic vaccine based on the predicted immunogenicity of the mutation and demonstrate that the identified mutations are indeed immunogenic.

Mutation detection

The rationale for using CTCs: the detection of circulating tumor cells (CTC) from the peripheral blood of cancer patients is a recognized independent prognostic marker for the clinical course of tumors (Pantel et al, Trends Mol Med 2010; 16(9):398-406). For many years, the clinical significance of CTCs has been the subject of intense scientific and clinical research in oncology. It has been shown that the detection of CTCs in the blood of patients with metastatic breast, prostate and colorectal cancer has prognostic relevance, providing additional information to conventional imaging techniques and other prognostic tumor biomarkers. Sequential blood samples drawn from a patient before, during an early stage, and after treatment with a therapeutic agent (systemic or targeted) provides information on treatment response/failure. The molecular analysis of drug-resistant CTCs may provide a further insight into resistance mechanisms (e.g. mutations in specific signaling pathways or loss of target expression) in individual patients. An additional possibility from the profiling and genetic characterization of CTCs is the identification of novel cancer targets for the development of new targeted therapies. This new diagnostic strategy is referred to as "Liquid Tumor Biopsy." As this profiling could be quickly and repetitively done, requiring only patient blood and no surgery, this would provide a "real time" view of the tumor state.

Mutations from tumor cells: We demonstrate our ability to identify mutations using B16 melanoma cells, exome capture to extract protein coding regions, next-generation sequencing using our HiSeq 2000, followed by bioinformatics analysis using our "iCAM" software pipeline (Figure 1). We identify 2448 non-synonymous mutations and selected 50 for confirmation. We were able to confirm all 50 somatic mutations.

The following is an example of the protein impact of a discovered somatic mutation in B16 melanoma cells:

```
Kif18b, NM_197959, exon 3

Mutation (+15 aa)
SPSKPSFQEFVDWENVSPELNSTDQPFLPS
Wild type (+15 aa)
SPSKPSFQEFVDWEKVSPELNSTDQPFLPS
```

Mutations from individual circulating tumor cells (CTCs): Next, we were able to identify tumor-specific somatic mutations from NGS profiling of RNA from single CTCs. Labeled B16 melanoma cells were intravenously injected into mouse tails, mice were sacrificed, blood was collected from hearts, cells sorted to retrieve labeled circulating B16 cells (CTCs), RNA extracted, a SMART-based cDNA synthesis and unspecific amplification performed, followed by the NGS RNA-Seq assay and subsequence data analysis (below).

We profiled eight individual CTCs and identified somatic mutations. Furthermore, in eight of eight cells, previously identified somatic mutations were identified. In multiple cases, the data showed heterogeneity at the individual cell level. For example, at position 144078227 on chromosome 2 (assembly mm9), in gene Snx15, two cells showed the reference nucleotide (C) while two cells showed the mutated nucleotide (T).

This demonstrates that we are able to profile individual CTCs to identify somatic mutations, a fundamental path to a "real-time" iVAC (individualized vaccine), in which patients are profiled repetitively and the results reflect the current patient status rather than the status at an earlier time point. Furthermore, this demonstrates that we are able to identify heterogeneous somatic mutations that are present in a subset of tumor cells, enabling evaluation of mutation frequency, such as for identification of major mutations and rare mutations.

Methods

Samples: For the profiling experiment, samples included 5-10mm tail samples from C57BL/6 mice ("Black6") and highly aggressive B16F10 murine melanoma cells ("B16"), which are originally derived from Black6 mice.

Circulating tumor cells (CTCs) were created using fluorescent labeled B16 melanoma cells. B16 cells were resuspended in PBS and an equal volume of freshly prepared CFSE-Solution (5 µM in PBS) was added to the cells. The sample was gentle mixed by vortex followed by incubation for 10 min at room temperature. To stop the labeling reaction, the equal amount of PBS containing 20% FSC was added to the sample and mixed gently by vortex. Following 20 min incubation at room temperature, the cells were washed twice using PBS. Finally, the cells were resuspended in PBS and injected intravenously (i.v.) in mice. After 3 minutes the mice were sacrificed and blood collected.

Erythrocytes from the blood samples were lysed by adding 1,5 ml fresh prepared PharmLyse Solution (Beckton Dickinson) per 100 µl blood. After one washing step, 7-AAD was added to the sample and incubated for 5 min at room temperature. The incubation was followed by two washing steps and the sample was resuspended in 500 µl PBS.

The CFSE labeled circulating B16 cells were sorted with an Aria I cells-sorter (BD). Single cells were sorted on 96-well-v-bottem plated prepared with 50 µl/well RLT buffer (Quiagen). After finishing the sorting the plates were stored at -80°C until the Nucleic acid extraction and sample preparation started.

Nucleic acid extraction and sample preparation: nucleic acids from B16 cells (DNA and RNA) and Black6 tail tissue (DNA) were extracted using Qiagen DNeasy Blood and Tissue kit (DNA) and Qiagen RNeasy Micro kit (RNA).

For individual sorted CTCs, RNA was extracted and a SMART-based cDNA synthesis and unspecific amplification performed. RNA from sorted CTC cells was extracted with the RNeasy Micro Kit (Qiagen, Hilden, Germany) according to the instructions of the supplier. A modified BD SMART protocol was used for cDNA synthesis: Mint Reverse Transcriptase (Evrogen, Moscow, Russia) was combined with oligo(dT)-T-primer long for priming of the first-strand synthesis reaction and TS-short (Eurogentec S.A., Seraing, Belgium) introducing an oligo(riboG) sequence to allow for creation of an extended template by the terminal transferase activity of the reverse transcriptase and for template switch [*Chenchik, A., Y. et al. 1998. Generation and use of high quality cDNA from small amounts of total RNA by SMART PCR. In Gene Cloning and Analysis by RT-PCR. P. L. J. Siebert, ed. BioTechniques Books, MA, Natick. 305-319*]. First strand cDNA synthesized according to the manufacturer's instructions was subjected to 35 cycles of amplification with 5 U PfuUltra Hotstart High-Fidelity DNA Polymerase (Stratagene, La Jolla, CA) and 0.48 µM primer TS-PCR primer in the presence of 200 µM dNTP (cycling conditions: 2 min at 95 °C for, 30 s at 94 °C, 30 s at 65 °C, 1 min at 72 °C for, final extension of 6 min at 72 °C). Successful amplification of the CTC genes was controlled with specific primers to monitor actin and GAPDH.

Next-generation sequencing, DNA sequencing: Exome capture for DNA resequencing was performed using the Agilent Sure-Select solution-based capture assay [*Gnirke A et al: Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat Biotechnol 2009, 27:182-189*], in this case designed to capture all mouse protein coding regions.

Shortly, 3 ug purified genomic DNA was fragmented to 150-200 bp's using a Covaris S2 ultrasound device. gDNA fragments were end repaired using T4 DNA polymerase, Klenow DNA polymerase and 5' phosphorylated using T4 polynucleotide kinase. Blunt ended gDNA fragments were 3' adenylated using Klenow fragment (3' to 5' exo minus). 3' single T-overhang Illumina paired end adapters were ligated to the gDNA fragments using a 10:1 molar ratio of adapter to genomic DNA insert using T4 DNA ligase. Adapter ligated gDNA fragments were enriched pre capture and flow cell specific sequences were added using Illumina PE PCR primers 1.0 and 2.0 and Herculase II polymerase (Agilent) using 4 PCR cycles.

500 ng of adapter ligated, PCR enriched gDNA fragments were hybridized to Agilent's SureSelect biotinylated mouse whole exome RNA library baits for 24 hrs at 65 °C. Hybridized gDNA/RNA bait complexes where removed using streptavidin coated magnetic beads. gDNA/RNA bait complexes were washed and the RNA baits cleaved off during elution in SureSelect elution buffer leaving the captured adapter ligated, PCR enriched gDNA fragments. gDNA fragments were PCR amplified post capture using Herculase II DNA polymerase (Agilent) and SureSelect GA PCR Primers for 10 cycles.

All cleanups were done using 1,8x volume of AMPure XP magnetic beads(Agencourt) 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 and 50 bps were sequenced on the Illumina HiSeq2000 using Truseq SBS kit-HS 50 bp.

Next-generation sequencing, RNA sequencing (RNA-Seq): Barcoded mRNA-seq cDNA libraries were prepared from 5 ug of total RNA using a modified version of the Illumina mRNA-seq protocol. mRNA was isolated using Seramag Oligo(dT) magnetic beads (Thermo Scientific). Isolated mRNA was fragmented using divalent cations and heat resulting in fragments ranging from 160-220 bp. Fragmented mRNA was converted to cDNA using random primers and SuperScriptII (Invitrogen) followed by second strand synthesis using DNA polymerase I and RNaseH. cDNA was end repaired using T4 DNA polymerase, Klenow DNA polymerase and 5' phosphorylated using T4 polynucleotide kinase. Blunt ended cDNA fragments were 3' adenylated using Klenow fragment (3' to 5' exo minus). 3' single T-overhang Illumina multiplex specific adapters were ligated using a 10:1 molar ratio of adapter to cDNA insert using T4 DNA ligase.

cDNA libraries were purified and size selected at 200-220 bp using the 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 (Finnzymes). All cleanups were done using 1,8x volume of AgencourtAMPure 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 on the cBot using Truseq SR cluster kit v2.5 using 7 pM and 50 bps were sequenced on the Illumina HiSeq2000 using Truseq SBS kit-HS 50 bp.

CTCs: For the RNA-Seq profiling of CTCs, a modified version of this protocol was used in which 500-700 ng SMART amplified cDNA was used, paired end adapters were ligated and PCR enrichment was done using Illumina PE PCR primers 1.0 and 2.0.

NGS data analysis, gene expression: To determine expression values, the output sequence reads from RNA samples from the Illumina HiSeq 2000 were preprocessed according to the Illumina standard protocol. This includes filtering for low quality reads and demultiplexing. For RNA-Seq transcriptome analysis, sequence reads were aligned to the reference genomic sequence [*Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome. Nature, 420, 520-562 (2002)*] using bowtie (version 0.12.5) [*Langmead B. et al. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25*] using parameters "-v2 –best" for genome alignments and default parameters for transcript alignments. The alignment coordinates were compared to the exon coordinates of the RefSeq transcripts [*Pruitt KD. et al. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res. 2005 Jan 1;33(Database issue):D501-4*] and for each transcript the counts of overlapping alignments were recorded. Sequence reads not alignable to the genomic sequence were aligned to a database of all possible exon-exon junction sequences of the RefSeq transcripts. The counts of reads aligning to the splice junctions were aggregated with the respective transcript counts obtained in the previous step and normalized to RPKM (number of reads which map per kilobase of exon model per million mapped reads [*Mortazavi, A. et al. (2008). Mapping and quantifying mammalian transcriptomes by rna-seq. Nat Methods, 5(7):621-628*]) for each transcript. Both gene expression and exon expression values were calculated based on the normalized number of reads overlapping each gene or exon, respectively.

Mutation discovery, bulk tumor: 50 nt, single end, reads from the Illumina HiSeq 2000 were aligned using bwa (version 0.5.8c) [*Li H. and Durbin R. (2009) Fast and accurate short read alignment with Burrows-Wheeler Transform. Bioinformatics, 25:1754-60*] using default options to the reference mouse genome assembly mm9. Ambiguous reads – those reads mapping to multiple locations of the genome - were removed, the remaining alignments were sorted, indexed and converted to a binary and compressed format (BAM) and the read quality scores converted from the Illumina standard phred+64 to standard Sanger quality scores using shell scripts.

For each sequencing lane, mutations were identified using three software programs: including samtools (version 0.1.8) [*Li H. Improving SNP discovery by base alignment quality. Bioinformatics. 2011 Apr 15;27(8):1157-8. Epub 2011 Feb 13*], GATK (version 1.0.4418) [*McKenna A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010 Sep;20(9):1297-303. Epub 2010 Jul 19*], and SomaticSniper . For samtools, the author-recommend options and filter criteria were used, including first round filtering, maximum coverage 200. For samtools second round filtering, the minimum indel qualtify score was 50, the point mutation minimum quality was 30. For GATK mutation calling, we followed the author-designed best practice guidelines presented on the GATK user manual. The variant score recalibration step was omitted and replaced by the hard-filtering option. For SomaticSniper mutation calling, the default options were used and only predicted mutations with a "somatic score" of 30 or more were considered further.

Mutation discovery, CTCs: As per the bulk tumor iCAM process, 50 nt, single end, reads from the Illumina HiSeq 2000 were aligned using bwa (version 0.5.8c) [5]) using default options to the reference mouse genome assembly mm9. As CTC NGS reads were derived from the RNA-Seq assay, reads were also aligned to transcriptome sequences, including exon-exon junctions, using bowtie (above). Using all alignments, the nucleotide sequences from the reads were compared to both the reference genome and the bulk-tumor derived B16 mutations. Identified mutations were evaluated both using perl scripts and manually using the software program samtools and the IGV (Integrated Genome Viewer) to image the results.

The output of "mutation discovery" is the identification of somatic mutations in tumor cells, from sample to NGS data to a list of mutations. In the B16 samples, we identified 2448 somatic mutations using exome resequencing.

Mutation prioritization

Next, we demonstrate a possibility of a mutation prioritization pipeline for vaccine inclusion. This method, called "individual cancer mutation detection pipeline" (iCAM) identifies and prioritizes somatic mutations through a series of steps incorporating multiple cutting edge algorithms and bioinformatics methods. The output of this process is a list of somatic mutations, prioritized based on likely immunogenicity.

Somatic mutation identification: Mutations are identified using three different algorithms, for both the B16 and Black6 samples (Mutation discovery, above). The first iCAM step is to combine the output lists from each algorithm to generate a high-confidence list of somatic mutations. GATK and samtools report variants in one sample relative to a reference genome. To select high confidence mutations with few false-positives for a given sample (i.e., tumor or normal), mutations are selected that are identified in all replicates. Then, variants are selected which are present in the tumor sample but not present in the normal sample. SomaticSniper automatically reports potential somatic variations from tumor and normal data pairs. We further filtered results through the intersection of the results obtained from replicates. To remove as many false positive calls as possible, we intersected the list of mutations derived from the use of all three algorithms and all replicates. The final step for each somatic mutation is to assign a confidence value (p-value) for each mutation based on coverage depth, SNP quality, consensus quality and mapping quality.

Mutation impact: the impact of the filtered, consensus, somatic mutations is determined by a script within the iCaM mutation pipeline. First, mutations that occur in genomic regions that are not unique within the genome, such as occur for some protein paralogs and pseudogenes, are excluded from analysis as sequence reads that align to multiple locations are removed. Second, whether the mutation occurs in a transcript is determined. Third, whether the mutation occurs in a protein-coding region is determined. Fourth, the transcript sequence is translated with and without the mutation to determine if there is a change in amino acid sequence.

Mutation expression: the iCAM pipeline selects somatic mutations that are found in genes and exons that are expressed in tumor cells. Expression levels are determined through NGS RNA-Seq of tumor cells (above). The number of reads that overlap a gene and an exon indicates expression levels. These counts are normalized to RPKM (Reads Per Kilobase of exon model per Million mapped reads, [*Mortazavi A. et al. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods. 2008 Jul;5(7):621-8. Epub 2008 May 30*]) and those expressed above 10 RPKM are selected.

MHC binding: to determine the likelihood that an epitope containing the mutated peptide is binds to an MHC molecule, the iCAM pipeline runs a modified version of the MHC prediction software from the Immune Epitope Database . The local installation includes modifications to optimize data flow through the algorithm. For the B16 and Black6 data, the prediction was run using all available black6 MHC class I alleles and all epitopes for the respective peptide lengths. Mutations are selected which fall in an epitope ranked in the 95th percentile of the prediction score distribution of the IEDB training data , considering all MHC alleles and all potential epitopes overlapping the mutation.

Mutation selection criteria: somatic mutations are selected by the following criteria: a) have unique sequence content, b) identified by all three programs, c) high mutation confidence, d) non-synonymous protein change, e) high transcript expression, f) and favorable MHC class I binding prediction.

The output of this process is a list of somatic mutations, prioritized based on likely immunogenicity. In B16 melanoma cells, there are 2448 somatic mutations. 1247 of these mutations are found in gene transcripts. Of these, 734 cause non-synonymous protein changes. Of these, 149 are in genes expressed in the tumor cells. Of these, 102 of these expressed, non-synonymous mutations are predicted to be presented on MHC molecules. These 102 likely immunogenic mutations are then passed to mutation confirmation (below).

Mutation confirmation

Somatic mutations from DNA exome-resequencing were confirmed by either of two methods, resequencing of the mutated region and RNA-Seq analysis.

For the confirmation of the mutations by resequencing, a genomic region containing the mutation was amplified by standard PCR from 50 ng of both the tumor DNA and the normal control DNA. The size of the amplified products was in the range of 150 to 400 nt. The specificity of the reaction was controlled by loading the PCR product on the Qiaxel device (Qiagen). PCR products were purified using the minElute PCR purification kit (Qiagen). Specific PCR products were sequenced using the standard Sanger sequencing method (Eurofins), followed by electropherogram analysis.

Mutation confirmation was also accomplished through examination of tumor RNA. Tumor gene and exon expression values were generated from RNA-Seq (NGS of RNA), which generates nucleotide sequences that were mapped to transcripts and counted. We examined sequence data itself to identify mutations in the tumor sample [*Berger MF. et al. Integrative analysis of the melanoma transcriptome. Genome Res. 2010 Apr;20(4):413-27. Epub 2010 Feb 23*], providing an independent confirmation of the DNA-derived identified somatic mutations.

Table 1: List of genes containing the 50 validated mutations

Genes containing the 50 identified and confirmed somatic mutations, with annotation regarding gene symbol, gene name, and predicted localization and function.

ID	Symbol	Entrez Gene Name	Location
NM_021895	ACTN4	actinin, alpha 4	Cytoplasm
NM_028840	ARMC1	armadillo repeat containing 1	unknown
NM_029291	ASCC2	activating signal cointegrator 1 complex subunit 2	unknown
NM_024184	ASF1B	ASF1 anti-silencing function 1 homolog B (S. cerevisiae)	Nucleus
NM_138679	ASH1L	ash1 (absent, small, or homeotic)-like (Drosophila)	Nucleus
NM_015804	ATP11A	ATPase, class VI, type 11A	Plasma Membrane
NM_009730	ATRN	attractin	Extracellular Space
NM_028020	CPSF3L	cleavage and polyadenylation specific factor 3-like	Nucleus
NM_010017	DAG1	dystroglycan 1 (dystrophin-associated glycoprotein 1)	Plasma Membrane
NM_015735	DDB1	damage-specific DNA binding protein 1, 127kDa	Nucleus
NM_001080981	DDX23	DEAD (Asp-Glu-Ala-Asp) box polypeptide 23	Nucleus
NM_054046	DEF8	differentially expressed in FDCP 8 homolog (mouse)	unknown
NM_019965	DNAJB12	DnaJ (Hsp40) homolog, subfamily B, member 12	Cytoplasm
NM_011262	DPF2	D4, zinc and double PHD fingers family 2	Nucleus
NM_007907	EEF2	eukaryotic translation elongation factor 2	Cytoplasm
NM_001081286	FAT1	FAT tumor suppressor homolog 1 (Drosophila)	Plasma Membrane
NM_173182	FNDC3B	fibronectin type III domain containing 3B	unknown
NM_008057	FZD7	frizzled homolog 7 (Drosophila)	Plasma Membrane
NM_201617	GNAS	GNAS complex locus	Plasma Membrane
NM_030035	GOLGB1	golgin B1	Cytoplasm
NM_011365	ITSN2	intersectin 2	Cytoplasm
NM_029841	KIAA2013	KIAA2013	unknown
NM_197959	KIF18B	kinesin family member 18B	unknown
NM_145479	KLHL22	kelch-like 22 (Drosophila)	unknown
NM_018810	MKRN1	makorin ring finger protein 1	unknown
NM_001170785	MTHFD1L	methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 1-like	Cytoplasm
NM_133947	NUMA1	nuclear mitotic apparatus protein 1	Nucleus
NM_178884	OBSL1	obscurin-like 1	unknown
NM_008765	ORC2	origin recognition complex, subunit 2	Nucleus
NM_023209	PBK	PDZ binding kinase	Cytoplasm
NM_033594	PCDHGA11	protocadherin gamma subfamily A, 11	Plasma Membrane
NM_025951	PI4K2B	phosphatidylinositol 4-kinase type 2 beta	Cytoplasm
NM_011961	PLOD2	procollagen-lysine, 2-oxoglutarate 5-dioxygenase 2	Cytoplasm
NM_023200	PPP1R7	protein phosphatase 1, regulatory (inhibitor) subunit 7	Nucleus
NM_008986	PTRF	polymerase I and transcript release factor	Nucleus
NM_011240	RANBP2	RAN binding protein 2	Nucleus
NM_009438	RPL13A	ribosomal protein L13a	Cytoplasm
NM_009113	S100A13	S100 calcium binding protein A13	Cytoplasm
NM_001081203	SBNO1	strawberry notch homolog 1 (Drosophila)	unknown
NM_009153	SEMA3B	sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3B	Extracellular Space
NM_026912	SNX15	sorting nexin 15	Cytoplasm
NM_024225	SNX5	sorting nexin 5	Cytoplasm
NM_008188	THUMPD3	THUMP domain containing 3	unknown
NM_133352	TM9SF3	transmembrane 9 superfamily member 3	Cytoplasm
NM_177296	TNPO3	transportin 3	Cytoplasm
NM_011640	TP53	tumor protein p53	Nucleus
NM_023279	TUBB3	tubulin, beta 3	Cytoplasm
NM_029896	WDR82	WD repeat domain 82	unknown
NM_025830	WWP2	WW domain containing E3 ubiquitin protein ligase 2	Cytoplasm
NM_001081056	XPOT	exportin, tRNA (nuclear export receptor for tRNAs)	Nucleus

Example 2: IVAC selection algorithm enables the detection of immunogenic mutations

To investigate if specific T-cell responses could be induced against the confirmed mutations from B16F10 melanoma cells, naïve C57BL/6 mice (n=5/peptide) were immunized twice (d0, d7) subcutaneously with 100 μg peptide (+ 50 μg PolyI:C as adjuvant) comprising either the mutated or the wild type aa sequence (see Table 2). All peptides had a length of 27 aa with the mutated/wild type aa at the central position. At day 12 mice were sacrificed and the spleen cells were harvested. As read-out method IFNγ ELISpot was performed using $5x10^5$ spleen cells/well as effectors and $5x10^4$ bone marrow dendritic cells loaded with peptides (2 μg/ml) as target cells. The effector spleen cells were tested against the mutated peptide, the wild type peptide and a control peptide (vesiculostomatitis virus nucleoprotein, VSV-NP).

With 44 sequences tested we observed that 6 of them induced a T-cell immunity directed against the mutated sequence only but not to the wild type peptide (Fig. 3).

The data prove that the identified and prioritized mutations can be utilized to induce tumor specific T-cell immunity after being utilized as peptide vaccine in antigen naïve mice.

Table 2: Listing of mutated sequences that induced a T-cell reactivity specific for the mutated versus the wild type peptide. The amino acid exchange is marked underlined.

Number	RefSeq ID	Sequence Type	Peptide Sequence	T-cell reactivity (mice)
12	NM_00107750, NM_010309, NM_201618, NM_201617	Mutated	TPPPEEAMPFEFNGPAQGDHSQPPLQV	5/5
		Wild Type	TPPPEEAMPFEFNEPAQGDHSQPPLQV	0/5
16	NM_008188	Mutated	RVTCNRAGEKHCFSSNEAARDFGGAIQ	3/5
		Wild Type	RVTCNRAGEKHCFTSNEAARDFGGAIQ	0/5
20	NM_023279	Mutated	FRRKAFLHWYTGEAMDEMEFTEAESNM	5/5
		Wild Type	FRRKAFLHWYTGEGMDEMEFTEAESNM	1/5
30	NM_197959	Mutated	PSKPSFQEFVDWENVSPELNSTDQPFL	5/5
		Wild Type	PSKPSFQEFVDWEKVSPELNSTDQPFL	1/5
34	NM_145479	Mutated	HLTQQLDTYILKNVVAFSRTDKYRQLP	3/5
		Wild Type	HLTQQLDTYILKNFVAFSRTDKYRQLP	0/5
36	NM_133352	Mutated	CGTAFFINFIAIYHHASRAIPFGTMVA	5/5
		Wild Type	CGTAFFINFIAIYYHASRAIPFGTMVA	0/5

Example 3: Identified mutations can provide therapeutic anti-tumor immunity

In order to validate whether the identified mutations have the potential to confer anti-tumor immunity after vaccination to naïve mice we investigated this question with the peptide for mutation number 30 that was shown to induce a mutation selective T-cell reactivity. B16F10 cells (7,5 x 10^4) were inoculated subcutaneously on d0. Mice were vaccinated with peptide 30 (see table 1; 100 µg peptide + 50 µg PolyI:C s.c.) on day -4, day +2, and day +9. The control group received only Poly I:C (50 µg s.c.). Tumor growth was monitored every other day. At day +16 we observed that only 1out of 5 mice in the peptide vaccine group had developed a tumor whereas in the control group 4 out of 5 mice showed tumor growth.

The data prove that a peptide sequence incorporating a B16F10 specific mutation can confer anti tumor immunity that is efficiently able to destroy tumor cells (see Fig. 4). Since B16F10 is a highly aggressive tumor cell line the finding that the methodology applied to identify and prioritize mutations finally led to the selection of a mutation that by itself already is potent as a vaccine is an important proof of concept for the whole process.

Example 4: Data supporting polyepitopic antigen presentation

Validated mutations from protein-coding regions of a patient constitute the pool from which candidates can be selected for assembly of the poly-neo-epitope vaccine template to be used as precursor for GMP manufacturing of the RNA vaccine. Suitable vector cassettes as vaccine backbone has been already described (Holtkamp, S. et al., Blood, *108*: 4009-4017, 2006; Kreiter, S. et al., Cancer Immunol. Immunother., *56*: 1577-1587, 2007; Kreiter, S. et al., J.Immunol., *180*: 309-318, 2008). The preferred vector cassettes are modified in their coding and untranslated regions (UTR) and ensure maximized translation of the encoded protein for extended periods (Holtkamp, S. et al., Blood, *108*: 4009-4017, 2006; Kuhn, A. N. et al., Gene Ther., *17*: 961-971, 2010). Furthermore, the vector backbone contains antigen routing modules for the simultaneous expansion of cytotoxic as well as helper T-cells (Kreiter, S. et al., Cancer Immunol. Immunother., *56*: 1577-1587, 2007; Kreiter, S. et al., J. Immunol., *180*: 309-318, 2008; Kreiter, S. et al., Cancer Research, 70 (22), 9031-9040, 2010 (Figure 5). Importantly, we have proven that such RNA vaccine can be used to present multiple MHC class I and class II epitopes simultaneously.

The IVAC poly-neo-epitope RNA vaccine sequences are built from stretches of up to 30 amino acids that include the mutation in the center. These sequences are connected head-to-tail via short linkers to form a poly-neo-epitope vaccine coding for up to 30 or more selected mutations and their flanking regions. These patient-specific individually tailored inserts are codon-optimized and cloned into the RNA backbone described above. Quality control of such constructs includes *in vitro* transcription and expression in cells for validation of functional transcription and translation. Analysis of translation will be performed with antibodies against the c-terminal targeting domain.

Example 5: Scientific proof of concept for the RNA poly-neo epitope construct

The RNA poly-neo epitope concept is based on a long in vitro transcribed mRNA which consists of sequentially arranged sequences coding for the mutated peptides connected by linker sequences (see Fig. 6). The coding sequences are chosen from the non synonymous mutations and are always built up of the codon for the mutated amino acid flanked by regions of 30 to 75 base-pairs from the original sequence context. The linker sequence codes for amino acids that are preferentially not proccssed by the cellular antigen processing machinery. *In vitro* transcription constructs are based on the pST1-A120 vector containing a T7 promotor, a tandem beta-globin 3' UTR sequence and a 120-bp poly(A) tail, which have been shown to increase the stability and translational efficiency of the RNA thereby enhancing the T-cell stimulatory capacity of the encoded antigen (Holtkamp S. et al., Blood 2006; PMID: 16940422). In addition, an MHC class I signal peptide fragment and the transmembrane and cytosolic domains including the stop-codon (MHC class I trafficking signal or MITD) flanking a poly-linker sequence for cloning the epitopes were inserted (Kreiter S. et al., J. Immunol., 180: 309-318, 2008). The latter have been shown to increase the antigen presentation, thereby enhancing the expansion of antigen-specific CD8+ and CD4+ T cells and improving effector functions.

For a first proof of concept, biepitopic vectors were used, i.e. encoding one polypeptide containing two mutated epitopes. Codon optimized sequences coding for (i) a mutated epitope of 20 to 50 amino acids, (ii) a glycine/serine-rich linker, (iii) a second mutated epitope of 20 to 50 amino acids, and (iv) an additional glycine/serine-rich linker – flanked by suitable recognition sites for restriction endonucleases to be cloned into the pST1-based construct as described above – were designed and synthesized by a commercial provider (Geneart, Regensburg, Germany). After verification of the sequence, these were cloned into the pST1-based vector backbone to obtain constructs as depicted in Figure 6.

The pST1-A120-based plasmids as described above were linearized with a class IIs restriction endonuclease. The linearized plasmid DNAs were purified by phenol chloroform extraction and ethanol precipitation. Linearized vector DNAs were quantified spectrophotometrically and subjected to *in vitro* transcription essentially as described by Pokrovskaya and Gurevich (1994, Anal. Biochem. 220: 420-423). A cap analog has been added to the transcription reaction to obtain RNAs with the correspondingly modified 5'-cap structures. In the reactions, GTP was present at 1.5 mM, while the cap-analog was present at 6.0 mM. All other NTPs were present at 7.5 mM. At the end of the transcription reaction, linearized vector DNA was digested with 0.1 U/µl TURBO DNase (Ambion, Austin/TX, USA) for 15 minutes at 37°C. RNAs were purified from these reactions using the MEGAclear Kit (Ambion, Austin/TX, USA) as per manufacturer's protocol. RNA concentration and quality were assessed by spectrophotometry and analysis on a 2100 Bioanalyzer (Agilent, Santa Clara, CA, USA).

In order to proof that a sequence incorporating a mutated amino acid and being 5′- as well as 3′-flanked by the linker sequence can be processed, presented and recognized by antigen specific T-cells we used T-cells from peptide vaccinated mice as effector cells. In an IFNγ ELISpot we tested whether the T-cells induced by peptide vaccination as described above are capable of recognizing the target cells (bone marrow dendritic cells, BMDC) either pulsed with peptide (2 µg/ml for 2h at 37°C and 5% CO_2) or transfected with RNA (20 µg produced as described above) by electroporation. As exemplified in Fig. 7 for mutation 12 and 30 (see table 2) we could observe that the RNA construct is able to give rise to the epitope recognized by mutation specific T-cells.

With the data provided we could demonstrate that an RNA encoded poly-neo epitope including glycine / serine rich linker can be translated and processed in antigen presenting cells leading to presentation of the correct epitope that is recognized by the antigen specific T-cells.

Example 6: Poly-neo-epitope vaccine design – The relevance of the linker

The poly-neo-epitope RNA construct contains a backbone construct into which multiple somatic mutation-encoding peptides connected with a linker peptide sequence are placed. In addition to codon optimization and increased RNA stability and translational efficiency due to the backbone, one embodiment of the RNA poly-neo-epitope vaccine contains linkers designed to increase MHC Class I and II presentation of antigenic peptides and decrease presentation of deleterious epitopes.

Linker: the linker sequence was designed to connect multiple mutation-containing peptides. The linker should enable creation and presentation of the mutation epitope while hinder creation of deleterious epitopes, such as those created at the junction suture between adjacent peptides or between linker sequence and endogenous peptides. These "junction" epitopes may not only compete with the intended epitopes to be presented on the cell surface, decreasing vaccine efficacy, but could generate an unwanted auto-immune reaction. Thus, we designed the linker sequence to a) avoid creating "junction" peptides that bind to MHC molecules, b) avoid proteasomal processing to create "junction" peptides, c) be efficiently translated and processed by the proteasome.

To avoid creation of "junction" peptides that bind MHC molecules, we compared different linker sequences. Glycine, for example, inhibits strong binding in MHC binding groove positions [Abastado JP. et al., J Immunol. 1993 Oct 1;151(7): 3569-75]. We examined multiple linker sequences and multiple linker lengths and calculated the number of "junction" peptides that bind MHC molecules. We used software tools from the Immune Epitope Database (IEDB) to calculate the likelihood that a given peptide sequence contains a ligand that will bind MHC Class I molecules.

In the B16 model, we identified 102 expressed, non-synonymous somatic mutations predicted to be presented on MHC Class I molecules. Using the 50 confirmed mutations, we computationally designed different vaccine constructs, including either the use of no linkers or the use of different linker sequences, and computed the number of deleterious "junction" peptides using the IEDB algorithm (Figure 8).

Table 5 shows the results of several different linkers, different linker lengths, and the use of no linker and five linkers. The number of MHC-binding junction peptides ranges from 2 to 91 for the 9 aa and 10 aa epitope predictions (top and middle). The size of the linker influences the number of junction peptides (bottom). For this sequence, the fewest 9 aa epitopes are predicted for the 7 aa linker sequence GGSGGGG.

The Linker 1 and Linker 2 used in the RNA poly-neo epitope vaccine constructs tested experimentally (see below) also had a favorably low number of predicted junctional neoepitopes. This holds true for predictions of 9-mers and 10-mers.

This demonstrates that the sequence of the linker is critically important for the creation of bad MHC binding epitopes. Furthermore, the length of the linker sequence impacts the number of bad MHC binding epitopes. We find that sequences that are G-rich hinder the creation of MHC-binding ligands.

Table 3. Impact of Linker (10 aa epitopes). The predicted number of bad epitopes defined as MHC Class I binding epitopes that contain junction sequences, for each peptide linker. Here, 10 amino acid epitopes are considered. Glycine-rich linkers have the fewest junction epitopes.

Linker	# bad epitopes (10 aa)
none	14
TSLNALLNAH	54
SIINFEKL	65
SSSSSSSSSS	85
GGGGGGGGGG	6
GGSGGGGSGG (Linker 1)	8
GGSGGGSGGG (Linker 2)	9

Table 4. Impact of Linker Part (9 aa epitopes). The predicted number of bad epitopes, defined as MHC Class I binding epitopes that contain junction sequences, for each peptide linker. Here, 9 amino acid epitopes are considered. Glycine-rich linkers have the fewest junction epitopes.

Linker	# bad epitopes (9 aa)
none	17
TSLNALLNAH	83
SIINFEKL	64
SSSSSSSSSS	33
GGGGGGGGGG	2
GGSGGGGSGG (Linker 1)	4
GGSGGGSGGG (Linker 2)	3

Table 5: Impact of Linker Part. The predicted number of bad epitopes, defined as MHC Class I binding epitopes that contain junction sequences, for each peptide linker. Here, 9 amino acid epitopes are considered. Top: the number of 9 aa junction epitopes for no linker and 5 diverse linkers. Middle: the number of 10 aa junction epitopes for no linker and 5 diverse linkers. Lower: the number of 99 aa junction epitopes for similar linkers of different lengths. Glycine-rich linkers have the fewest junction epitopes.

Linker sequence	# junction epitopes (9aa)
none	17
TSLNALLNA	91
SIINFEKL	64
SSSSSSSSS	33
GGGGGGGGG	2
GGSGGGGSG	4

Linker sequence	# junction epitopes (10aa)
none	14
TSLNALLNA	63
SIINFEKL	65
SSSSSSSSS	85
GGGGGGGGG	6
GGSGGGSGG	9

Linker sequence	# junction epitopes (9aa)
GGSGG	5
GGSGGG	4
GGSGGGG	2
GGSGGGGS	7
GGSGGGGSG	4
GGSGGGGSGG	4

To avoid proteasomal processing that may create "junction" peptides, we explored usage of different amino acids in the linker. Glycine rich sequences impair proteasomal processing [Hoyt MA et al. (2006). EMBO J 25 (8): 1720–9; Zhang M. and Coffino P. (2004) J Biol Chem 279 (10): 8635–41]. Thus glycine rich linker sequences act to minimize the number of linker-containing peptides that can be processed by the proteasome.

The linker should allow the mutation-containing peptides to be efficiently translated and processed by the proteasome. Amino acids glycine and serine are flexible [Schlessinger A and Rost B., Proteins. 2005 Oct 1;61(1):115-26]; including them in a linker results in a more flexible protein. We incorporate glycine and serine into the linker to increase protein flexibility which should allow more efficient translation and processing by the proteasome, in turn enabling better access to the encoded antigenic peptides.

Thus, the linker should be glycine rich to hinder the creation of MHC binding bad epitopes; should hinder the ability of the proteasome to process linker peptides, which can be accomplished through inclusion of glycine; and should be flexible to increase access to mutation containing peptides, which can be accomplished through the combination of glycine and serine amino acids. Therefore, in one embodiment of the vaccine construct of the invention, the sequences GGSGGGGSGG and GGSGGGSGGS are preferably included as linker sequences.

Example 7: RNA poly-neo epitope vaccine

The RNA poly-neo epitope vaccine constructs are based on the pST1-A120 vector containing a T7 promotor, a tandem beta-globin 3' UTR sequence and a 120-bp poly(A) tail, which have been shown to increase the stability and translational efficiency of the RNA thereby enhancing the T-cell stimulatory capacity of the encoded antigen ((Holtkamp S. et al., Blood 2006; PMID: 16940422). In addition, an MHC class I signal peptide fragment and the transmembrane and cytosolic domains including the stop-codon (MHC class I trafficking signal or MITD) flanking a poly-linker sequence for cloning the epitopes were inserted (Kreiter S. et al., J. Immunol., 180: 309-318, 2008). The latter have been shown to increase the antigen presentation, thereby enhancing the expansion of antigen-specific CD8+ and CD4+ T cells and improving effector functions.

To provide RNA poly-neo epitope constructs for the 50 identified and validated mutations of B16F10 3 RNA constructs were generated. The construct consists of codon optimized sequences coding for (i) a mutated epitope of 25 amino acids, (ii) a glycine/serine-rich linker, (iii) repetitions of mutated epitope sequence followed by a glycine/serine-rich linker. The chain of mutated epitope containing sequences and linkers is flanked by suitable recognition sites for restriction endonucleases to be cloned into the pST1-based construct as described above. The vaccine constructs were designed and synthesized by GENEART. After verification of the sequence, these were cloned into the pST1-based vector backbone to obtain the RNA poly-neo epitope vaccine constructs.

Description of the Clinical Approach

The Clinical Application will cover following steps:

- Eligible patients must consent to DNA analysis by next generation sequencing.

- Tumor specimen obtained from routine diagnostic procedures (paraffin embedded formalin fixed tissue) and peripheral blood cells will be obtained and used for mutation analysis as described.
- Discovered mutations will be confirmed
- Based on Prioritization vaccine will be designed. For RNA vaccines a master plasmid template will be generated by gene synthesis and cloning
- Plasmids will be used for clinical grade RNA production, quality control and release of the RNA vaccine.
- The vaccine drug product will be sent to the respective trial center for clinical application.
- The RNA vaccine can be used as a naked vaccine in formulation buffer or encapsulated into nanoparticles or liposomes for direct injection into e.g. lymph nodes, s.c., i.v., i.m.. Alternatively, the RNA vaccine can be used for in vitro transfection e.g of dendritic cells for adoptive transfer.

The whole clinical process takes less than 6 weeks. The "lag phase" between patient informed consent and availability of the drug will be carefully addressed by the clinical trial protocol, including allowing the standard treatment regimen to be continued until the investigational drug product is available.

Example 8: Identification of tumor mutations and exploiting them for tumor vaccination

We applied NGS exome resequencing for mutation discovery in the B16F10 murine melanoma cell line and identified 962 non-synonymous somatic point mutations, 563 in expressed genes. Potential driver mutations occur in classical tumor suppressor genes (*Pten*, *Trp53*, *Tp63*, *Pml*) and genes involved in proto-oncogenic signaling pathways that control cell proliferation (e.g. *Mdm1*, *Pdgfra*), cell adhesion and migration (e.g. *Fdz7*, *Fat1*) or apoptosis (*Casp9*). Moreover, B16F10 harbors mutations in *Aim1* and *Trrap* that were previously described to be frequently altered in human melanoma.

The immunogenicity and specificity of 50 validated mutations were assayed using C57BL/6 mice immunized with long peptides encoding the mutated epitopes. One third (16/50) of them were shown to be immunogenic. Of these, 60% elicited immune responses preferentially directed against the mutated sequence as compared to the wild type sequence.

We tested the hypothesis in tumor transplant models. Immunization with peptides conferred *in vivo* tumor control in protective and therapeutic settings, qualifying mutated epitopes containing single amino acid substitutions as effective vaccines.

Animals

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

Cells

B16F10 melanoma cell line was purchased in 2010 from the American Type Culture Collection (Product: ATCC CRL-6475, Lot Number: 58078645). Early (3rd, 4th) passages of cells were used for tumor experiments. Cells were routinely tested for *Mycoplasma*. Re-authentification of cells has not been performed since receipt.

Next-generation sequencing

Nucleic acid extraction and sample preparation: DNA and RNA from bulk B16F10 cells and DNA from C57BL/6 tail tissue were extracted in triplicate using Qiagen DNeasy Blood and Tissue kit (for DNA) and Qiagen RNeasy Micro kit (for RNA).

DNA exome sequencing: Exome capture for DNA resequencing was performed in triplicate using the Agilent Sure-Select mouse solution-based capture assay (Gnirke A et al., Nat Biotechnol 2009;27:182-9), designed to capture all mouse 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 and 5' phosphorylated and 3' adenylated according to the maufacturer's instructions. Illumina paired end adapters were ligated to the gDNA fragments using a 10:1 molar ratio of adapter to gDNA. Enriched pre capture and flow cell specific sequences were added using Illumina PE PCR primers 1.0 and 2.0 for 4 PCR cycles. 500 ng of adapter ligated, PCR enriched gDNA fragments were hybridized to Agilent's SureSelect biotinylated mouse whole exome RNA library 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. These eluted gDNA fragments were PCR amplified post capture 10 cycles. Exome enriched gDNA libraries were clustered on the cBot using Truseq SR cluster kit v2.5 using 7 pM and 50 bps were sequenced on the Illumina HiSeq2000 using Truseq SBS kit-HS 50 bp.

RNA gene expression "transcriptome" profiling (RNA-Seq): Barcoded mRNA-seq cDNA libraries were prepared in triplicate, from 5 µg of total RNA (modified Illumina mRNA-seq protocol). 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 SuperScriptII (Invitrogen) followed by second strand synthesis using DNA polymerase I and RNaseH. cDNA was end repaired, 5' phosphorylated and 3' adenylated according to the manufacturer's 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 200-220 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 (Finnzymes). All cleanups up to this step were done with 1,8x volume of AgencourtAMPure 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 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 reference genomic sequence (Waterston RH et al., Nature 2002;420:520-62) with bowtie (version 0.12.5) (Langmead B et al., Genome Biol 2009;10:R25). For genome alignments, two mismatches were allowed and only the best alignment ("-v2 –best") was recorded; for transcriptome alignments the default parameters were used. Reads not alignable to the genomic sequence were aligned to a database of all possible exon-exon junction sequences of RefSeq transcripts (Pruitt KD et al., Nucleic Acids Res 2007;35:D61-D65). 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) (Mortazavi A et al., Nat Methods 2008;5:621-8).

NGS data analysis, somatic mutation discovery: Somatic mutations were identified as described in Example 9. 50 nucleotide (nt), single-end reads were aligned to the mm9 reference mouse genome using bwa (default options, version 0.5.8c) (Li H and Durbin R, Bioinformatics 2009;25:1754-60). Ambiguous reads mapping to multiple locations of the genome were removed. Mutations were identified using three software programs: samtools (version 0.1.8) (Li H, Bioinformatics 2011;27:1157-8), GATK (version 1.0.4418) (McKenna A et al, Genome Res 2010;20:1297-303), and SomaticSniper

(Ding L et al., Hum Mol Genet 2010;19:R188-R196). Potential variations identified in all B16F10 triplicates were assigned a "false discovery rate" (FDR) confidence value (cf. Example 9).

Mutation selection, validation, and function

Selection: Mutations had to fulfill following criteria to be selected: (i) present in all B16F10 and absent in all C57BL/6 triplicates, (ii) FDR ≤0.05, (iii) homogeneous in C57BL/6, (iv) occur in a RefSeq transcript, and (v) cause non-synonymous changes to be scored as an authentic mutation. Selection for validation and immunogenicity testing required that mutations are expressed genes (median RPKM across replicates >10).

Validation: DNA-derived mutations were classified as validated if confirmed by either Sanger sequencing or the B16F10 RNA-Seq reads. All selected variants were amplified from 50 ng of DNA from B16F10 cells and C57BL/6 tail tissue using flanking primers, products visualized (QIAxcel system, Qiagen) and purified (QIAquick PCR Purification Kit, Qiagen). The amplicon of the expected size was excised from the gel, purified (QIAquick Gel Extraction Kit, Qiagen) and subjected to Sanger sequencing (Eurofins MWG Operon, Ebersberg, Germany) with the forward primer used for PCR amplification.

Functional impact: The programs SIFT (Kumar P et al., Nat Protoc 2009;4:1073-81) and POLYPHEN-2 (Adzhubei IA et al., Nat Methods 2010;7:248-9), which predict the functional significance of an amino acid on protein function based on the location of protein domains and cross-species sequence conservation, were employed to assess the impact of selected mutations. Ingenuity IPA tools were used to infer gene function.

Synthetic peptides and adjuvants

All peptides including ovalbumin class I ($OVA_{258-265}$), class II (OVA class $II_{330-338}$), influenza nucleoprotein ($Inf\text{-}NP_{366-374}$), vesiculo-stomatitis virus nucleoprotein ($VSV\text{-}NP_{52-59}$) and tyrosinase-related protein 2 ($Trp2_{180-188}$) were purchased from Jerini Peptide Technologies (Berlin, Germany). Synthetic peptides were 27 amino acids long with the mutated (MUT) or wild type (WT) amino acid on position 14. Polyinosinic:polycytidylic acid (poly(I:C), InvivoGen) was used as subcutaneously injected adjuvant. MHC-Pentamer specific for the $Inf\text{-}NP_{366-374}$ peptide was purchased from ProImmune Ltd..

Immunization of mice

Age-matched female mice C57BL/6 mice were injected subcutaneously with 100 µg peptide and 50 µg poly(I:C) formulated in PBS (200 µl total volume) into the lateral flank (5 mice per group). Every group was immunized on day 0 and day 7 with two different mutation coding peptides, one peptide per flank. Twelve days after the initial injection mice were sacrificed and splenocytes were isolated for immunological testing.

Alternatively, age-matched female mice C57BL/6 mice were injected intravenously with 20 µg *in vitro* transcribed RNA formulated with 20 µl Lipofectamine™ RNAiMAX (Invitrogen) in PBS in a total injection volume of 200 µl (3 mice per group). Every group was 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. DNA-sequences representing one (Monoepitope), two (Biepitope), or 16 mutations (Polyepitope) were constructed using 50 amino acids (aa) with the mutation on position 25 (Biepitope) or 27 aa with the mutation on position 14 (Mono- and Polyepitope), were separated by a glycin/serine linker of 9aa and cloned into the pST1-2BgUTR-A120 backbone (Holtkamp et al., Blood 2006;108:4009-17). *In vitro* transcription from this template and purification were previously described (Kreiter et al., Cancer Immunol Immunother 2007;56:1577-87).

Enzyme-linked immunospot assay

Enzyme-linked immunospot (ELISPOT) assay (Kreiter S et al., Cancer Res 2010;70:9031-40) and generation of syngeneic bone marrow derived dendritic cells (BMDCs) as stimulators were previously described (Lutz MB et al., J Immunol Methods 1999;223:77-92). BMDCs were either peptide pulsed (2 µg/ml), or transfected with *in vitro* transcribed (IVT) RNA coding for the indicated mutation or for control RNA (eGFP-RNA). Sequences representing two mutations, each comprising 50 amino acids with the mutation on position 25 and separated by a glycin/serine linker of 9aa were cloned into the pST1-2BgUTR-A120 backbone (Holtkamp S et al., Blood 2006;108:4009-17). *In vitro* transcription from this template and purification were previously described (Kreiter S et al., Cancer Immunol Immunother 2007;56:1577-87). For the assay, 5×10^4 peptide or RNA engineered BMDCs were coincubated with 5×10^5 freshly isolated splenocytes in a microtiter plate coated with anti-IFN-γ antibody (10 µg/mL, clone AN18; 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, when either the test gave a p-value < 0.05 and the mean spot numbers were >30 spots/5x10^5 effector cells. Reactivities were rated by mean spot numbers (-: <30; +: >30; ++: >50; +++ >200 spots/well).

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 BMDCs. After incubation the cells were washed with PBS, resuspended in 50µl PBS 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 100µL 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 50µL/sample 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-α 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).

B16 melanoma tumor model

For tumor vaccination experiments 7.5 × 10^4 B16F10 melanoma cells were inoculated s.c. into the flanks of C57BL/6 mice. In the prophylactic setting, immunization with mutation-specific peptide was performed 4 days before and on days 2 and 9 after tumor inoculation. For the therapeutic experiment the peptide vaccine was administered on days 3 and 10 after tumor injection. The tumor sizes were measured every three days and mice were sacrificed when tumor diameter reached 15 mm.

Alternatively, for tumor vaccination experiments 1 × 10^5 B16F10 melanoma cells were inoculated s.c. into the flanks of age-matched female C57BL/6 mice. Peptide vaccination was performed on days 3, 10 and 17 after tumor inoculation with 100 µg peptide and 50 µg poly(I:C) formulated in PBS (200 µl total volume) injected subcutaneously into the lateral flank. RNA immunizations were performed using 20 µg in vitro transcribed mutation-encoding RNA formulated with 20 µl Lipofectamine™ RNAiMAX (Invitrogen) in PBS in a total injection volume of 200 µl. As control one group of animals was injected with RNAiMAX (Invitrogen) in PBS. The animals were immunized on days 3, 6, 10, 17 and 21 after tumor inoculation. The tumor sizes were measured every three days using a caliper and mice were sacrificed when tumor diameter reached 15 mm.

Identification of non-synonymous mutations in B16F10 mouse melanoma

Our objective was to identify potentially immunogenic somatic point mutations in B16F10 mouse melanoma by NGS and to test these for *in vivo* immunogenicity by peptide vaccination of mice measuring elicited T-cell responses by ELISPOT assay (Figure 9A). We sequenced the exomes of the C57BL/6 wild type background genome and of B16F10 cells, each with triplicate extractions and captures. For each sample, more than 100 million single-end 50 nt reads were generated. Of these 80%, align uniquely to the mouse mm9 genome and 49% align on target, demonstrating successful target enrichment and resulting in over 20-fold coverage for 70% of the target nucleotides in each of the triplicate samples. RNA-Seq of B16F10 cells, also profiled in triplicate, generated a median of 30 million single-end 50 nt reads, of which 80% align to the mouse transcriptome.

DNA reads (exome-capture) from B16F10 and C57BL/6 were analyzed to identify somatic mutations. Copy number variation analysis (Sathirapongsasuti JF et al., Bioinformatics 2011;27:2648-54) demonstrated DNA amplifications and deletions in B16F10, including the homozygous deletion of tumor suppressor *Cdkn2a* (Cyclin-dependent kinase inhibitor 2A, p16Ink4A). Focusing on point mutations to identify possible immunogenic mutations, we identified 3570 somatic point mutations at FDR ≤ 0.05 (Figure 9B). The most frequent class of mutations were C>T / G>A transitions, typically resulting from ultraviolet light (Pfeifer GP et al., Mutat Res 2005;571:19-31). Of these somatic mutations, 1392 occur in transcripts, with 126 mutations in untranslated regions. Of the 1266 mutations in coding regions, 962 cause non-synonymous protein changes and 563 of these occur in expressed genes (Figure 9B).

Assignment of identified mutations to carrier genes and validation

Noteworthy, many of the mutated genes (962 genes containing non-synonymous somatic point mutations) have been previously associated with the cancer phenotypes. Mutations were found in established tumor suppressor genes, including *Pten*, *Trp53 (*also called *p53)*, and *Tp63*. In *Trp53*, the best established tumor suppressor (Zilfou JT et al., Cold Spring Harb Perspect Biol 2009;1:a001883), the asparagine to aspartic acid mutation at protein position 127 (p.N127D) is localized in the DNA binding domain and is predicted by SIFT to alter function. *Pten* contained two mutations (p.A39V, p.T131P), both of which are predicted to have deleterious impact on protein function. The p.T131P mutation is adjacent to a mutation (p.R130M) shown to diminish phosphatase activity (Dey N et al., Cancer Res 2008;68:1862-71). Moreover, mutations were found in genes associated with DNA repair pathways, such as *Brca2* (breast cancer 2, early onset), *Atm* (ataxia telangiectasia mutated), *Ddb1* (damage-specific DNA binding protein 1) and *Rad9b* (RAD9 homolog B). Furthermore, mutations occur in other tumor associated genes, including *Aim1* (tumor suppressor "Absent In Melanoma 1"), *Flt1* (oncogene *Vegr1*, fms-related tyrosine kinase 1), *Pml* (tumor suppressor "promyelocytic leukemia"), *Fat1* ("FAT tumor suppressor homolog 1"), *Mdm1* (TP53 binding nuclear protein), *Mta3* (metastasis associated 1 family, member 3), and *Alk* (anaplastic lymphoma receptor tyrosine kinase). We found a mutation at p.S144F in *Pdgfra* (platelet-derived growth factor receptor, alpha polypeptide), a cell-membrane-bound receptor tyrosine kinase of the MAPK/ERK pathway, previously identified in tumors (Verhaak RG et al., Cancer Cell 2010;17:98-110). A mutation occurs at p.L222V in *Casp9* (caspase 9, apoptosis-related cysteine peptidase). CASP9 proteolytically cleaves poly(ADP-ribose) polymerase (PARP), regulates apoptosis, and has been linked to several cancers (Hajra KM et al., Apoptosis 2004;9:691-704). The mutation we found may potentially impact PARP and apoptosis signaling. Most interestingly, no mutations were found in *Braf*, *c-Kit*, *Kras* or *Nras*. However, mutations were identified in *Rassf7* (RAS-associated protein) (p.S90R), *Ksr1* (kinase suppressor of ras 1) (p.L301V), and *Atm* (PI3K pathway) (p.K91T), all of which are predicted to have significant impact on protein function. *Trrap* (transformation/transcription domain-associated protein) was identified earlier this year in human melanoma specimens as a novel potential melanoma target (Wei X et al., Nat Genet 2011;43:442-6). In B16F10, a *Trrap* mutation occurs at p.K2783R and is predicted to disturb the overlapping phosphatidylinositol kinase (PIK)-related kinase FAT domain.

From the 962 non-synonymous mutations identified using NGS, we selected 50 mutations, including 41 with FDR < 0.05, for PCR-based validation and immunogenicity testing. Selection criteria were location in an expressed gene (RPKM > 10) and predicted immunogenicity. Noteworthy, we were able to validate all 50 mutations (Table 6, Figure 9B).

Table 6: Mutations selected for validation. From left: assigned ID, gene symbol, amino acid substitution and position, gene name, predicted subcellular localization and type (Ingenuity).

ID	Symbol	Change	Entrez Gene Name	Subcellular localization	Type
MUT1	*Fzd7*	p.G304A	frizzled family receptor 7	Plasma Membrane	G-protein coupled receptor
MUT2	*Xpot*	p.I830S	exportin, tRNA (nuclear export receptor for tRNAs)	Nucleus	other
MUT3	*Ranbp2*	p.Q2871H	RAN binding protein 2	Nucleus	enzyme
MUT4	*Dnajb12*	p.P54T	DnaJ (Hsp40) homolog, subfamily B, member 12	Cytoplasm	other
MUT5	*Eef2*	p.G795A	eukaryotic translation elongation factor 2	Cytoplasm	translation regulator
MUT6	*Ptrf*	p.D382G	polymerase I and transcript release factor	Nucleus	transcription regulator
MUT7	*Trp53*	p.N128D	tumor protein p53	Nucleus	transcription regulator
MUT8	*Ddx23*	p.V602A	DEAD (Asp-Glu-Ala-Asp) box polypeptide 23	Nucleus	enzyme
MUT9	*Golgb1*	p.E2855D	golgin B1	Cytoplasm	other
MUT10	*Pcdhga11*	p.G82R	protocadherin gamma subfamily A, 11	Plasma Membrane	other
MUT11	*Snx15*	p.E211G	sorting nexin 15	Cytoplasm	transporter
MUT12	*Gnas*	p.S112G	GNAS (guanine nucleotide binding protein, alpha stimulating) complex locus	Plasma Membrane	enzyme
MUT13	*Fndc3b*	p.C561W	fibronectin type III domain containing 3B	Cytoplasm	other
MUT14	*Sbno1*	p.P309T	strawberry notch homolog 1 (Drosophila)	unknown	enzyme
MUT15	*Pi4k2b*	p.R344Q	phosphatidylinositol 4-kinase type 2 beta	Cytoplasm	kinase
MUT16	*Thumpd3*	p.T243S	THUMP domain containing 3	unknown	other
MUT17	*Tnpo3*	p.G504A	transportin 3	Cytoplasm	other
MUT18	*Numa1*	p.Q447K	nuclear mitotic apparatus protein 1	Nucleus	other
MUT19	*Wwp2*	p.E742K	WW domain containing E3 ubiquitin protein ligase 2	Cytoplasm	enzyme
MUT20	*Tubb3*	p.G402A	tubulin, beta 3	Cytoplasm	other
MUT21	*Atp11a*	p.R522S	ATPase, class VI, type 11A	Plasma Membrane	transporter
MUT22	*Asf1b*	p.A141P	ASF1 anti-silencing function 1 homolog B (S. cerevisiae)	Nucleus	other
MUT23	*Wdr82*	p.I221L	WD repeat domain 82	Nucleus	other
MUT24	*Dag1*	p.P425A	dystroglycan 1 (dystrophin-associated glycoprotein 1)	Plasma Membrane	transmembrane receptor
MUT25	*Plod2*	p.F530V	procollagen-lysine, 2-oxoglutarate 5-dioxygenase 2	Cytoplasm	enzyme
MUT26	*Orc2*	p.F278V	origin recognition complex, subunit 2	Nucleus	other
MUT27	*Obsl1*	p.T1764M	obscurin-like 1	unknown	other
MUT28	*Ppp1r7*	p.L170P	protein phosphatase 1, regulatory (inhibitor) subunit 7	Nucleus	phosphatase
MUT29	*Mthfd1l*	p.F294V	methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 1-like	Cytoplasm	enzyme
MUT30	*Kif18b*	p.K739N	kinesin family member 18B	unknown	other
MUT31	*Ascc2*	p.A59G	activating signal cointegrator 1 complex subunit 2	unknown	other
MUT32	*Itsn2*	p.S1551R	intersectin 2	Cytoplasm	other
MUT33	*Pbk*	p.V145D	PDZ binding kinase	Cytoplasm	kinase
MUT34	*Klhl22*	p.F179V	kelch-like 22 (Drosophila)	unknown	other
MUT35	*Ddb1*	p.L438I	damage-specific DNA binding protein 1, 127kDa	Nucleus	other
MUT36	*Tm9sf3*	p.Y382H	transmembrane 9 superfamily member 3	Cytoplasm	transporter
MUT37	*Dpf2*	p.F275V	D4, zinc and double PHD fingers family 2	Nucleus	other
MUT38	*Atrn*	p.S745N	attractin	Extracellular Space	other
MUT39	*Snx5*	p.R373Q	sorting nexin 5	Cytoplasm	transporter
MUT40	*Armc1*	p.S85I	armadillo repeat containing 1	Cytoplasm	other
MUT41	*Ash1l*	p.L632I	ash1 (absent, small, or homeotic)-like (Drosophila)	Nucleus	transcription regulator
MUT42	*S100a13*	p.S18C	S100 calcium binding protein A13	Cytoplasm	other
MUT43	*2510039O18 Rik*	p.E391K	KIAA2013	unknown	other
MUT44	*Cpsf3l*	p.D314N	cleavage and polyadenylation specific factor 3-like	Nucleus	other
MUT45	*Mkrn1*	p.N346Y	makorin ring finger protein 1	unknown	other
MUT46	*Actn4*	p.F835V	actinin, alpha 4	Cytoplasm	other
MUT47	*Rpl13a*	p.A24G	ribosomal protein L13a	Cytoplasm	other
MUT48	*Def8*	p.R255G	differentially expressed in FDCP 8 homolog (mouse)	unknown	other
MUT49	*Fat1*	p.I1940M	FAT tumor suppressor homolog 1 (Drosophila)	Plasma Membrane	other
MUT50	*Sema3b*	p.L663V	sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3B	Extracellular Space	other

Figure 9C shows the locations of the B16F10 chromosomes, genes density, gene expression, mutations, and filtered mutations (inner rings).

***In vivo* testing of immunogenicity testing with mutation-representing long peptides**

To provide antigens for immunogenicity testing of these mutations, we employed long peptides which have many advantages over other peptides for immunization (Melief CJ and van der Burg SH, Nat Rev Cancer 2008;8:351-60). Long peptides are capable of inducing antigen-specific CD8+ as well as CD4+ T-cells (Zwaveling S et al., Cancer Res 2002;62:6187-93; Bijker MS et al., J Immunol 2007;179:5033-40). Moreover, long peptides require processing to be presented on MHC molecules. Such uptake is most efficiently done by dendritic cells, which are optimal for priming a potent T-cell response. Fitting peptides, in contrast, do not require trimming and are loaded exogenously on all cells expressing MHC molecules, including non-activated B and T-cells, leading to induction of tolerance and fratricide (Toes RE et al., J Immunol 1996;156:3911-8; Su MW et al., J Immunol 1993;151:658-67). For each of the 50 validated mutations, we designed peptides of 27 amino acids length with the mutated or wild type amino acid positioned centrally. Thus, any potential MHC class I and class II epitope of 8 to 14 amino acid length carrying the mutation could be processed from this precursor peptide. As adjuvant for peptide vaccination we used poly(I:C) which is known to promote cross presentation and increase vaccine efficacy (Datta SK et al., J Immunol 2003;170:4102-10; Schulz O et al., Nature 2005;433:887-92). The 50 mutations were tested *in vivo* in mice for induction of T-cells. Impressively, 16 out of 50 mutation-coding peptides were found to elicit immune responses in immunized mice. The induced T-cells displayed different reactivity patterns (Table 7).

Table 7: Summary of T-cell reactivities determined consecutive to vaccination with mutation encoding peptide. Statistical analysis was done by student's t-test and Mann-Whitney test (non-parametric test). Responses were considered significant, when either test gave a p-value < 0,05 and the mean spot numbers were >30 spots/5x10^5 effector cells. Reactivities were rated by mean spot numbers -: <30; +: >30; ++: >50; +++ >200 spots/well.

Mutation	Gene Symbol	Reactivity against mutation	Reactivity against WT	Mutation	Gene Symbol	Reactivity against mutation	Reactivity against WT
MUT01	*Fzd7*	-	-	MUT26	Orc2	-	-
MUT02	*Xpot*	-	-	MUT27	Obsl1	-	-
MUT03	*Ranbp2*	-	-	MUT28	Ppp1r7	–	–
MUT04	*Dnajb12*	-	-	MUT29	Mthfd1l	–	-
MUT05	*Eef2*	——	——	MUT30	Kif18b	+—	-
MUT06	*Ptrf*	-	-	MUT31	Ascc2	-	-
MUT07	*Trp53*	-	-	MUT32	Itsn2	-	-
MUT08	*Ddx23*	-	-	MUT33	Pbk	-	-
MUT09	*Golgb1*	-	-	MUT34	Klhl22	-	-
MUT10	*Pcdhga11*	-	-	MUT35	Ddb1	-	-
MUT11	*Snx15*	-	-	MUT36	Tm9sf3	+	-
MUT12	*Gnas*	+	-	MUT37	Dpf2	-	-
MUT13	*Fndc3b*	-	-	MUT38	Atrn	-	-
MUT14	*Sbno1*	-	-	MUT39	Snx5	-	-
MUT15	*Pi4k2b*	-	-	MUT40	Armc1	-	-
MUT16	*Thumpd3*	-	-	MUT41	Ash1l	-	-
MUT17	*Tnpo3*	+++	++	MUT42	S100a13	-	-
MUT18	*Numa1*	-	-	MUT43	Rik	-	-
MUT19	*Wwp2*	-	-	MUT44	Cpsf3l	+++	++
MUT20	*Tubb3*	+++	-	MUT45	Mkrn1	++	++
MUT21	*Atp11a*	-	-	MUT46	Actn4	++	+
MUT22	*Asf1b*	++	++	MUT47	Rpl13a	-	-
MUT23	*Wdr82*	-	-	MUT48	Def8	++	++
MUT24	*Dag1*	++	+	MUT49	Fat1	-	-
MUT25	*Plod2*	+—	++	MUT50	Sema3b	+—	++

Eleven peptides induced an immune response preferentially recognizing the mutated epitope. This is exemplified for mice immunized with mutations 30 (MUT30, *Kif18b*) and 36 (MUT36, *Plod2*) (Figure 10A). ELISPOT testing revealed strong mutation-specific immune responses without cross reactivity against the wild-type peptide or an unrelated control peptide (VSV-NP). With five peptides, including mutations 05 (MUT05, *Eef2*) and 25 (MUT25, *Plod2*) (Figure 10A), immune responses with comparable recognition of both the mutated as well as the wild-type peptide were obtained. The majority of mutated peptides were not capable of inducing significant T-cell responses as exemplified by mutations 01 (MUT01, *Fzd7*), 02 (MUT02, *Xpot*), and 07 (MUT07, *Trp53*). Immune responses induced by several of the discovered mutations were well in the range of immunogenecity (500 spots/5x10^5 cells) generated by immunizing mice as a positive control with a described MHC-class I epitope from the murine melanoma tumor antigen tyrosinaserelated protein 2 (Trp2180-188, Figure 10A) (Bloom MB et al., Exp Med 1997;185:453-9; Schreurs MW et al. Cancer Res 2000;60:6995-7001). For selected peptides that induce a strong mutation-specific T-cell response, we confirmed immune recognition by an independent approach. Instead of long peptides, *in vitro* transcribed RNA (IVT RNA) coding for the mutated peptide fragments

MUT17, MUT30 and MUT44 was used for the immunological read-out. BMDCs transfected with mutation-coding RNA or irrelevant RNA served as antigen presenting cells (APCs) in an ELISPOT assay, whereas spleen cells of immunized mice served as effector cell population. BMDCs transfected with MUT17, MUT30 and MUT44 encoding mRNA were specifically and strongly recognized by splenocytes of mice immunized with the respective long peptides (Figure 10B). Significantly lower reactivity against control RNA-transfected BMDCs was recorded, which is likely due to the unspecific activation of the BMDCs by the single stranded RNA (student's t-test; MUT17: $p = 0.0024$, MUT30: $p = 0.0122$, MUT44: $p = 0.0075$). These data confirm that the induced mutation-specific T-cells in effect recognize endogenously processed epitopes. Two mutations that induce a preferred recognition of mutated epitopes are in genes *Actn4* and *Kif18b*. The somatic mutation in ACTN4 (actinin, alpha 4) is at p.F835V in the calcium binding "EF-hand" protein domain. While both SIFT and POLYPHEN predict a significant impact of this mutation on protein function, the gene is not an established oncogene. However, mutation-specific T-cells against ACTN4 have been recently associated with a positive patient outcome (Echchakir H et al., Cancer Res 2001;61:4078-83). KIF18B (kinesin family member 18B) is a kinesin with microtubule motor activity and ATP and nucleotide binding that is involved in regulation of cell division (Lee YM et al., Gene 2010;466:16-25) (Figure 10C). The DNA sequence at the position encoding p.K739 is homogeneous in the reference C57BL/6, whereas B16F10 DNA reads reveal a heterozygous somatic mutation. Both nucleotides were detected in the B16F10 RNA-Seq reads and validated by Sanger sequencing. KIF18B has not been previously associated with a cancer phenotype. The mutation p.K739N is not localized in a known functional or conserved protein domain (Figure 10C, bottom) and thus most likely is a passenger rather than a driver mutation. These examples suggest a lack of correlation between the capability of inducing mutation-recognizing immune response and a functional or immunological relevance.

In vivo assessment of antitumoral activity of vaccine candidates

To assess whether immune responses elicited *in vivo* translate in anti-tumoral effects in tumor bearing mice, we chose MUT30 (mutation in Kifl8b) and MUT44 as examples. These mutations had been shown to induce a strong immune reaction preferentially against the mutated peptide and to be endogenously processed (Figure 10A, B). The therapeutical potential of vaccinating with mutated peptides was explored by immunizing mice with either MUT30 or MUT44 and adjuvant 3 and 10 days after grafting with $7.5x10^5$ B16F10. Growth of tumors was inhibited by both peptide vaccinations as compared to the control group

(Figure 11A). As B16F10 is a very aggressively growing tumor, we also tested protective immune responses. Mice were immunized with MUT30 peptide, inoculated s.c. with $7.5x10^5$ B16F10 cells 4 days later and boosted with MUT30 2 and 9 days after tumor challenge. Complete tumor protection and survival of 40% of the mice treated with MUT30 were observed, whereas all mice in the control treated group died within 44 days (Figure 11B left). In those mice, developing tumors despite immunization with MUT30, growth of tumors was slower resulting in an elongation of the median survival by 6 days as compared to the control group (Figure 11B right). These data imply that already vaccination against a single mutation is able to confer anti-tumoral effects.

Immunization with mutation-coding RNAs

The 50 validated mutations from the B16F10 melanoma cell line were used to construct different RNA vaccines. DNA-sequences representing one (Monoepitope), two (Biepitope), or 16 different mutations (Polyepitope), were constructed using 50 amino acids (aa) with the mutation on position 25 (Biepitope) or 27 aa with the mutation on position 14 (Mono- and Polyepitope) and were separated by a glycine/serine linker of 9aa. These constructs were cloned into the pST1-2BgUTR-A120 backbone for *in vitro* transcription of mRNA (Holtkamp et al., Blood 2006;108:4009-17).

To test the *in vivo* ability to induce T-cell responses against the different RNA-vaccines groups of three C57BL/6 mice were immunized by formulation of the RNA with RNAiMAX lipofectamine and subsequent intravenous injection. After 5 immunizations the mice were sacrificed and splenocytes were analyzed for mutation-specific T-cell responses using intracellular cytokine staining and IFN-γ ELISPOT analysis after restimulation with the corresponding mutation coding peptide or control peptide (VSV-NP).

Figure 12 shows one example for each vaccine design. In the upper row the mice were vaccinated with the Monoepitope-RNA coding for MUT30 (mutation in *Kif18b*), which induces MUT30-specific $CD4^+$ T-cells (see exemplary FACS-plot). In the middle row the graph and FACS-plot show induction of MUT08-specific (mutation in *Ddx23*) $CD4^+$ T-cells after immunization with the Biepitope coding for MUT33 and MUT08. In the lower row mice were immunized with a Polyepitope encoding 16 different mutations including MUT08, MUT33 and MUT27 (see Table 8). The graph and FACS-plot illustrate that MUT27 reactive T-cells are of a CD8 phenotype.

Table 8. Overview of mutations and gene names encoded by Mono-, Bi- and Polyepitope RNA-vaccines.

Construct	Encoded mutation	Gene annotation
Monoepitope	MUT30	*Kif18b*
Biepitope	MUT33	*Pbk*
	MUT08	*Ddx23*
Polyepitope	MUT01	*Fzd7*
	MUT02	*Xpot*
	MUT03	*Ranbp2*
	MUT04	*Dnajb12*
	MUT05	*Eef2*
	MUT06	*Ptrf*
	MUT07	*Trp53*
	MUT08	*Ddx23*
	MUT26	*Orc2*
	MUT27	*Obsl1*
	MUT28	*Ppp1r7*
	MUT29	*Mthfd1l*
	MUT30	*Kif18b*
	MUT31	*Ascc2*
	MUT32	*Itsn2*
	MUT33	*Pbk*

The same Polyepitope was used to generate the data shown in Figure 13. The graph shows ELISPOT data after restimulation of splenocytes with control (VSV-NP), MUT08, MUT27 and MUT33 peptides, proving that the Polyepitope vaccine can induce specific T-cell responses against several different mutations.

Taken together the data show the possibility to induce mutation-specific T-cells using RNA-encoded Mono-,Bi- and Polyepitopes. Furthermore, the data show induction of CD4$^+$ and CD8$^+$ T cells and the induction of several different specificities from one construct.

Immunization with model epitopes

To further characterize the polyepitopic RNA-vaccine design a DNA-sequence was constructed, which included five different known model epitopes including one MHC class II epitope (ovalbumin class I (SIINFEKL), class II (OVA class II), influenza nucleoprotein (Inf-NP), vesiculo-stomatitis virus nucleoprotein (VSV-NP) and tyrosinase-related protein 2 (Trp2)). The epitopes were separated with the same glycine/serine linker of 9aa used for the mutation Polyepitope. This constructs was cloned into the pST1-2BgUTR-A120 backbone for *in vitro* transcription of mRNA.

The *in vitro* transcribed RNA was used to vaccinate five C57BL/6 mice by intranodal immunization (four immunizations with 20µg of RNA into the inguinal lymphnodes). Five days after the last immunization blood samples and splenocytes were taken from the mice for analysis. Figure 14A shows IFN-γ ELISPOT analysis of the splenocytes restimulated with the indicated peptides. It can be clearly seen that all three MHC-class I epitope (SIINFEKL, Trp2 and VSV-NP) induce a very high number of antigen-specific CD8^{+} T cells. Also the MHC-class II epitope OVA class II induces a strong CD4^{+} T-cell response. The fourth MHC class I epitope was analyzed by staining of Inf-NP-specific CD8^{+} T-cells with a fluorescence-labeled pentameric MHC-peptide complex (Pentamer) (Figure 14B).

These data prove that the polyepitope design using the glycine/serine linker to separate different immunogenic MHC-class I and -class II epitopes is able to induce specific T-cells against every encoded epitope, regardless of its immunodominance.

Anti-tumoral response after therapy with a mutation-encoding polyepitopic RNA vaccine

The same Polyepitope which was analyzed in Figure 13 for immunogenicity was used to investigate the anti-tumoral activity of the mutation-encoding RNAs against the B16F10 tumor cells. In detail, groups of C57BL/6 mice (n=10) were subcutaneously inoculated with 1 x 10^{5} B16F10 melanoma cells into the flank. On days 3, 6, 10, 17 and 21 the mice were immunized with the polytopic RNA using a liposomal transfection reagent. The control group was injected with liposomes alone.

Figure 21 shows the survival curves of the groups, revealing a strongly improved median survival of 27 days with 1 of 10 mice surviving without tumor compared to 18,5 days median survival in the control group.

Anti-tumoral response after therapy with a combination of mutated and normal peptide

Anti-tumoral activity of the validated mutations was evaluated by a therapeutic *in vivo* tumor experiment by using the MUT30 as a peptide vaccine. In detail, groups of C57BL/6 mice (n=8) were subcutaneously inoculated with 1 x 10^5 B16F10 melanoma cells into the flank. On day 3, 10 and 17 the mice were immunized using polyI:C as adjuvant with MUT30, tyrosinase-related protein 2 ($Trp2_{180-188}$) or a combination of both peptides. Trp2 is a known $CD8^+$ epitope expressed by the B16F10 melanoma cells.

Figure 15 A shows the mean tumor growth of the groups. It can be clearly seen that until day 28 the tumor growth is almost completely inhibited in the group which was immunized with the combination of the known $CD8^+$ T-cell epitope and the $CD4^+$ T-cell inducing MUT30. The known Trp2 epitope alone is not sufficient to provide a good anti-tumoral effect in this setting, but both single therapy groups (MUT30 and Trp2) still provide a tumor growth inhibition in comparison to the untreated group in the beginning of the experiment up to day 25. These data are strengthened by the survival curves shown in Figure 15 B. Clearly the median survival is increased by the mice injected with the single peptides, with 1/8 mice surviving in the group with Trp2 vaccination. In addition the group treated with both peptides shows an even better median survival with 2/8 mice surviving.

Taken together both epitopes act in a synergistic manner to provide a strong anti-tumoral effect.

Example 9: Framework for confidence-based somatic mutation detection and application to B16-F10 melanoma cells

NGS is unbiased in that it enables a high throughput discovery of variations within an entire genome or targeted regions, such as protein coding exons.

However, while revolutionary, the NGS platform is still prone to errors leading to erroneous variation calls. Furthermore, the quality of results is dependent on experimental design parameters and analysis methodologies. While variation calls typically include scores designed to differentiate true variations from errors, the utility of these scores is not fully understood, nor is their interpretation with regard to optimization of experiments. This is particularly true when comparing tissue states, such comparing tumor and normal for somatic mutations. As a consequence, researchers are forced to rely on personal experience to determine experimental parameters and arbitrary filtering thresholds for selecting mutations.

Our study aims a) to establish a framework for comparing parameters and methods to identify somatic mutations and b) to assign a confidence value to identified mutations. We sequence triplicate samples from C57BL/6 mice and the B16-F10 melanoma cell line. Using these data, we formulate the false discovery rate of detected somatic mutations, a measure that we then use to evaluate existing mutation discovery software and lab protocols.

Various experimental and algorithmic factors contribute to the false positive rate for variations found by NGS [Nothnagel, M. et al., Hum. Genet. 2011 Feb 23 [Epub ahead of print]]. The error sources include PCR artifacts, biases in priming [Hansen, K.D., et al., Nucleic. Acids. Res. 38, e131 (2010); Taub, M.A. et al., Genome Med. 2, 87 (2010)] and targeted enrichment [Bainbridge, M.N. et al., Genome Biol. 11, R62 (2010)], sequence effects [Nakamura, K. et al., Acids Res.(2011) first published online May 16, 2011 doi:10.1093/nar/gkr344], base calling causing sequence errors [Kircher, M. et al., Genome Biol. 10, R83 (2009). Epub 2009 Aug 14] and read alignment [Lassmann, T. et al., Bioinformatics 27, 130–131 (2011)], causing variation in coverage and sequencing errors which influence the further downstream analysis, e.g. variant calling around indels [Li, H., *Bioinformatics* 27, 1157-1158 (2011)].

No general statistical model has been described to describe the impact of different error sources on somatic mutation calls; only individual aspects are covered without removing all bias. Recent computational methods to measure the expected amount of false positive mutation calls include utilization of the transition/transversion ratio of a set of variations [Zhang, Z., Gerstein, M., Nucleic Acids Res 31, 5338-5348 (2003); DePristo, M.A. et al., Nature Genetics 43, 491–498 (2011)], machine learning [DePristo, M.A. *et al.*, *Nature Genetics* **43**, 491–498 (2011)] and inheritance errors when working with family genomes [Ewen, K.R. et al., Am. J. Hum. Genet. 67, 727-736 (2000)] or pooled samples [Druley, T.E. et al., Nature Methods 6, 263 - 265 (2009); Bansal, V., Bioinformatics 26, 318-324 (2010)]. For optimization purposes, Druley et al. [Druley, T.E. *et al.*, *Nature Methods* 6, 263 - 265 (2009)] relied on short plasmid sequence fragments, which however might not be representative for the sample. For a set of single nucleotide variations (SNVs) and selected experiments, a comparison to SNVs identified by other techniques is feasible [Van Tassell, C.P. et al., Nature Methods 5, 247 - 252 (2008)] but is difficult to evaluate in terms of novel somatic mutations.

Using an exome sequencing project as an example, we propose the calculation of a false discovery rate (FDR) based on NGS data alone. The method is not only applicable to the selection and prioritization of diagnostic and therapeutic targets, but also supports algorithm and method development by allowing us to define confidence-driven recommendations for similar experiments.

To discover mutations, DNA from tail tissue of three C57BL/6 (black6) mice (litter mates) and DNA from B16-F10 (B16) melanoma cells, in triplicate, were individually enriched for protein coding exons (Agilent Sure Select Whole Mouse Exome), resulting in 6 samples. RNA was extracted from B16 cells in triplicate. Single end 50 nt (1x50 nt) and paired end 100 nt (2x100 nt) reads were generated on an Illumina HiSeq 2000. Each sample was loaded into an individual lane, resulting in an average of 104 million reads per lane. DNA reads were aligned to the mouse reference genome using bwa [Li, H. Durbin, R., Bioinformatics 25, 1754-1760 (2009)] and RNA reads were aligned with bowtie [Langmead, B. et al., *Genome Biol.* **10**, R25 (2009)]. A mean coverage of 38 fold of 97% of the targeted regions was achieved for the 1x50 nt libraries, while the 2x100 nt experiment yielded an average coverage of 165 fold for 98% of the targeted regions.

Somatic variations were independently identified using the software packages SAMtools [Li, H. *et al.*, *Bioinformatics* **25**, 2078-2079 (2009)], GATK [DePristo, M.A. *et al.*, *Nature Genetics* **43**, 491–498 (2011)] and SomaticSNiPer [Ding, L. et al., *Hum. Mol. Genet* (2010) first published online September 15, 2010] (Fig. 16) by comparing the single nucleotide variations found in B16 samples to the corresponding loci in the black6 samples (B16 cells were originally derived from a black6 mouse). The potential mutations were filtered according to recommendations by the respective software authors (SAMtools and GATK) or by selecting an appropriate lower threshold for the somatic score of SomaticSNiPer, respectively.

To create a false discovery rate (FDR) for mutation discovery, we first intersected the mutation sites and obtained 1,355 high quality somatic mutations as consensus among all three programs (Fig. 17). However, the observed differences in the results of the applied software tools are substantial. To avoid erroneous conclusions, we developed a method to assign a FDR to each mutation using the replicates. Technical repeats of a sample should generate identical results and any detected mutation in this "same vs. same comparison" is a false positive. Thus, to determine the false discovery rate for somatic mutation detection in a tumor sample relative to a normal sample ("tumor comparison"), we can use a technical repeat of the normal sample as a reference to estimate the number of false positives.

Figure 18A shows examples of variations found in the black6/B16 data, including a somatic mutation (left), non-somatic variation to the reference (middle), and possible false positive (right). Each somatic mutation can be associated with a quality score Q. The number of false positives in the tumor comparison indicates a number of false positives in the same vs. same comparison. Thus, for a given mutation with quality score Q detected in the tumor comparison, we estimate the false discovery rate by computing the ratio of same vs. same mutations with a score of Q or better to the overall number of mutations found in the tumor comparison with a score of Q or better.

A challenge arises in defining Q since most mutation detection frameworks compute multiple quality scores. Here, we apply a random forest classifier [Breiman, L., *Statist. Sci.* **16**, 199-231 (2001)] to combine multiple scores into a single quality score Q. We refer to the methods section for details regarding details of the quality score and FDR computation.

A potential bias in comparing methods is differential coverage; we thus normalize the false discovery rate for the coverage:

$$FDR(Q) = \frac{\#Same\ vs. Same\ SNVs\ with\ score\ \geq Q}{\#Tumor\ SNVs\ with\ score\ \ \geq Q}\ x\ \frac{\#common\ coverage\ tumor\ comparison}{\#common\ coverage\ same\ vs. same\ comparison}$$

We calculate the common coverage by counting all bases of the reference genome which are covered by both the tumor and normal sample or by both "same vs. same" samples, respectively.

By estimating the number of false positives and positives at each FDR (see Methods), we generate receiver operating characteristic (ROC) curves and calculate the AUC (area under the curve) for each mutation discovery method, thus enabling a comparison of strategies for mutation discovery (Fig. 18B).

Furthermore, the selection of the reference data might influence the calculation of the FDRs. Using the available black6/B16 data it is possible to create 18 triplets (combinations of black6 vs. black6 and black6 vs. b16). When comparing the resulting FDR distributions for the sets of somatic mutations, the results are consistent (Fig. 18B).

Using this definition of a false discovery rate, we have established a generic framework for evaluating the influence of numerous experimental and algorithmic parameters on the resulting set of somatic mutations. Next, we apply this framework to study the influence of software tools, coverage, paired end sequencing and the number of technical replicates on somatic mutation identification.

First, the choice of the software tool has a clear impact on the identified somatic mutations (Fig. 19A). On the tested data, SAMtools produces the highest enrichment of true positives in a set of somatic mutations ranked by the FDR. However, we note that all tools offer many parameters and quality scores for the individual mutations. Here, we have used the default settings as specified by the algorithm developers; we expect that the parameters could be optimized and emphasize that the FDR framework defined here is designed for running and evaluating such an optimization.

For the described B16 sequencing experiment, we sequenced each sample in an individual flowcell lane and achieved a target region mean base coverage of 38 fold for the individual samples. However, this coverage might not be needed to obtain an equally good set of somatic mutations, possibly reducing costs. Also, the impact of the depth of caverage on whole genome SNV detection has been discussed recently [Ajay, S.S. et al., Genome Res. 21, 1498-1505 (2011)]. In order to study the effect of the coverage on exon capture data, we downsampled the number of aligned sequence reads for every 1x50 nt library to generate an approximate coverage of 5, 10 and 20 fold, respectively, and then reapplied the mutation call algorithms. As expected, a higher coverage results in a better (i.e. fewer false positives) somatic mutation set, although the improvement from the 20 fold coverage to the maximum is marginal (Fig. 19B).

It is straightforward to simulate and rank different experimental settings using the available data and framework. Comparing duplicates to triplicates, triplicates do not offer a benefit compared to the duplicates (Fig. 19C), while duplicates offer a clear improvement compared to a study without any replicates. In terms of the ratio of somatic mutations in the given sets, we see enrichment at a FDR of 5% from 24.2% for a run without replicates to 71.2% for duplicates and 85.8% for triplicates. Despite the enrichment, using the intersection of triplicates removes more mutations with a low FDR than ones with a high FDR, as indicated by the lower ROC AUC and the shift of the curve to the left (Fig. 19C): the specificity is slightly increased at the cost of a lower sensitivity.

The additionally sequenced 2x100 nt library was used to simulate a 1x100, two 2x50 and two 1x50 nt libraries, respectively, by *in silicio* removal of the second read and/or the 3' and 5' ends of the reads, resulting in a total of 5 simulated libraries. These libraries were compared using the calculated FDRs of predicted mutations (Fig. 19D). Despite the much higher mean coverage (more than 77 vs. 38), the somatic mutations found using the 2x50 5' and 1x100 nt libraries have a lower ROC AUC and thus a worse FDR distribution than the 1x50 nt library. This phenomenon results from the accumulation of high FDR mutations in low coverage regions as the sets of low FDR mutations found are highly similar. The consequence is that the optimal sequencing length is either small so that the sequenced bases are concentrated around the capture probe sequences (potentially losing information on the somatic status of mutations in non-covered regions, though) or should be close to the fragment length (2x100 nt = 200 nt total length for ~250 nt fragments in our case), effectively filling up the coverage gaps. This is also supported by the ROC AUC of the 2x50 nt 3' library (simulated by using only the 3' ends of the 2x100 nt library) which is higher than the one of the 2x50 nt 5' library (simulated by using only the 5' ends of the 2x100 nt library) despite the lower base quality of the 3' read ends.

These observations allow us to define best practice procedures for the discovery of somatic mutations. Across all evaluated parameters, 20 fold coverage in both samples and using a technical duplicate achieves close to the optimum results in these relatively homogeneous samples, while also considering costs. A 1x50 nt library resulting in approximately 100 million reads seems to be the most pragmatic choice to achieve this coverage. This remains true across all possible dataset pairings. We retrospectively applied those parameter settings, used no additional filtering of the raw variant calls, and calculated the FDRs for 50 selected mutations from the intersection of all three methods as shown in Figure 17. All mutations were confirmed by a combination of Sanger resequencing and the B16 RNA-Seq sequence reads. 44 of those mutations would have been found using a FDR cutoff of 5% (Fig. 20). As a negative control, we re-sequenced the loci of 44 predicted mutations with high FDRs (> 50%) and examined the respective sequences in the RNA-Seq data. We found 37 of these mutations to be not validated while the remaining seven loci of potential mutations were both not covered by RNA-Seq reads and yielded in not sequencing reaction.

While we show application of the framework to four specific questions, it is by no means limited to these parameters, but can be applied to study the influence of all experimental or algorithmic parameters, e.g. the influence of the alignment software, the choice of a mutation metric, or the choice of vendor for exome selection.

We performed all experiments on a set of B16 melanoma cell experiments; however, the method is not restricted to these data. The only requirement is the availability of a 'same-vs-same' reference data set, meaning at least a single technical repeat of a non-tumorous sample should be performed for each new protocol. While our experiments indicate that the method is robust with regard to the choice of the technical repeat within certain limits, so that a repeat is not necessarily required in every single experiment. However, the method does require that the various quality measures are comparable between the reference data set and remaining datasets.

Within this contribution, we have pioneered a statistical framework for a false-discovery-rate driven detection of somatic mutations. This framework is not only applicable for the diagnostic or therapeutic target selection, but also allows a generic comparison of experimental and computational protocol steps on a generated quasi ground truth data. Here, we applied this idea to make protocol decisions with regard to software tools, coverage, replicates as well as paired end sequencing.

Methods

Library capture and sequencing

Next-generation sequencing, DNA sequencing: Exome capture for DNA resequencing was performed using the Agilent Sure-Select solution-based capture assay [Gnirke, A., *et al.*, *Nat. Biotechnol.* **27**, 182-189 (2009)], in this case designed to capture all known mouse exons.

3 μg purified genomic DNA was fragmented to 150-200 nt using a Covaris S2 ultrasound device. gDNA fragments were end repaired using T4 DNA polymerase, Klenow DNA polymerase and 5' phosphorylated using T4 polynucleotide kinase. Blunt ended gDNA fragments were 3' adenylated using Klenow fragment (3' to 5' exo minus). 3' single T-overhang Illumina paired end adapters were ligated to the gDNA fragments using a 10:1 molar ratio of adapter to genomic DNA insert using T4 DNA ligase. Adapter ligated gDNA fragments were enriched pre capture and flow cell specific sequences were added using Illumina PE PCR primers 1.0 and 2.0 and Herculase II polymerase (Agilent) using 4 PCR cycles.

500 ng of adapter ligated, PCR enriched gDNA fragments were hybridized to Agilent's SureSelect biotinylated mouse whole exome RNA library baits for 24 hrs at 65 °C. Hybridized gDNA/RNA bait complexes where removed using streptavidin coated magnetic beads. gDNA/RNA bait complexes were washed and the RNA baits cleaved off during elution in SureSelect elution buffer leaving the captured adapter ligated, PCR enriched gDNA fragments. gDNA fragments were PCR amplified post capture using Herculase II DNA polymerase (Agilent) and SureSelect GA PCR Primers for 10 cycles.

Cleanups were performed using 1.8x volume of AMPure XP magnetic beads (Agencourt). For quality controls we used 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 and sequenced on the Illumina HiSeq2000 using Truseq SBS kit.

Exome data analysis

Sequence reads were aligned using bwa (version 0.5.8c) [Li, H. Durbin, R., *Bioinformatics* 25, 1754-1760 (2009)] using default options to the reference mouse genome assembly mm9 [Mouse Genome Sequencing Consortium, *Nature* 420, 520-562 (2002)]. Ambiguous reads –

those reads mapping to multiple locations of the genome as provided by the bwa output - were removed. The remaining alignments were sorted, indexed and converted to a binary and compressed format (BAM) and the read quality scores converted from the Illumina standard phred÷64 to standard Sanger quality scores using shell scripts.

For each sequencing lane, mutations were identified using three software programs: SAMtools pileup (version 0.1.8) [Li, H. *et al.*, *Bioinformatics* 25, 2078-2079 (2009)], GATK (version 1.0.4418) [DePristo, M.A. *et al.*, *Nature Genetics* 43, 491–498 (2011)], and SomaticSniper [Ding, L. et al., *Hum. Mol. Genet* (2010) first published online September 15, 2010]. For SAMtools, the author-recommend options and filter criteria were used, including first round filtering, maximum coverage 200. For SAMtools second round filtering, the minimum indel quality score was 50, the point mutation minimum quality was 30. For GATK mutation calling, we followed the author-designed best practice guidelines presented on the GATK user manual. For each sample a local realignment around indel sites followed by a base quality recalibration was performed. The UnifiedGenotyper module was applied to the resultant alignment data files. When needed, the known polymorphisms of the dbSNP [Sherry, S.T. *et al.*, *Nucleic Acids Res.* 29, 308-311 (2009)] (version 128 for mm9) were supplied to the individual steps. The variant score recalibration step was omitted and replaced by the hard-filtering option. For SomaticSniper mutation calling, the default options were used and only predicted mutations with a "somatic score" of 30 or more were considered further. Additionally, for each potentially mutated locus we required a non-zero coverage in the normal tissue and removed all mutations located in repetitive sequences as defined by the RepeatMasker track of the UCSC Genome Browser for the mouse genome assembly mm9 [Fujita, P.A. *et al.*, *Nucleic Acids Res.* 39, 876-882 (2011)].

RNA-Seq

Barcoded mRNA-seqcDNA libraries were prepared from 5 ug of total RNA using a modified version of the Illumina mRNA-seq protocol. mRNA was isolated using SeramagOligo(dT) magnetic beads (Thermo Scientific). Isolated mRNA was fragmented using divalent cations and heat resulting in fragments ranging from 160-200 bp. Fragmented mRNA was converted to cDNA using random primers and SuperScriptII (Invitrogen) followed by second strand synthesis using DNA polymerase I and RNaseH. cDNA was end repaired using T4 DNA polymerase, Klenow DNA polymerase and 5' phosphorylated using T4 polynucleotide kinase. Blunt ended cDNA fragments were 3' adenylated using Klenow fragment (3' to 5' exo minus). 3' single T-overhang Illumina multiplex specific adapters were ligated on the cDNA fragments using T4 DNA ligase. cDNA libraries were purified and size selected at 300 bp using the 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 (Finnzymes). All cleanups were performed using 1,8x volume of Agencourt AMPure XP magnetic beads.

Barcoded RNA-seq libraries were clustered on the cBot using Truseq SR cluster kit v2.5 using 7 pM and sequenced on the Illumina HiSeq2000 using Truseq SBS kit.

The raw output data of the HiSeq was processed according to the Illumina standard protocol, including removal of low quality reads and demultiplexing. Sequence reads were then aligned to the reference genome sequence [Mouse Genome Sequencing Consortium, *Nature* 420, 520-562 (2002)] using bowtie [Langmead, B. et al., *Genome Biol.* **10**, R25 (2009)]. The alignment coordinates were compared to the exon coordinates of the RefSeq transcripts [Pruitt, K.D. et al., *Nucleic Acids Res.* **33**, 501-504 (2005)] and for each transcript the counts of overlapping alignments were recorded. Sequence reads not aligning to the genomic sequence were aligned to a database of all possible exon-exon junction sequences of the RefSeq transcripts [Pruitt, K.D. et al., *Nucleic Acids Res.* **33**, 501-504 (2005)]. The alignment coordinates were compared to RefSeq exon and junction coordinates, reads counted, and normalized to RPKM (number of reads which map per nucleotide kilobase of transcript per million mapped reads [Mortazavi, A. et al., *Nat. Methods* **5**, 621-628 (2008)]) for each transcript.

Validation of SNVs

We selected SNVs for validation by Sanger re-sequencing and RNA. SNVs were identified which were predicted by all three programs, non-synonymous, and found in transcripts having a minimum 10 RPKM. Of these, we selected the 50 with the highest SNP quality scores as provided by the programs. As a negative control, 44 SNVs were selected which have a FDR of 50% or more, are present in only one cell line sample and are predicted by only one mutation calling program. Using DNA, the selected variants were validated by PCR amplification of the regions using 50 ng of DNA, followed by Sanger sequencing (Eurofins MWG Operon, Ebersberg, Germany). The reactions were successful for 50 and 32 loci of positive and negative controls, respectively. Validation was also done by examination of the tumor RNA-Seq reads.

Calculation of FDRs and machine learning

Random Forest Quality Score Computation: Commonly-used mutation calling algorithms (DePristo, M.A. *et al.*, *Nature Genetics* 43, 491–498 (2011), Li, H. *et al.*, *Bioinformatics* 25, 2078-2079 (2009), Ding, L. et al., *Hum. Mol. Genet* (2010) first published online September 15, 2010) output multiple scores, which all are potentially influential for the quality of the mutation call. These include - but are not limited to - the quality of the base of interest as assigned by the instrument, the quality alignment for this position, the number of reads covering this position or a score for the difference between the two genomes compared at this position. For the computation of the false discovery rate we require an ordering of mutations, however this is not directly feasible for all mutations since we might have contradicting information from the various quality scores.

We use the following strategy to achieve a complete ordering. In a first step, we apply a very rigorous definition of superiority by assuming that a mutation has better quality than another if and only if it is superior in all categories. So a set of quality properties S=($s_1,\ldots,s_n$) is preferable to T=($t_1,\ldots,t_n$), denoted by S>T, iff $s_i > t_i$ for all i=1,...,n. We define an intermediate FDR (IFDR) as follows

$$IFDR(T) = \frac{\#Same\ vs.Same\ SNVs\ with\ score\ S > T}{\#Tumor\ SNVs\ with\ score\ \ S > T}\ x\ \frac{\#common\ coverage\ tumor\ comparison}{\#common\ coverage\ same\ vs.same\ comparison}$$

However, we regard the IFDR only as an intermediate step since in many closely related cases, no comparison is feasible and we are thus not benefitting from the vast amount of data available. Thus, we take advantage of the good generalization property of random forest regression [Breiman, L., Statist. Sci. 16, 199-231 (2001)] and train a random forest as implemented in R (R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2010, Liaw, A., Wiener, M., *R News* **2**, 18-22 (2002)).

For *m* input mutations with *n* quality properties each, the value range for each property was determined and up to *p* values were sampled with uniform spacing out of this range; when the set of values for a quality property was smaller than *p*, this set was used instead of the sampled set. Then each possible combination of sampled or selected quality values is created, which results in a maximum of p^n data points in the *n*-dimensional quality space. A random sample of 1% of these points and the corresponding IFDR values were used as predictor and response, respectively, for the random forest training.

The resulting regression score is our generalized quality score Q; it can be regarded as a locally weighted combination of the individual quality scores. It allows direct, single value comparison of any two mutations and the computation of the actual false discovery rate:

$$FDR(Q) = \frac{\#Same\ vs. Same\ SNVs\ with\ score\ \geq Q}{\#Tumor\ SNVs\ with\ score\ \geq Q} x \frac{\#common\ coverage\ tumor\ comparison}{\#common\ coverage\ same\ vs.\ same\ comparison}$$

For the training of the random forest model used to create the results for this study, we calculate the sample IFDR on the somatic mutations of all samples before selecting the random 1% subset. This ensures the mapping of the whole available quality space to FDR values. We used the quality properties “SNP quality”, “coverage depth”, “consensus quality” and “RMS mapping quality” (SAMtools, $p = 20$); “SNP quality”, “coverage depth”, “Variant confidence/unfiltered depth” and “RMS mapping quality” (GATK, $p = 20$); or SNP quality”, “coverage depth”, “consensus quality”, “RMS mapping quality” and “somatic score” (SomaticSNiPer, $p = 12$), respectively. The different values of p ensure a set size of comparable magnitude.

Common coverage computation: The number of possible mutation calls can introduce a major bias in the definition of a false discovery rate. Only if we have the same number of possible locations for mutations to occur for our tumor comparison and for our same vs. same comparison, the number of called mutations is comparable and can serve as a basis for a false discovery rate computation. To correct for this potential bias, we use the common coverage ratio. As common coverage we define the number of bases with coverage of at least one in both samples which are used for the mutation calling. We compute the common coverage individually for the tumor comparison as well as for the same vs. same comparison.

ROC estimation

Receiver operating characteristic (ROC) curves and the corresponding area under curve (AUC) are useful for organizing classifiers and visualizing their performance [Fawcett, T., Pattern Recogn. Lett. 27, 861–874 (2006)]. We extend this concept for evaluating the performance of experimental and computational procedures. However, plotting ROC graphs requires knowledge of all true and false positive (TP and FP) examples in a dataset, information which is usually not given and hard to establish for high throughput data (such as NGS data). Thus, we use the calculated FDRs to estimate the respective TP and FP rates and plot a ROC graph and calculate an AUC. The central idea is that the FDR of a single mutation in the dataset gives the proportion how much this mutation contributes to the sum of TP/FP mutations, respectively. Also, for a list of random assignments to TP and FP, the resultant

ROC AUC will be equal to 0.5 with our method, indicating a completely random prediction. We start with two conditions:

$$FDR = \frac{FPR}{FPR + TPR} \qquad [1]$$

and

$$FPR + TPR = 1 \qquad [2]$$

with *FPR* and *TPR* being the needed false positive true positive ratios, respectively, for the given mutation, defining the corresponding point in ROC space. [1] and [2] can be rearranged to

$$TPR = 1 - FPR \qquad [3]$$

and

$$FPR = FDR \qquad [4]$$

To obtain an estimated ROC curve, the mutations in dataset are sorted by FDR and for each mutation a point is plotted at the cumulative TPR and FPR values up to this mutation, divided by the sum of all TPR and TPR values, respectively. The AUC is calculated by summing up the areas of all consecutive trapezoids between the curve and the x-axis.

CLAIMS:

1. A method for preparing an individualized cancer vaccine comprising the steps:

(a) identifying in a tumor specimen of a cancer patient cancer specific somatic mutations to provide a cancer mutation signature of the patient comprising cancer specific somatic mutations, wherein the cancer cells of the patient express antigens comprising said mutations and wherein said mutations are not found in non-cancerous cells of the patient; and

(b) preparing an RNA vaccine featuring the cancer mutation signature obtained in step (a), wherein the RNA vaccine comprises RNA encoding a recombinant polyepitopic polypeptide comprising neo-epitopes comprising said mutations fused together by peptide bonds or linkers.

2. The method according to claim 1, wherein the step of identifying cancer specific somatic mutations comprises identifying the cancer mutation signature of the exome of one or more cancer cells.

3. The method according to 1 or 2, wherein the step of identifying cancer specific somatic mutations comprises single cell sequencing of one or more cancer cells.

4. The method according to claim 3, wherein the cancer cells are circulating tumor cells.

5. The method according to any one of claims 1 to 4, wherein the step of identifying cancer specific somatic mutations involves using next generation sequencing (NGS).

6. The method according to any one of claims 1 to 5, wherein the step of identifying cancer specific somatic mutations comprises sequencing genomic DNA and/or RNA of the tumor specimen.

7. The method according to claim 6, wherein the step of identifying cancer specific somatic mutations is replicated at least in duplicates.

8. The method according to any one of claims 1 to 7, comprising the further step of determining the usability of the identified mutations in epitopes for cancer vaccination, wherein determining the usability comprises one or more of the following: (i) assessing whether the

identified mutations are located in known or predicted MHC presented epitopes, (ii) *in vitro* and/or *in silico* testing whether the identified mutations are located in MHC presented epitopes, and (iii) *in vitro* testing whether the identified mutations are able to stimulate T cells of the patient having the desired specificity.

9. The method according to any one of claims 1 to 8, wherein the polypeptide comprises 5 or more, 10 or more, or 20 or more neo-epitopes.

10. The method according to any one of claims 1 to 9, wherein the polypeptide comprises up to 30 neo-epitopes.

11. The method according to any one of claims 1 to 10, wherein the polypeptide further comprises epitopes not containing cancer specific somatic mutations which are expressed by cancer cells.

12. The method according to any one of claims 1 to 11, wherein the polypeptide comprises neo-epitopes that are based on primary basal mutations.

13. The method according to any one of claims 1 to 12, wherein the neo-epitopes are flanked by amino acid sequences flanking said neo-epitopes in the naturally occurring protein so as to form a vaccine sequence.

14. The method according to any one of claims 1 to 13, wherein the neo-epitopes are lined up head-to-tail.

15. The method according to any one of claims 1 to 14, wherein the neo-epitopes are spaced by linkers.

16. The method according to any one of claims 1 to 15, wherein the RNA further comprises an unmasked poly-A sequence.

17. The method according to claim 16, wherein the unmasked poly-A sequence has a length of approximately 120 adenosine residues.

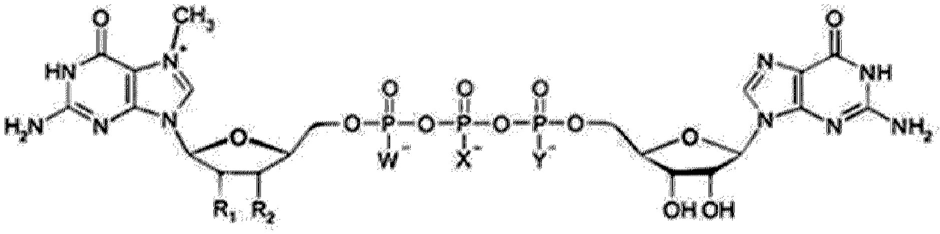
18. The method according to any one of claims 1 to 17, wherein the RNA further comprises a 3' UTR.

19. The method according to any one of claims 1 to 18, wherein the recombinant polyepitopic polypeptide encoded by the RNA further comprises an MHC class I signal peptide fragment, a transmembrane domain or a cytosolic domain.

20. The method according to any one of claims 1 to 19, wherein at least 50 % of the amino acids of the linkers are glycine and/or serine residues.

21. The method according to any one of claims 1 to 20, wherein the RNA further comprises a 5' cap.

22. The method according to claim 21, wherein the 5' cap is:



,

wherein R_1 and R_2 are independently hydroxy or methoxy, and W, X, and Y are independently oxygen, sulfur, selenium or BH_3.

23. The method according to claim 22 , wherein R_1 and R_2 are hydroxy, and W, X, and Y are oxygen.

24. The method according to claim 22, wherein one of R_1 and R_2 is hydroxy, the other one is methoxy, and W, X, and Y are oxygen.

25. The method according to claim 22, wherein X is sulfur.

26. The method according to claim 25, wherein W and Y are oxygen.

27. The method according to claim 21 or 22, wherein the 5' cap is $m_2^{7,2'-O}Gpp_spG$.

28. The method according to claim 27, wherein $\mathrm{m}_2^{7,2'\text{-}\mathrm{O}}\mathrm{Gpp}_\mathrm{s}\mathrm{pG}$ is in the Rp diastereoisomeric form.

29. The method according to claim 27, wherein $\mathrm{m}_2^{7,2'\text{-}\mathrm{O}}\mathrm{Gpp}_\mathrm{s}\mathrm{pG}$ is in the Sp diastereoisomeric form.

30. An RNA vaccine which is obtained by the method according to any one of claims 1 to 29, wherein the vaccine, when administered to the patient of claim 1, provides a collection of MHC presented epitopes incorporating sequence changes based on the identified mutations.

31. The RNA vaccine according to claim 30, wherein the MHC presented epitopes are MHC class II-presented epitopes that elicit a CD4+ helper T cell response against cells expressing antigens from which the MHC presented epitopes are derived and/or MHC class I-presented epitopes that elicit a CD8+ T cell response against cells expressing antigens from which the MHC presented epitopes are derived.

32. An RNA vaccine comprising RNA encoding a recombinant polypeptide comprising neo-epitopes fused together by peptide bonds or linkers, said neo-epitopes resulting from cancer specific somatic mutations in a tumor specimen of a cancer patient.

33. The RNA vaccine according to claim 32, wherein the polypeptide comprises 5 or more, 10 or more, or 20 or more neo-epitopes.

34. The RNA vaccine according to claim 32 or 33, wherein the polypeptide comprises up to 30 neo-epitopes.

35. The RNA vaccine according to any one of claims 32 to 34, wherein the polypeptide further comprises epitopes not containing cancer specific somatic mutations which are expressed by cancer cells.

36. The RNA vaccine according to any one of claims 32 to 35, wherein the polypeptide comprises neo-epitopes that are based on primary basal mutations.

37. The RNA vaccine according to any one of claims 32 to 36, wherein the neo-epitopes are flanked by amino acid sequences flanking said neo-epitopes in the naturally occurring protein so as to form a vaccine sequence.

38. The RNA vaccine according to any one of claims 32 to 37, wherein the neo-epitopes are lined up head-to-tail.

39. The RNA vaccine according to any one of claims 32 to 38, wherein the neo-epitopes are spaced by linkers.

40. The RNA vaccine according to any one of claims 32 to 39, wherein the vaccine, when administered to the patient of claim 32, provides a collection of MHC presented epitopes incorporating sequence changes based on the tumor specific mutations.

41. The RNA vaccine according to claim 40, wherein the MHC presented epitopes are MHC class II-presented epitopes that elicit a CD4+ helper T cell response against cells expressing antigens from which the MHC presented epitopes are derived and/or MHC class I-presented epitopes that elicit a CD8+ T cell response against cells expressing antigens from which the MHC presented epitopes are derived.

42. The RNA vaccine according to any one of claims 32 to 41, wherein the RNA further comprises an unmasked poly-A sequence.

43. The RNA vaccine according to claim 42, wherein the unmasked poly-A sequence has a length of approximately 120 adenosine residues.

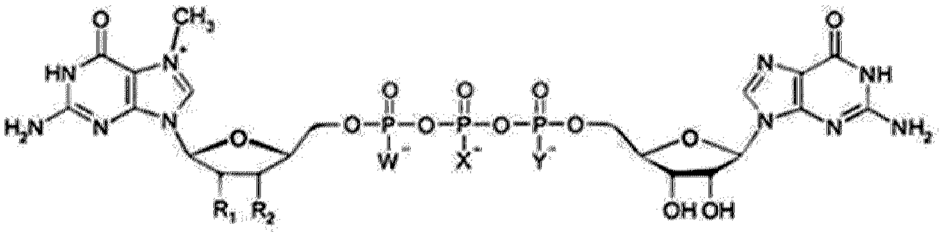
44. The RNA vaccine according to any one of claims 32 to 43, wherein the RNA further comprises a 3' UTR.

45. The RNA vaccine according to any one of claims 32 to 44, wherein the polyepitopic polypeptide encoded by the RNA further comprises an MHC class I signal peptide fragment, a transmembrane domain, or a cytosolic domain.

46. The RNA vaccine according to any one of claims 32 to 45, wherein at least 50 % of the amino acids of the linkers are glycine and/or serine residues.

47. The RNA vaccine according to any one of claims 32 to 46, wherein the RNA further comprises a 5' cap.

48. The RNA vaccine according to claim 47, wherein the 5' cap is:



,

wherein R_1 and R_2 are independently hydroxy or methoxy, and W, X, and Y are independently oxygen, sulfur, selenium or BH_3.

49. The RNA vaccine according to claim 48, wherein R_1 and R_2 are hydroxy, and W, X, and Y are oxygen.

50. The RNA vaccine according to claim 48, wherein one of R_1 and R_2 is hydroxy, the other one is methoxy, and W, X, and Y are oxygen.

51. The RNA vaccine according to claim 48, wherein X is sulfur.

52. The RNA vaccine according to claim 51, wherein W and Y are oxygen.

53. The RNA vaccine according to claim 47 or 48, wherein the 5' cap is $m_2^{7,2'\text{-}O}Gpp_spG$.

54. The RNA vaccine according to claim 53, wherein $m_2^{7,2'\text{-}O}Gpp_spG$ is in the Rp diastereoisomeric form.

55. The RNA vaccine according to claim 53, wherein $m_2^{7,2'\text{-}O}Gpp_spG$ is in the Sp diastereoisomeric form.

56. A method for preparing an individualized cancer vaccine comprising the steps:

(a) identifying cancer specific somatic mutations in a tumor specimen of a cancer patient to provide a cancer mutation signature of the cancer patient comprising cancer specific somatic mutations, comprising

(aa) obtaining nucleic acid sequence information by sequencing genomic DNA and/or RNA of the tumor specimen of the cancer patient,

(bb) obtaining reference nucleic acid sequence information by sequencing DNA and/or RNA of normal non-cancerous cells, and

(cc) comparing the nucleic acid sequence information from the tumor specimen obtained in step (aa) with the reference nucleic acid sequence information obtained in step (bb); and

(b) preparing an RNA vaccine featuring the cancer mutation signature obtained in step (a), wherein the RNA vaccine comprises RNA encoding a recombinant polyepitopic polypeptide comprising neo-epitopes fused together by peptide bonds or linkers.

57. The method according to claim 56, wherein the step of identifying cancer specific somatic mutations comprises identifying the cancer mutation signature of the exome of one or more cancer cells.

58. The method according to claim 56 or 57, wherein the step of identifying cancer specific somatic mutations comprises single cell sequencing of one or more cancer cells.

59. The method according to claim 58, wherein the cancer cells are circulating tumor cells.

60. The method according to any one of claims 56 to 59, wherein the step of identifying cancer specific somatic mutations involves using next generation sequencing (NGS).

61. The method according to any one of claims 56 to 60, wherein the normal non-cancerous cells are obtained from the cancer patient.

62. The method according to any one of claims 56 to 61, wherein the reference nucleic acid sequence information is obtained from genomic DNA obtained from peripheral blood mononuclear cells (PBMCs).

63. The method according to claim 62, wherein the tumor specimen is from a primary tumor, and wherein the step of identifying cancer specific somatic mutations in the tumor specimen further comprises the steps:

(dd) preparing a phylogenetic tree of cancer specific somatic mutations, wherein the reference nucleic acid sequence information obtained in step (bb) is used to root the tree,

(ee) reproducing ancestral sequences, wherein the ancestral sequences are sequences of nodes near the root of the phylogenetic tree containing primary basal mutations, wherein the primary basal mutations are the earliest mutations predicted to exist in the primary tumor; and

(ff) selecting the primary basal mutations from the ancestral sequences identified in step (ee).

64. The method according to any one of claims 56 to 63, wherein the step of identifying cancer specific somatic mutations is replicated at least in duplicates.

65. The method according to any one of claims 56 to 64, comprising the further step of determining the usability of the identified mutations in epitopes for cancer vaccination, wherein determining the usability comprises one or more of the following: (i) assessing whether the identified mutations are located in known or predicted MHC presented epitopes, (ii) *in vitro* and/or *in silico* testing whether the identified mutations are located in MHC presented epitopes, and (iii) *in vitro* testing whether the identified mutations are able to stimulate T cells of the patient having the desired specificity.

66. The method according to any one of claims 56 to 65, wherein the polypeptide comprises 5 or more, 10 or more, or 20 or more neo-epitopes.

67. The method according to any one of claims 56 to 66, wherein the polypeptide comprises up to 30 neo-epitopes.

68. The method according to any one of claims 56 to 67, wherein the polypeptide further comprises epitopes not containing cancer specific somatic mutations which are expressed by cancer cells.

69. The method according to any one of claims 56 to 68, wherein the polypeptide comprises neo-epitopes that are based on primary basal mutations.

70. The method according to any one of claims 56 to 69, wherein the neo-epitopes are flanked by amino acid sequences flanking said neo-epitopes in the naturally occurring protein so as to form a vaccine sequence.

71. The method according to any one of claims 56 to 70, wherein the neo-epitopes are lined up head-to-tail and/or are spaced by linkers.

72. The method according to any one of claims 56 to 71, wherein the RNA further comprises an unmasked poly-A sequence.

73. The method according to claim 72, wherein the unmasked poly-A sequence has a length of approximately 120 adenosine residues.

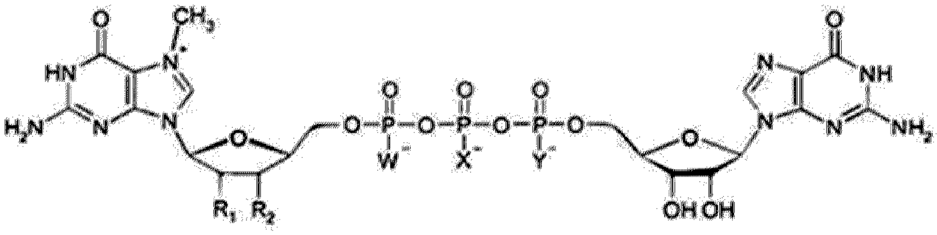
74. The method according to any one of claims 56 to 73, wherein the RNA further comprises a 3' UTR.

75. The method according to any one of claims 56 to 74, wherein the polyepitopic polypeptide encoded by the RNA further comprises an MHC class I signal peptide fragment, a transmembrane domain or a cytosolic domain.

76. The method according to any one of claims 56 to 75, wherein at least 50 % of the amino acids of the linkers are glycine and/or serine residues.

77. The method according to any one of claims 56 to 76, wherein the RNA further comprises a 5' cap.

78. The method according to claim 77, wherein the 5' cap is:



,

wherein R_1 and R_2 are independently hydroxy or methoxy, and W, X, and Y are independently oxygen, sulfur, selenium or BH_3.

79. The method according to claim 78, wherein R_1 and R_2 are hydroxy, and W, X, and Y are oxygen.

80. The method according to claim 78, wherein one of R_1 and R_2 is hydroxy, the other one is methoxy, and W, X, and Y are oxygen.

81. The method according to claim 78, wherein X is sulfur.

82. The method according to claim 81, wherein W and Y are oxygen.

83. The method according to claim 77 or 78, wherein the 5' cap is $m_2^{7,2'\text{-}O}Gpp_spG$.

84. The method according to claim 83, wherein $m_2^{7,2'\text{-}O}Gpp_spG$ is in the Rp diastereoisomeric form.

85. The method according to claim 83, wherein $m_2^{7,2'\text{-}O}Gpp_spG$ is in the Sp diastereoisomeric form.

86. An RNA vaccine which is obtained by the method according to any one of claims 56 to 85, wherein the vaccine, when administered to the patient of claim 56, provides a collection of MHC presented epitopes incorporating sequence changes based on the identified mutations.

87. The RNA vaccine according to claim 86, wherein the MHC presented epitopes are MHC class II-presented epitopes that elicit a CD4+ helper T cell response against cells expressing antigens from which the MHC presented epitopes are derived and/or MHC class I-presented

epitopes that elicit a CD8+ T cell response against cells expressing antigens from which the MHC presented epitopes are derived.

88. Use of the individualized cancer vaccine prepared by the method according to any one of claims 56 to 85 for treating the cancer of said patient.

89. Use of the RNA vaccine according to claim 86 or 87 for treating the cancer of said patient.

90. Use of the individualized cancer vaccine prepared by the method according to any one of claims 1 to 29 for treating the cancer of said patient.

91. Use of the RNA vaccine according to any one of claims 30 to 55 for treating the cancer of said patient.

Fig. 1

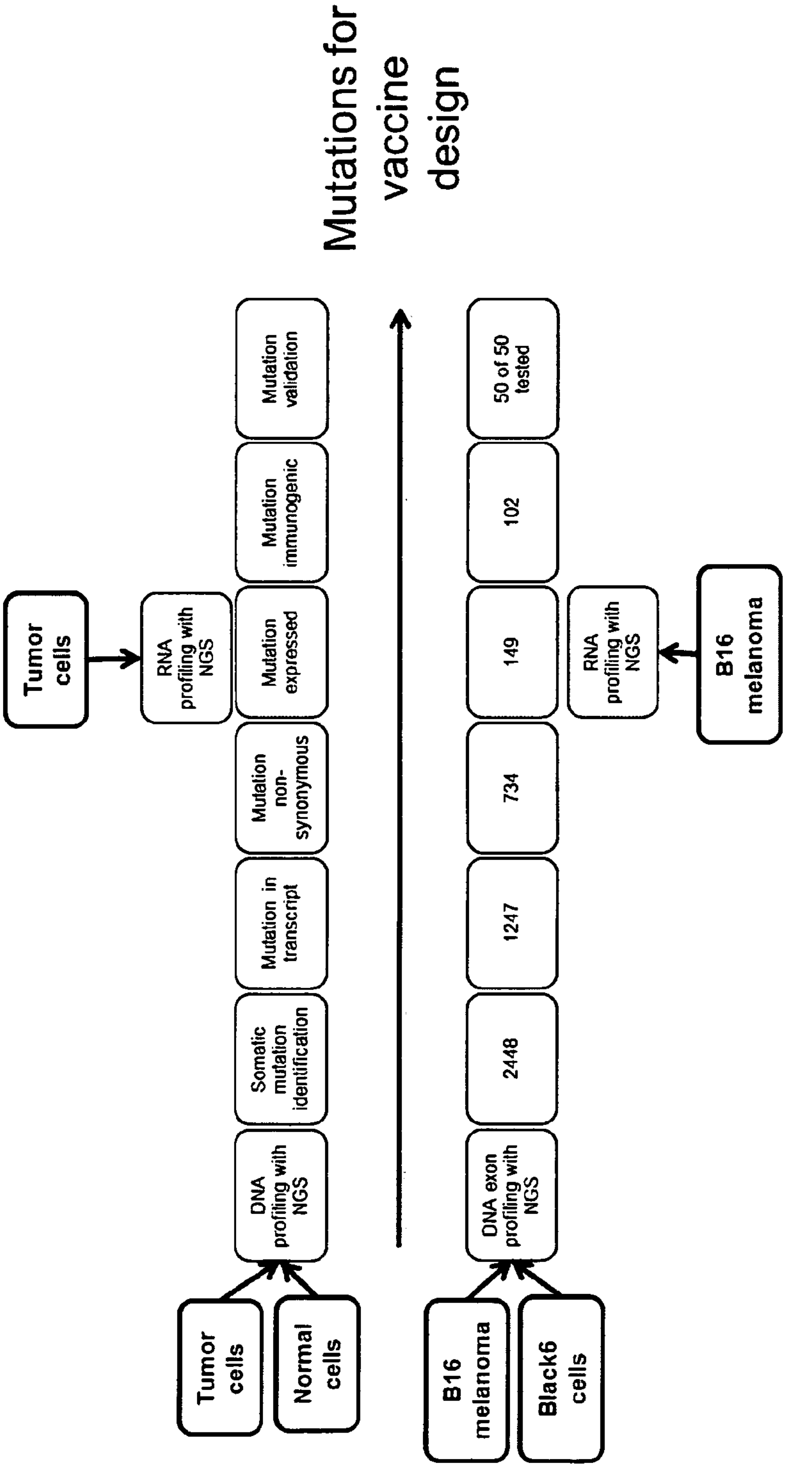

Tumor cells
Normal cells
DNA profiling with NGS
Somatic mutation identification
Mutation in transcript
Mutation non-synonymous
Tumor cells
RNA profiling with NGS
Mutation expressed
Mutation immunogenic
Mutation validation
Mutations for vaccine design
B16 melanoma
Black6 cells
DNA exon profiling with NGS
2448
1247
734
B16 melanoma
RNA profiling with NGS
149
102
50 of 50 tested

Fig. 2

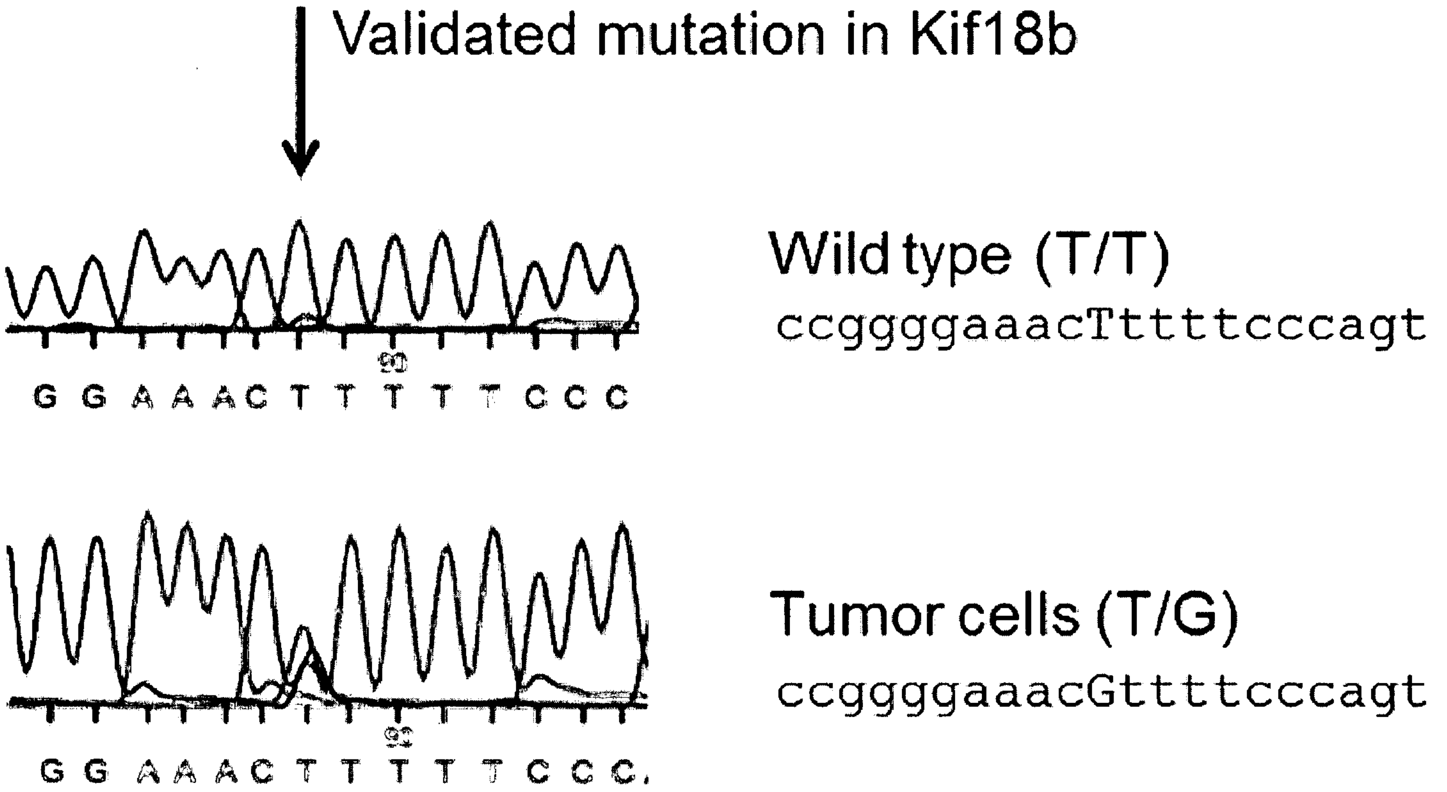
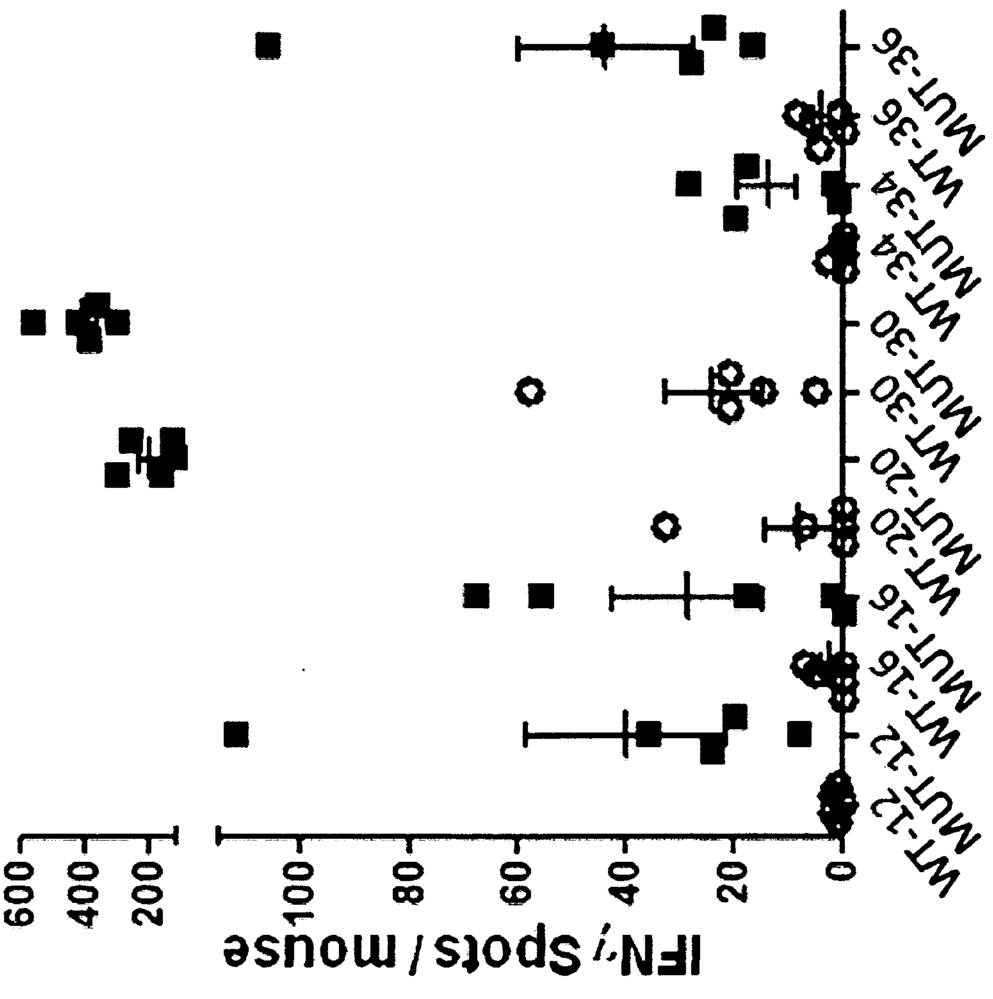

Validated mutation in Kif18b
Wild type (T/T)
ccggggaaacTttttcccagt
G G A A A C T T T T T C C C
Tumor cells (T/G)
ccggggaaacGttttcccagt
G G A A A C T T T T T C C C

Fig. 3


IFNγ Spots / mouse
600
400
200
100
80
60
40
20
0
WT-12
MUT-12
WT-16
MUT-16
WT-20
MUT-20
WT-30
MUT-30
WT-34
MUT-34
WT-36
MUT-36

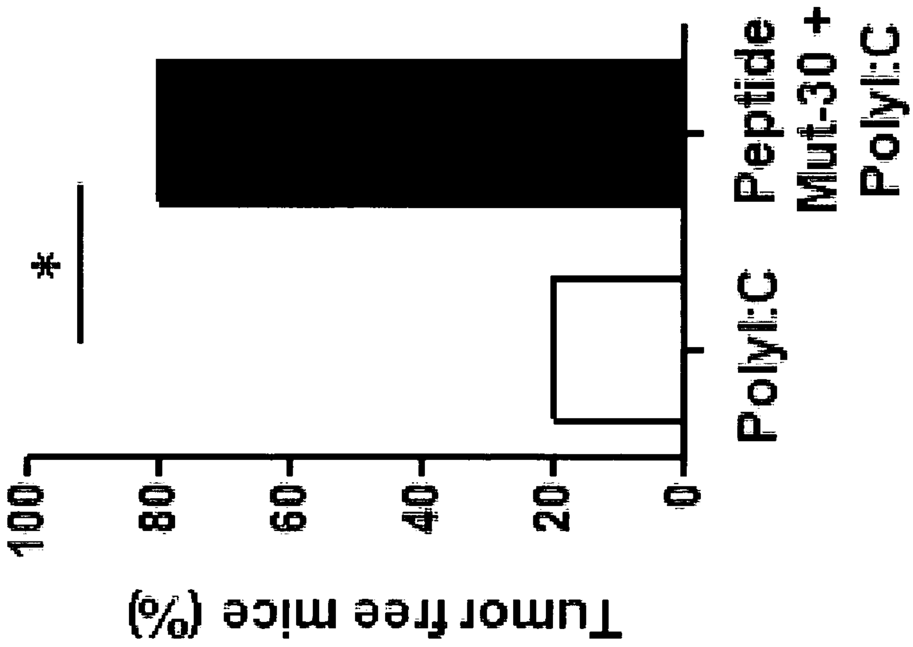


Fig. 4

Fig. 5

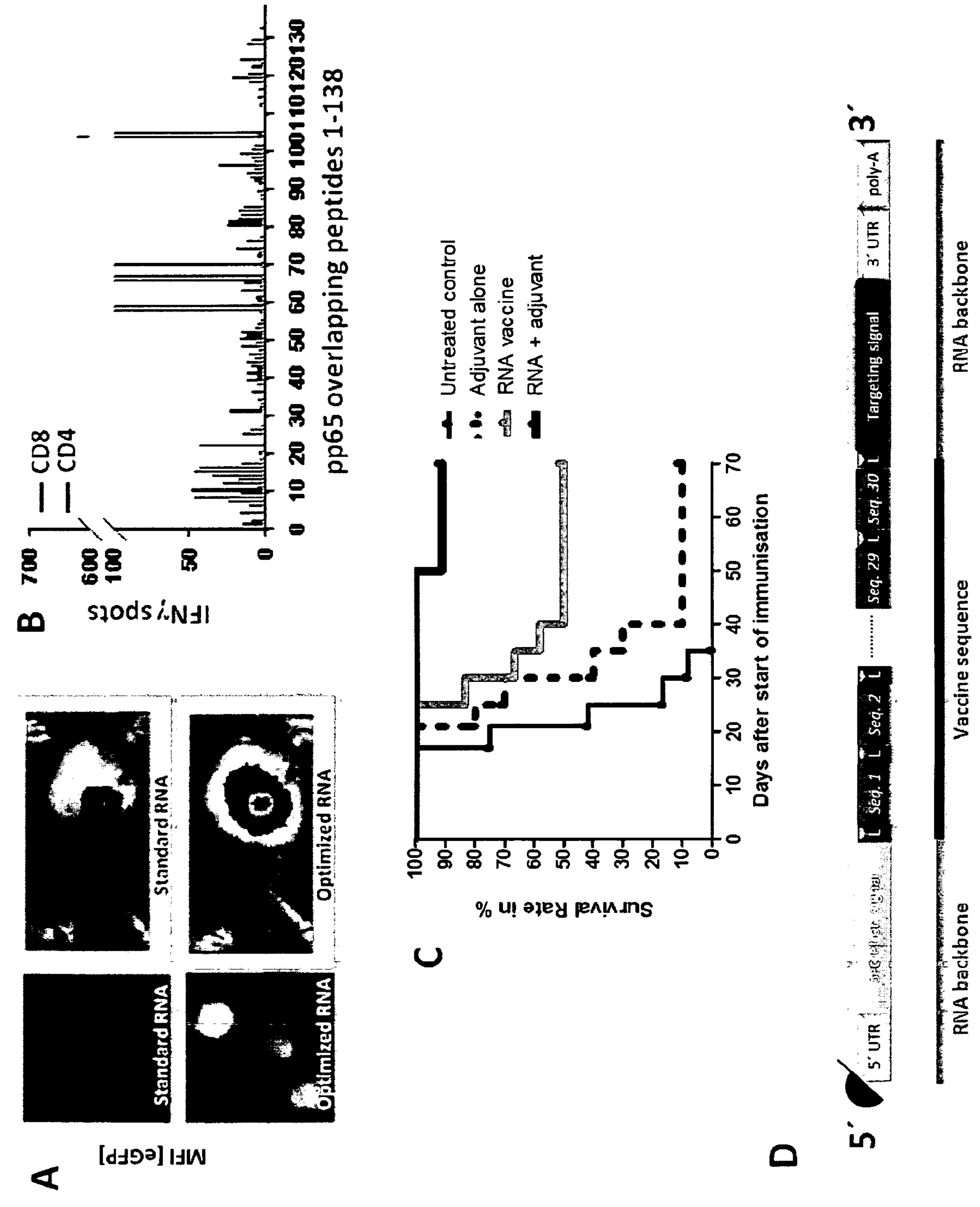

A
MFI [eGFP]
Standard RNA
Optimized RNA
Standard RNA
Optimized RNA
B
IFNγ spots
700
600
100
50
0
CD8
CD4
0 10 20 30 40 50 60 70 80 90 100 110 120 130
pp65 overlapping peptides 1-138
C
Survival Rate in %
100
90
80
70
60
50
40
30
20
10
0
0 10 20 30 40 50 60 70
Days after start of immunisation
Untreated control
Adjuvant alone
RNA vaccine
RNA + adjuvant
D
5´
5´ UTR
L
Seq. 1
L
Seq. 2
L
Seq. 29
L
Seq. 30
L
Targeting signal
3´ UTR
poly-A
3´
RNA backbone
Vaccine sequence
RNA backbone

Fig. 6

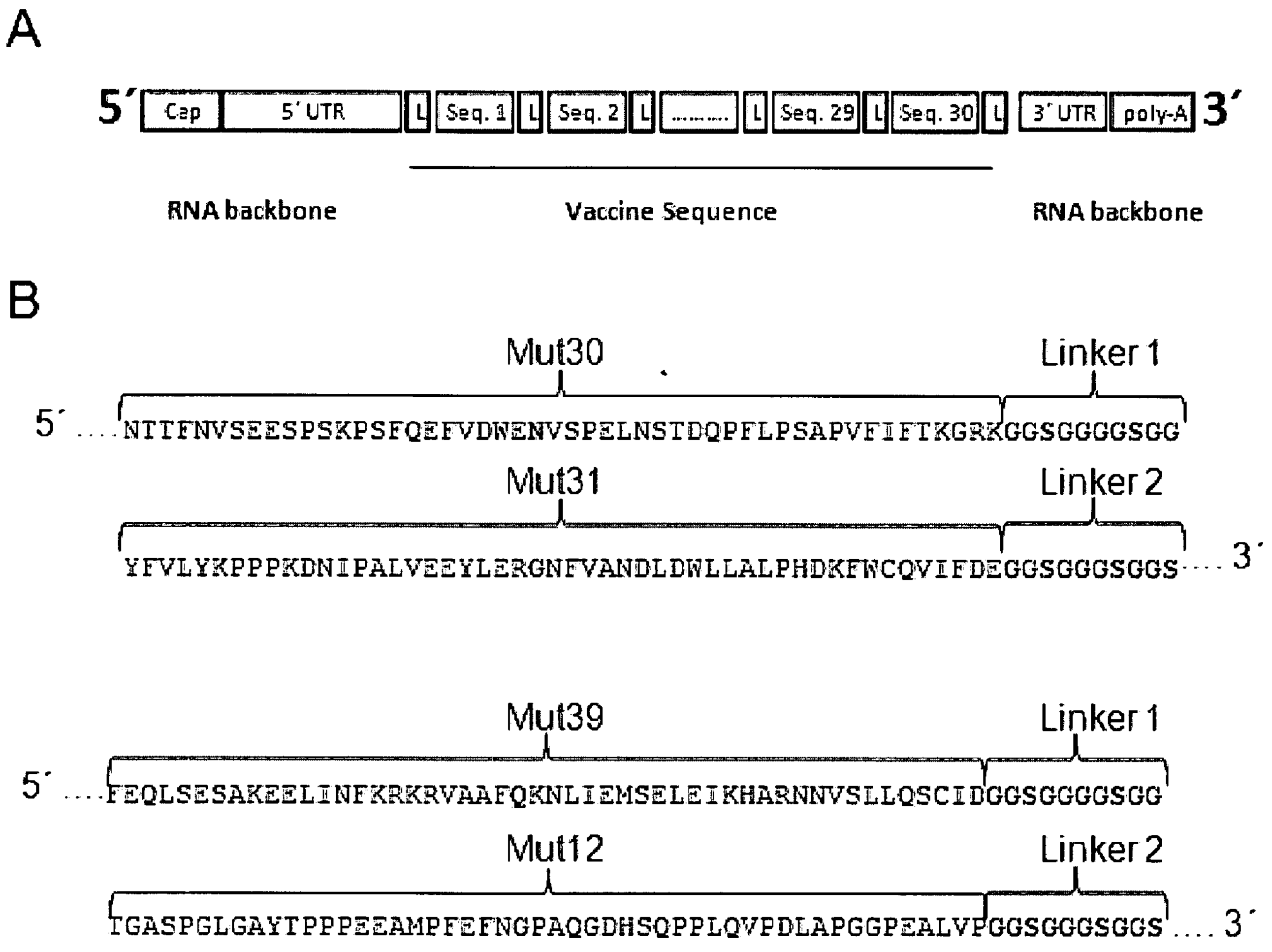
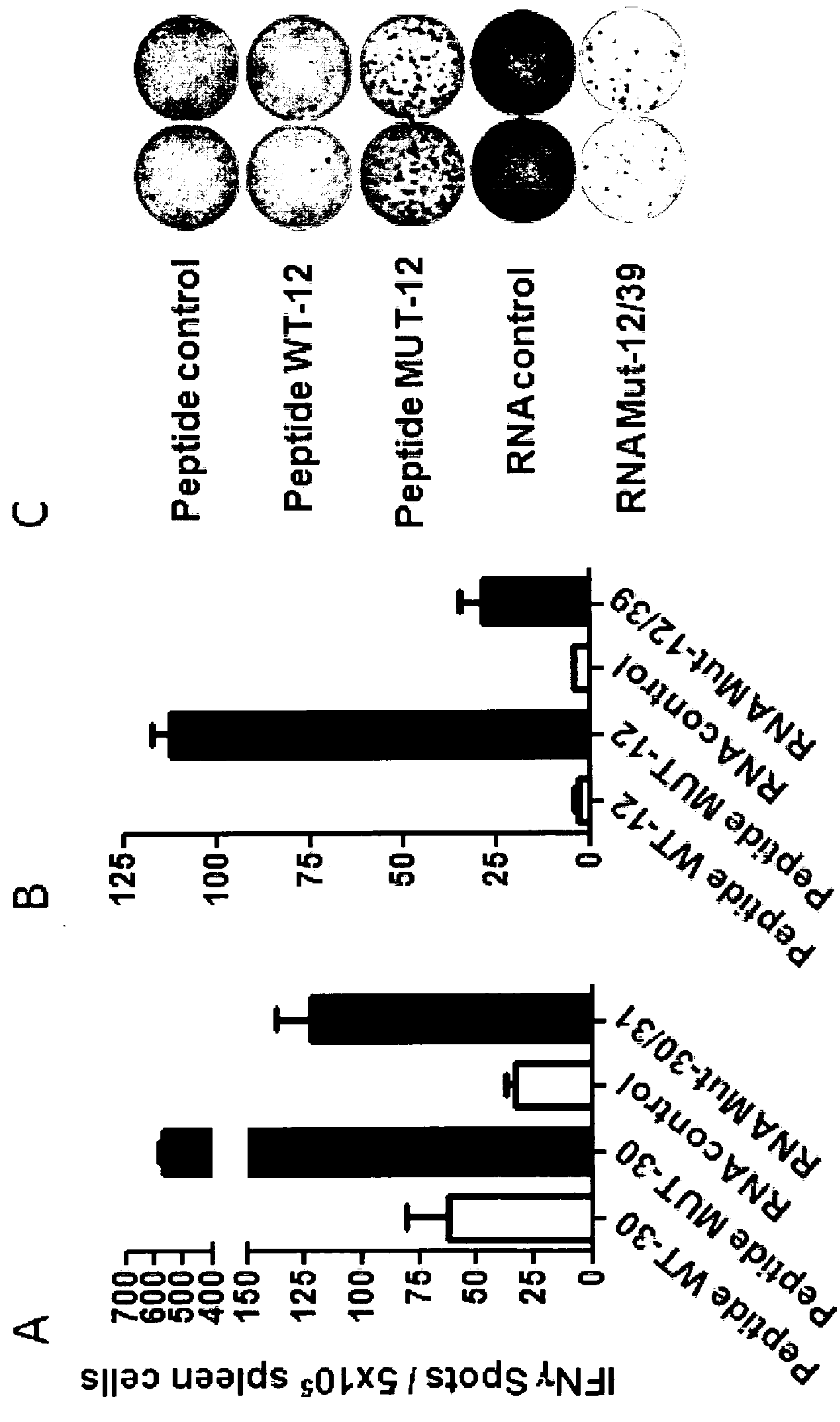

A
5´
Cap
5´UTR
L
Seq. 1
L
Seq. 2
L
....
L
Seq. 29
L
Seq. 30
L
3´UTR
poly-A
3´
RNA backbone
Vaccine Sequence
RNA backbone
B
Mut30
Linker 1
5´ ....NTTFNVSEESPSKPSFQEFVDWENVSPELNSTDQPFLPSAPVFIFTKGRKGGSGGGGSGG
Mut31
Linker 2
YFVLYKPPPKDNIPALVEEYLERGNFVANDLDWLLALPHDKFWCQVIFDEGGSGGGSGGS.... 3´
Mut39
Linker 1
5´ ....FEQLSESAKEELINFKRKRVAAFQKNLIEMSELEIKHARNNVSLLQSCIDGGSGGGGSGG
Mut12
Linker 2
TGASPGLGAYTPPPEEAMPFEFNGPAQGDHSQPPLQVPDLAPGGPEALVPGGSGGGSGGS.... 3´

Fig. 7


A
IFNγ Spots / 5x10⁵ spleen cells
700
600
500
400
150
125
100
75
50
25
0
Peptide WT-30
Peptide MUT-30
RNA control
RNA Mut-30/31
B
125
100
75
50
25
0
Peptide WT-12
Peptide MUT-12
RNA control
RNA Mut-12/39
C
Peptide control
Peptide WT-12
Peptide MUT-12
RNA control
RNA Mut-12/39

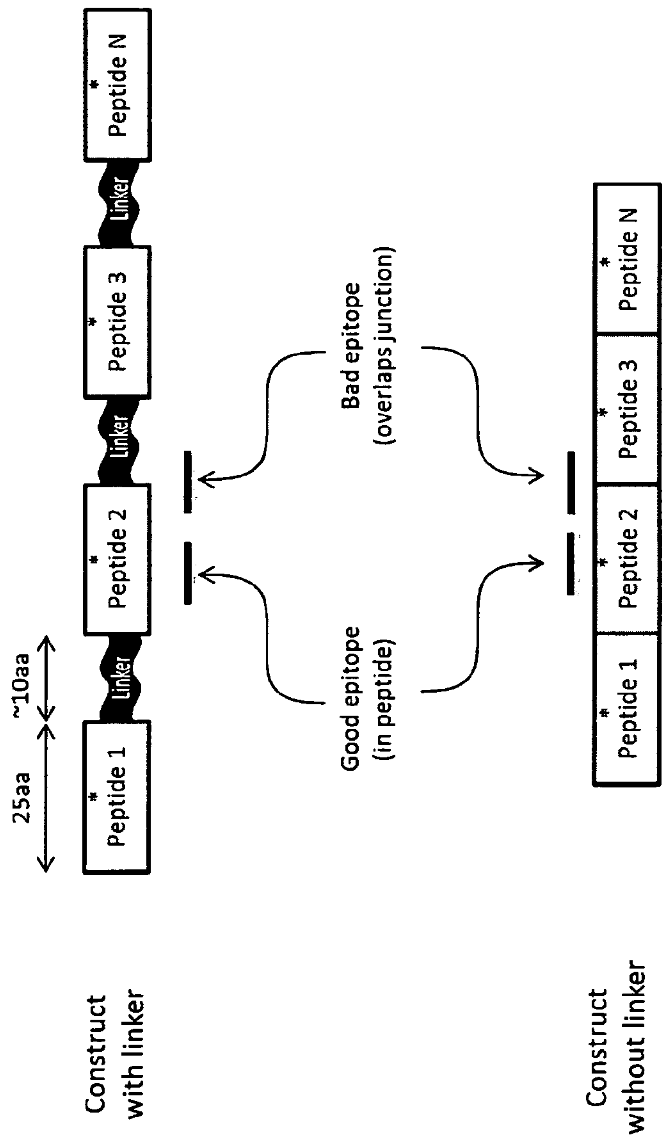

Construct with linker
25aa
~10aa
Peptide 1
Linker
Peptide 2
Linker
Peptide 3
Linker
Peptide N
Good epitope (in peptide)
Bad epitope (overlaps junction)
Construct without linker
Peptide 1
Peptide 2
Peptide 3
Peptide N

Fig. 8

Fig. 9

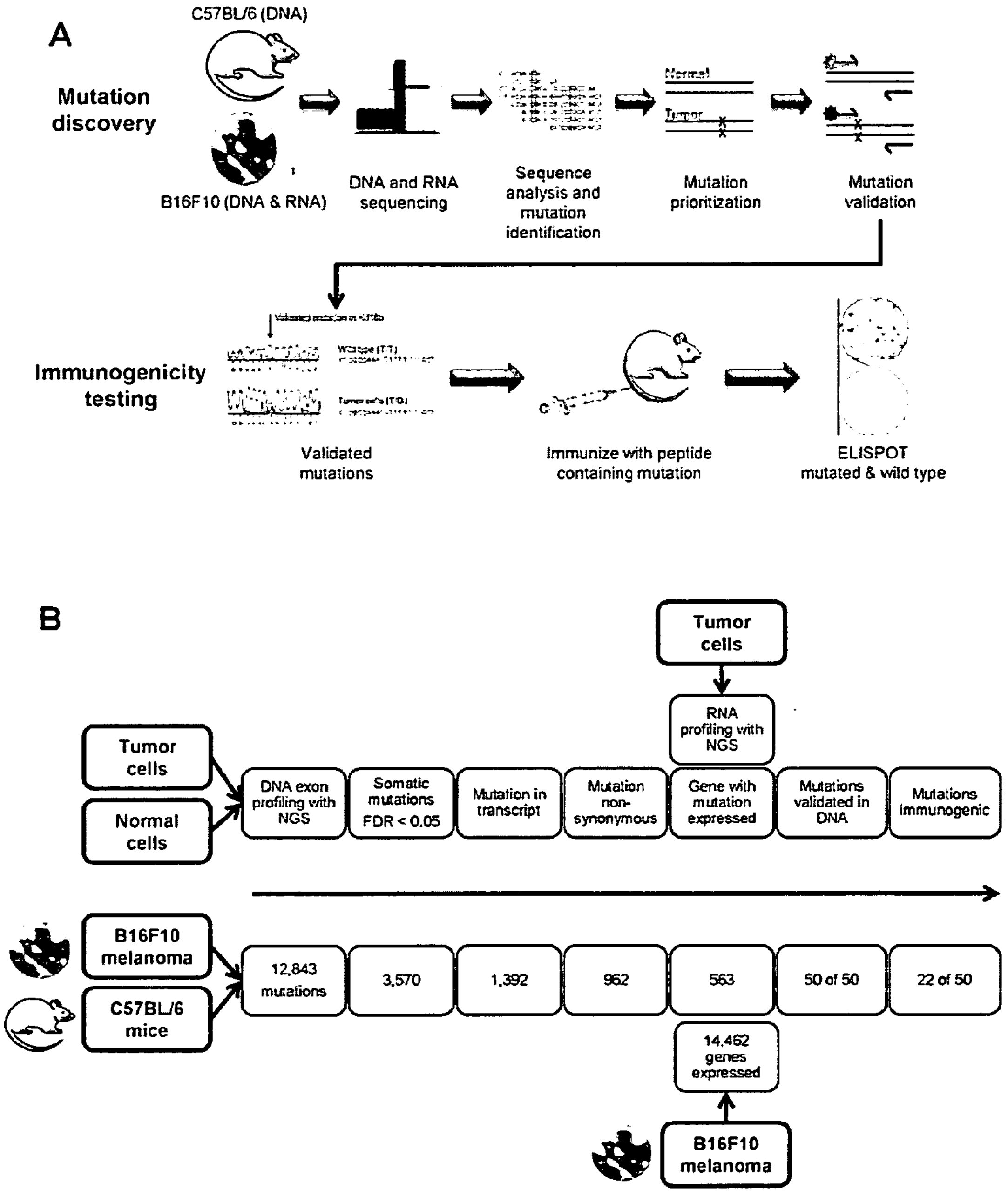

A
Mutation discovery
C57BL/6 (DNA)
B16F10 (DNA & RNA)
DNA and RNA sequencing
Sequence analysis and mutation identification
Normal
Tumor
Mutation prioritization
Mutation validation
Immunogenicity testing
Wt type (TT)
Tumor cells (TG)
Validated mutations
Immunize with peptide containing mutation
ELISPOT mutated & wild type
B
Tumor cells
RNA profiling with NGS
Tumor cells
Normal cells
DNA exon profiling with NGS
Somatic mutations FDR < 0.05
Mutation in transcript
Mutation non-synonymous
Gene with mutation expressed
Mutations validated in DNA
Mutations immunogenic
B16F10 melanoma
C57BL/6 mice
12,843 mutations
3,570
1,392
962
563
50 of 50
22 of 50
14,462 genes expressed
B16F10 melanoma

Fig. 9

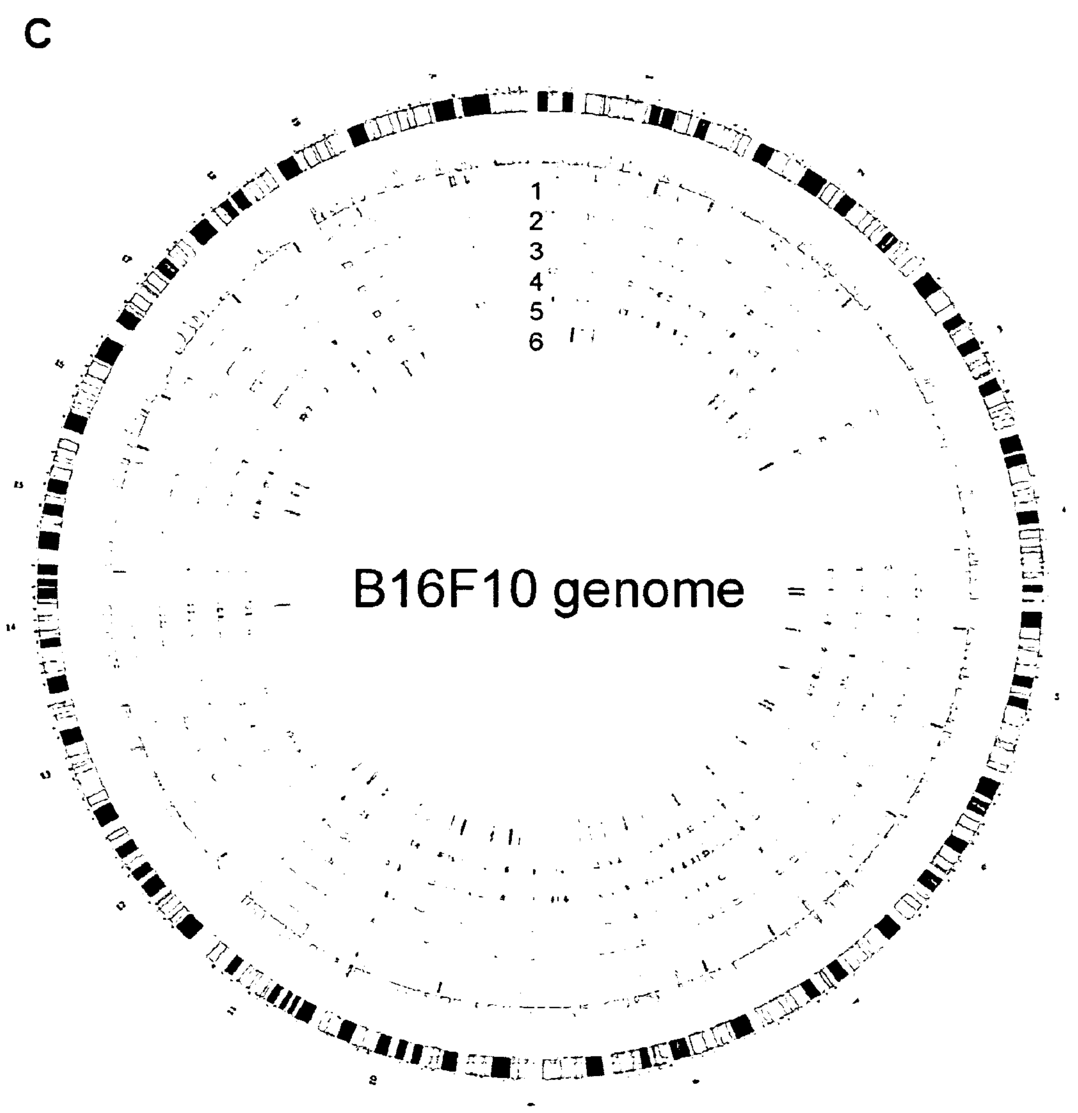


Fig. 10

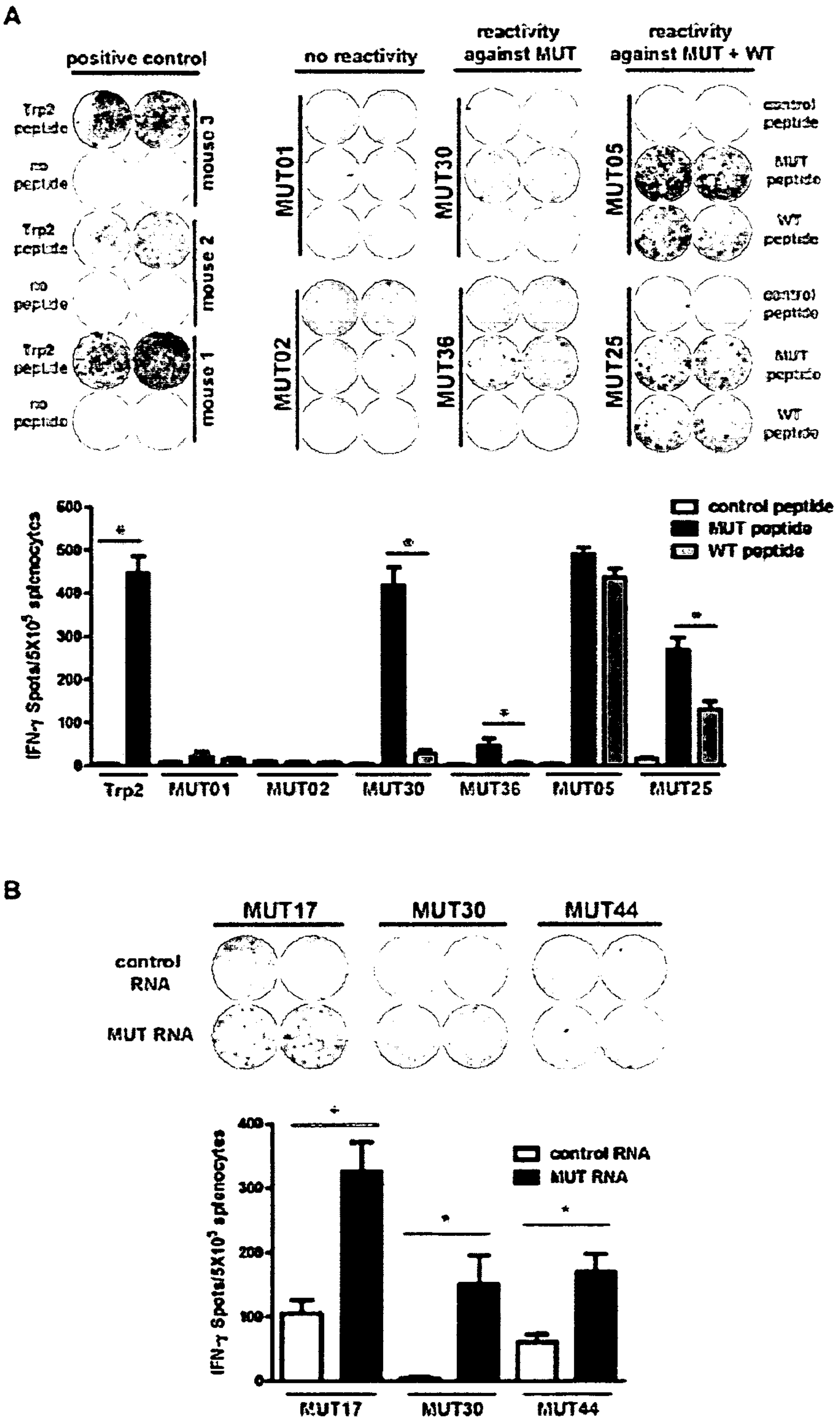
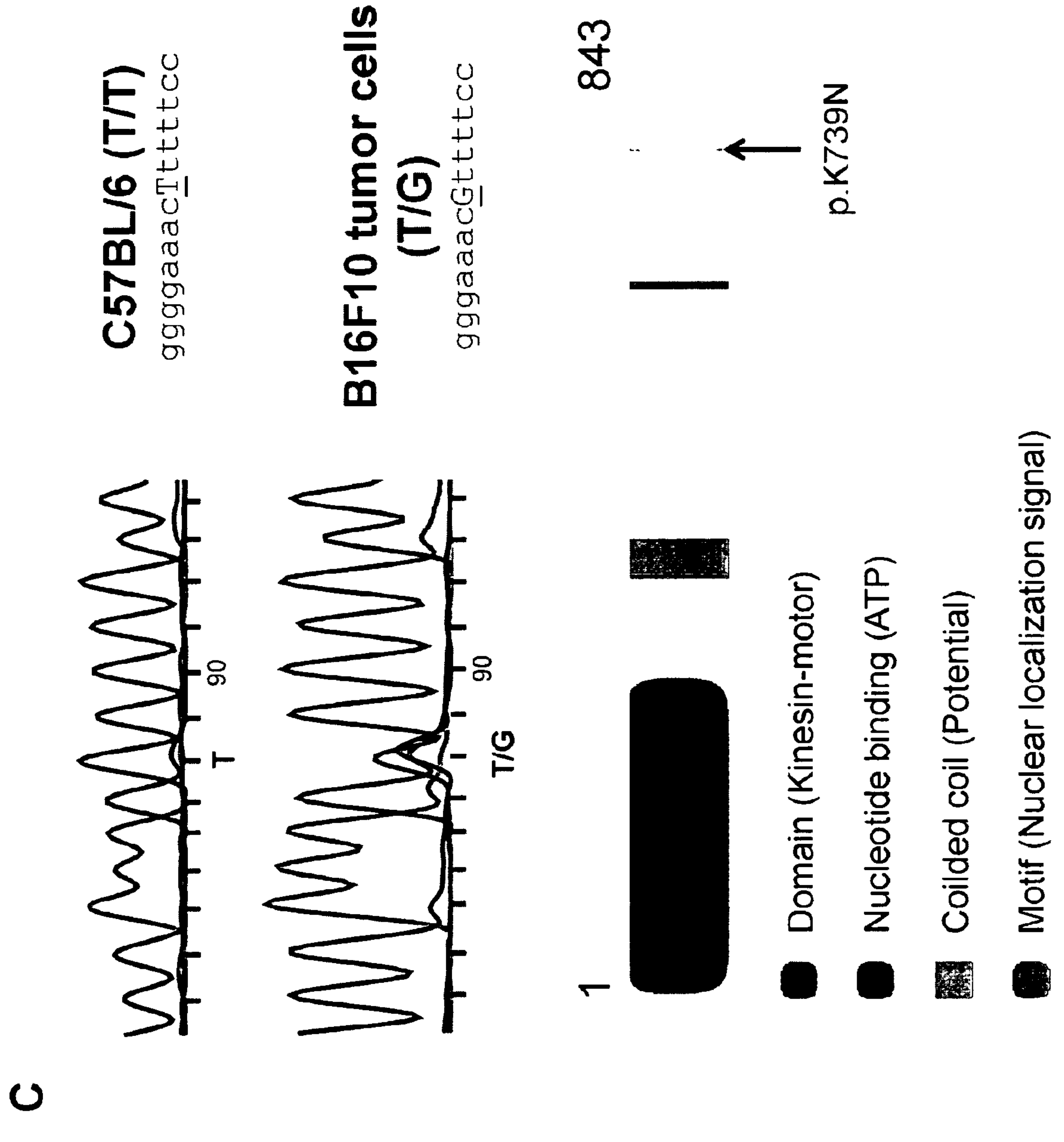

A
positive control
no reactivity
reactivity against MUT
reactivity against MUT + WT
Trp2 peptide
no peptide
mouse 3
mouse 2
mouse 1
MUT01
MUT02
MUT30
MUT36
MUT05
MUT25
control peptide
MUT peptide
WT peptide
IFN-γ Spots/5X10^5 splenocytes
0
100
200
300
400
500
600
*
control peptide
MUT peptide
WT peptide
Trp2
MUT01
MUT02
MUT30
MUT36
MUT05
MUT25
B
MUT17
MUT30
MUT44
control RNA
MUT RNA
IFN-γ Spots/5X10^5 splenocytes
0
100
200
300
400
*
control RNA
MUT RNA
MUT17
MUT30
MUT44

Fig. 10


C
C57BL/6 (T/T)
ggggaaacTttttcc
T
90
B16F10 tumor cells (T/G)
gggaaacGttttcc
T/G
90
1
843
p.K739N
● Domain (Kinesin-motor)
● Nucleotide binding (ATP)
▨ Coilded coil (Potential)
● Motif (Nuclear localization signal)

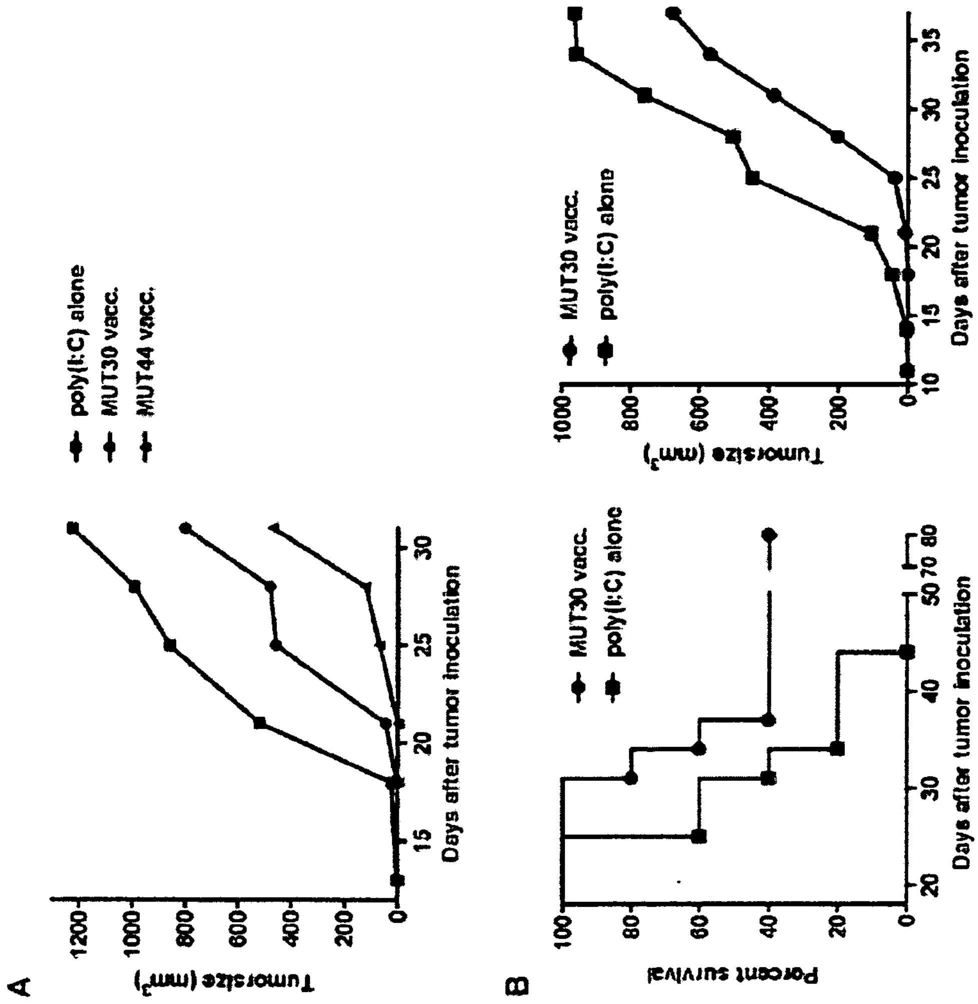

A
Tumorsize (mm³)
1200
1000
800
600
400
200
0
15
20
25
30
Days after tumor inoculation
poly(I:C) alone
MUT30 vacc.
MUT44 vacc.
B
Percent survival
100
80
60
40
20
0
20
30
40
50
70
80
Days after tumor inoculation
MUT30 vacc.
poly(I:C) alone
Tumorsize (mm³)
1000
800
600
400
200
0
10
15
20
25
30
35
Days after tumor inoculation
MUT30 vacc.
poly(I:C) alone

Fig. 11

Fig. 12

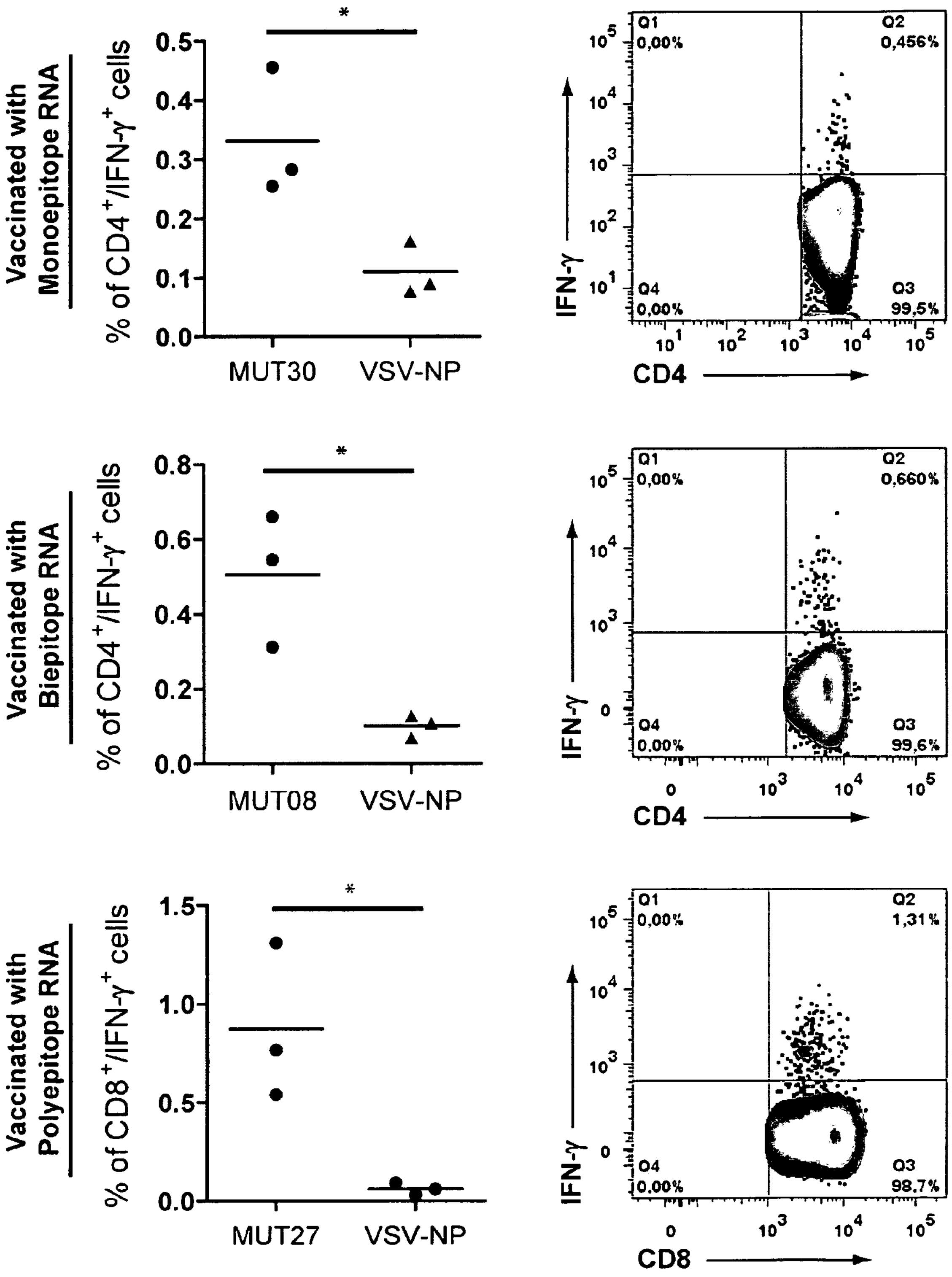

Vaccinated with Monoepitope RNA
% of CD4⁺/IFN-γ⁺ cells
0.0
0.1
0.2
0.3
0.4
0.5
*
MUT30
VSV-NP
IFN-γ
CD4
Q1 0,00%
Q2 0,456%
Q3 99,5%
Q4 0,00%
Vaccinated with Biepitope RNA
% of CD4⁺/IFN-γ⁺ cells
0.0
0.2
0.4
0.6
0.8
*
MUT08
VSV-NP
IFN-γ
CD4
Q1 0,00%
Q2 0,660%
Q3 99,6%
Q4 0,00%
Vaccinated with Polyepitope RNA
% of CD8⁺/IFN-γ⁺ cells
0.0
0.5
1.0
1.5
*
MUT27
VSV-NP
IFN-γ
CD8
Q1 0,00%
Q2 1,31%
Q3 98,7%
Q4 0,00%

Fig. 13

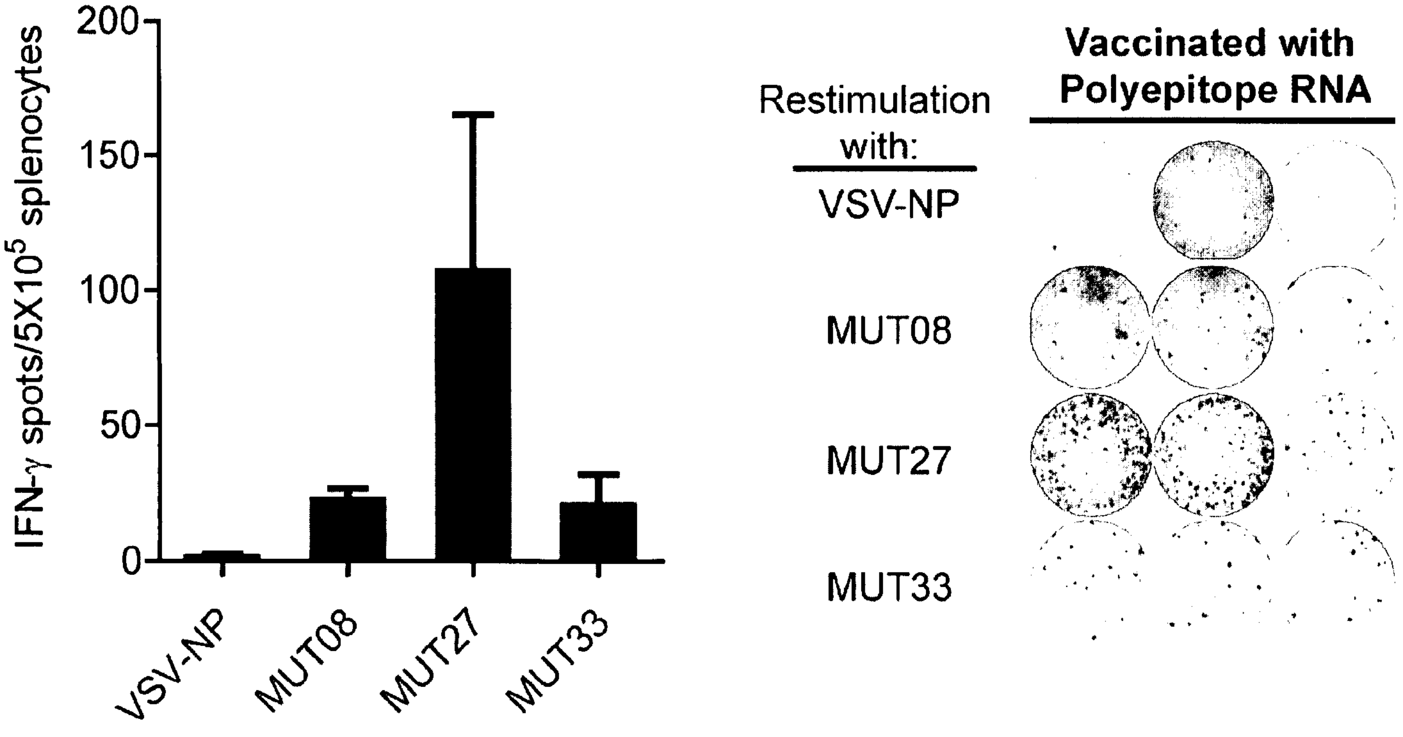

IFN-γ spots/5X10⁵ splenocytes
200
150
100
50
0
VSV-NP
MUT08
MUT27
MUT33
Restimulation with:
Vaccinated with Polyepitope RNA
VSV-NP
MUT08
MUT27
MUT33

Fig. 14

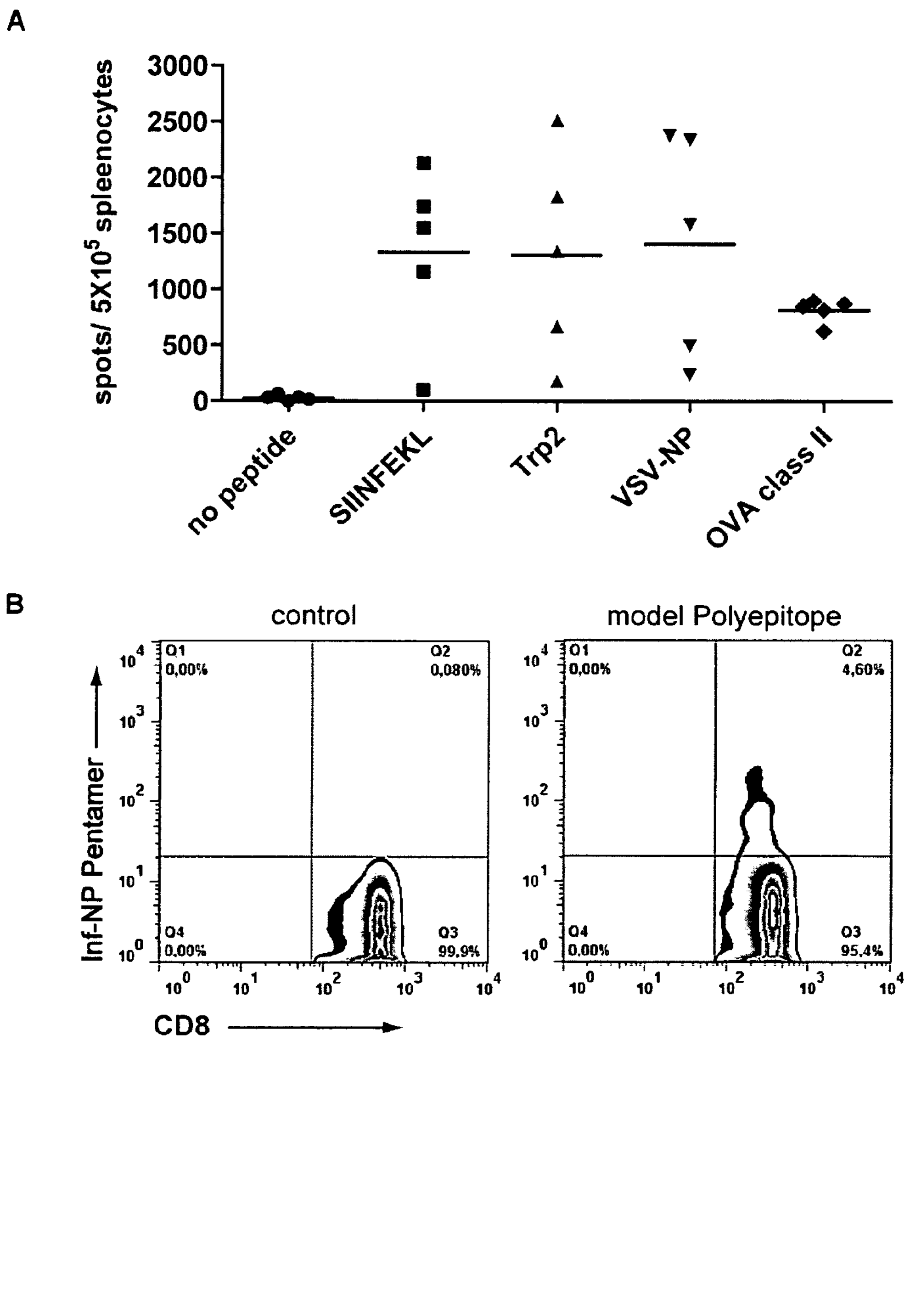

A
spots/ 5X10^5 spleenocytes
3000
2500
2000
1500
1000
500
0
no peptide
SIINFEKL
Trp2
VSV-NP
OVA class II
B
control
model Polyepitope
Inf-NP Pentamer
CD8
10^4
10^3
10^2
10^1
10^0
Q1
0,00%
Q2
0,080%
Q4
0,00%
Q3
99,9%
Q1
0,00%
Q2
4,60%
Q4
0,00%
Q3
95,4%

Fig. 15 A

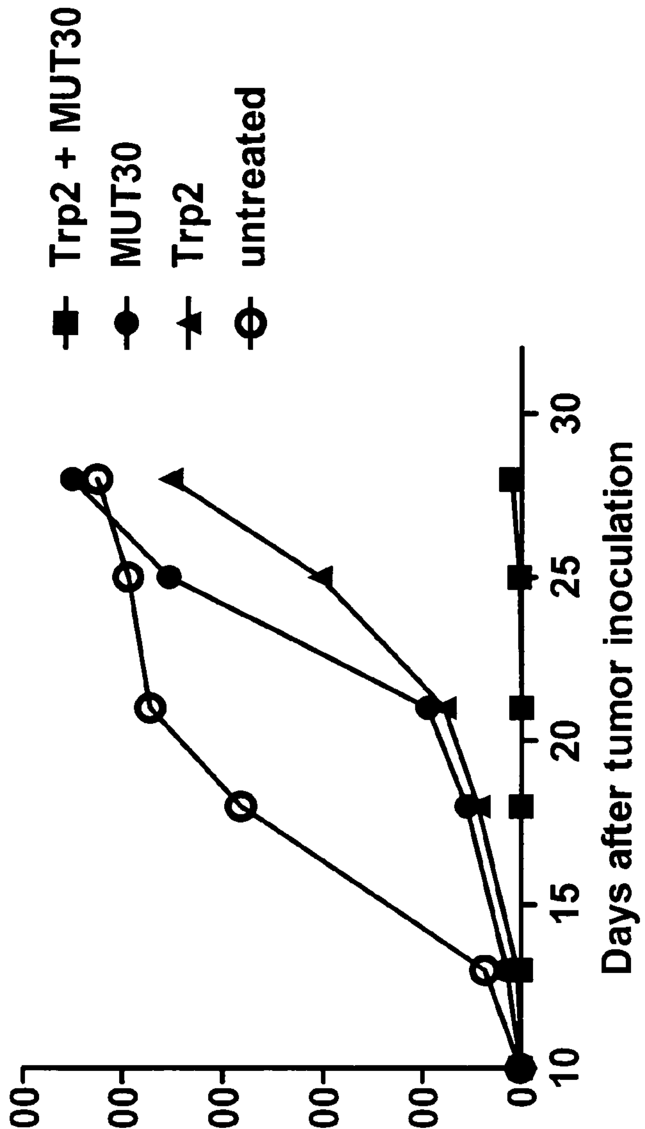

Tumorsize (mm³)
1000
800
600
400
200
0
10
15
20
25
30
Days after tumor inoculation
Trp2 + MUT30
MUT30
Trp2
untreated

Fig. 15 B

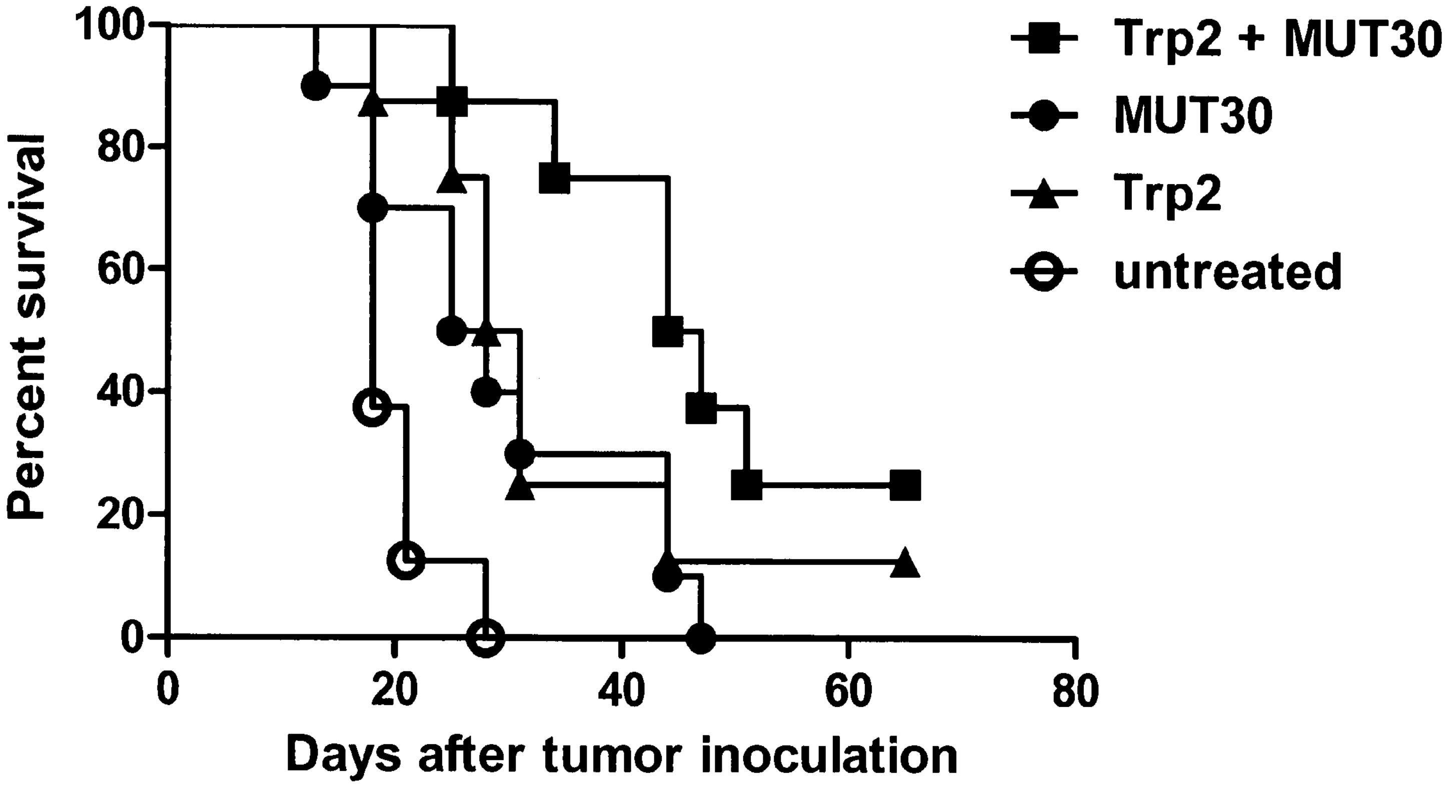
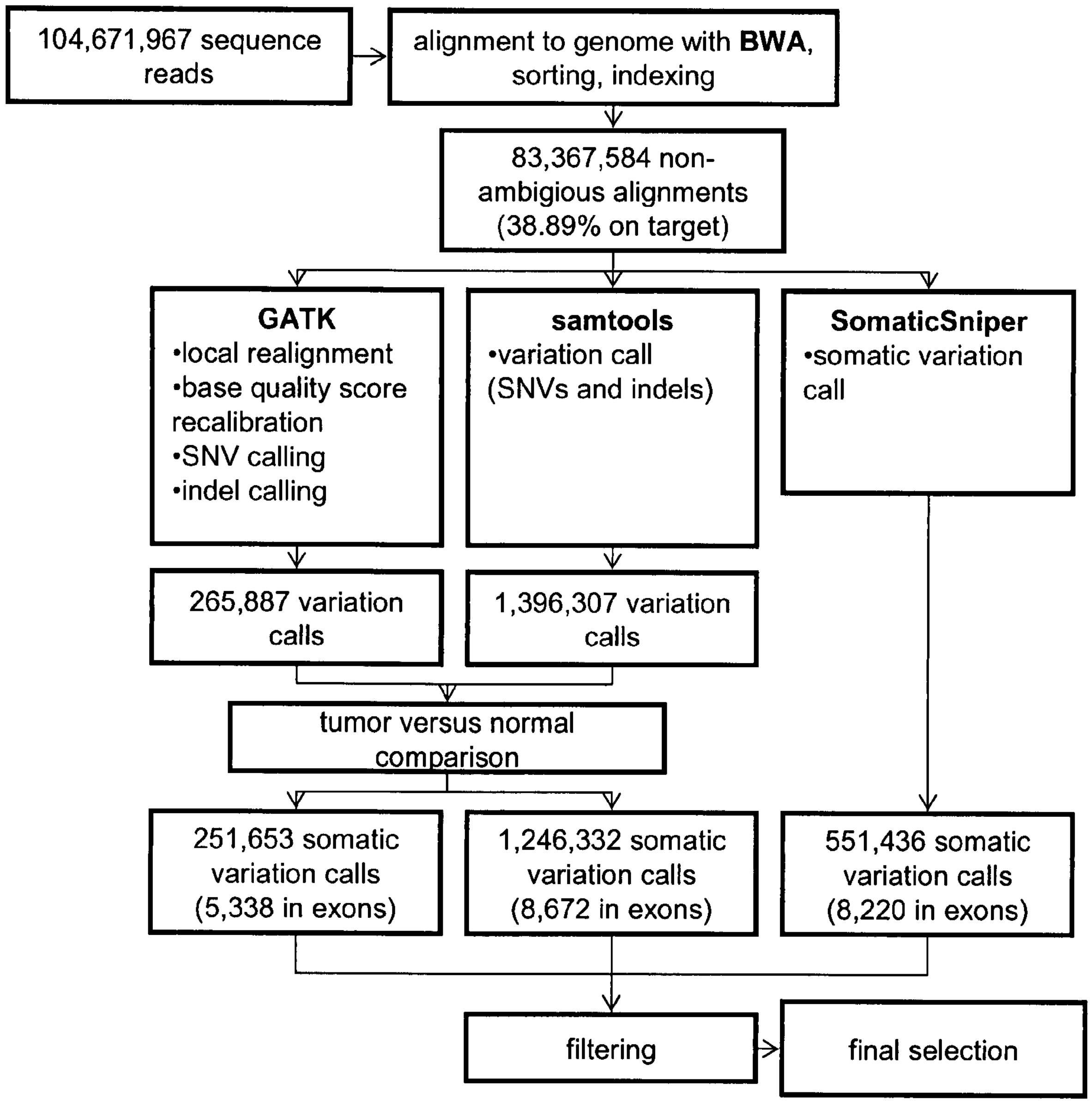

Percent survival
100
80
60
40
20
0
0
20
40
60
80
Days after tumor inoculation
■ Trp2 + MUT30
● MUT30
▲ Trp2
○ untreated

Fig. 16


104,671,967 sequence reads
alignment to genome with **BWA**, sorting, indexing
83,367,584 non-ambigious alignments (38.89% on target)
GATK
•local realignment
•base quality score recalibration
•SNV calling
•indel calling
samtools
•variation call (SNVs and indels)
SomaticSniper
•somatic variation call
265,887 variation calls
1,396,307 variation calls
tumor versus normal comparison
251,653 somatic variation calls (5,338 in exons)
1,246,332 somatic variation calls (8,672 in exons)
551,436 somatic variation calls (8,220 in exons)
filtering
final selection

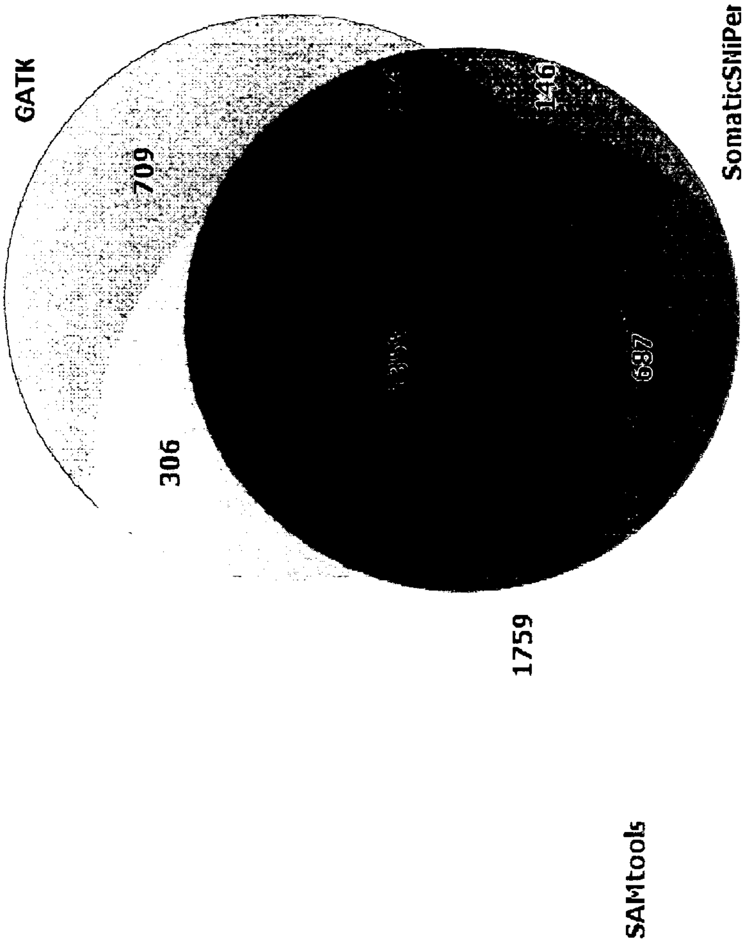


Fig. 17

Fig. 18

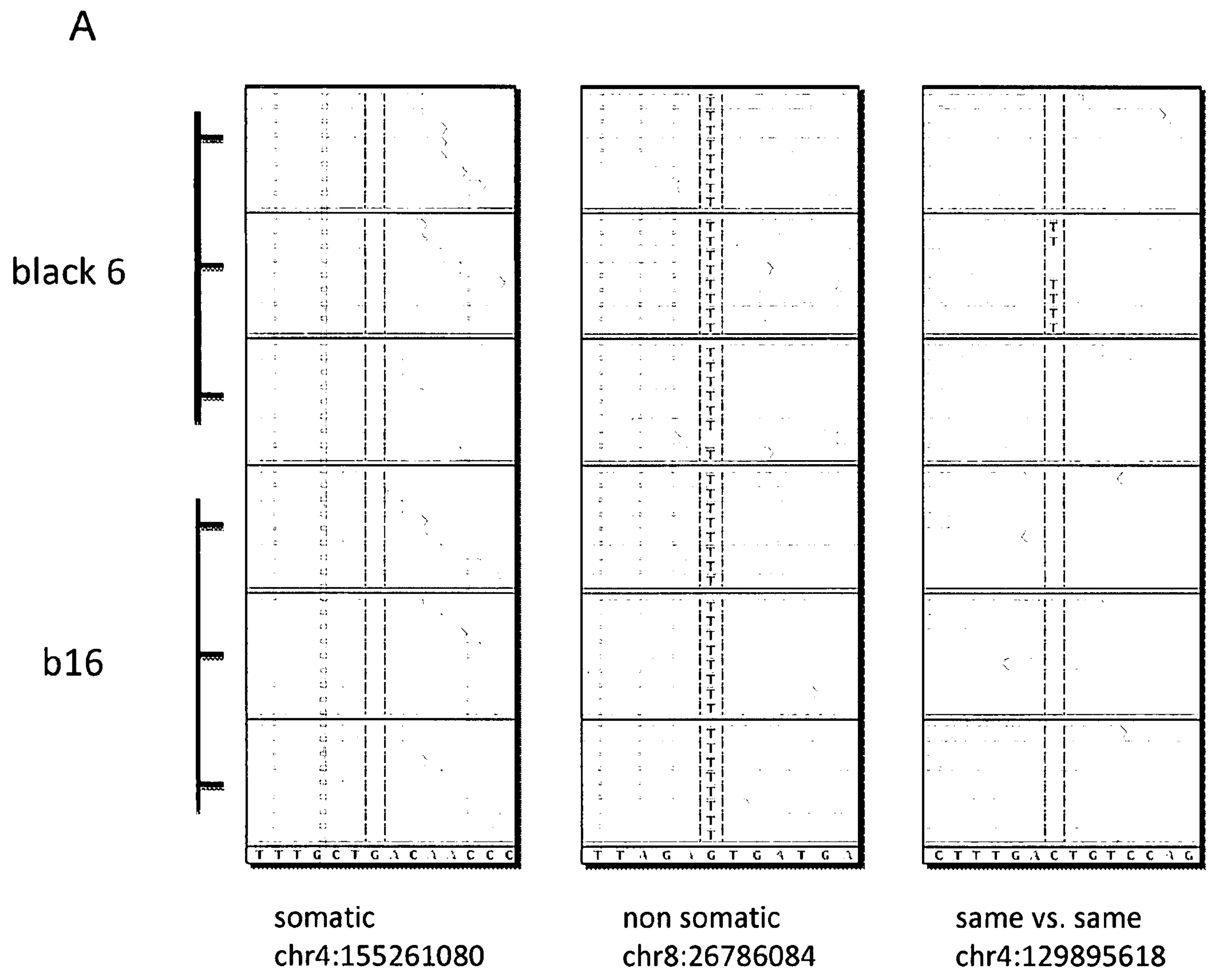

A
black 6
b16
T T T G C T G A C A A C C C
T T A G A G T G A T G A
C T T T G A C T G T C C A G
somatic
chr4:155261080
non somatic
chr8:26786084
same vs. same
chr4:129895618

Fig. 18

B

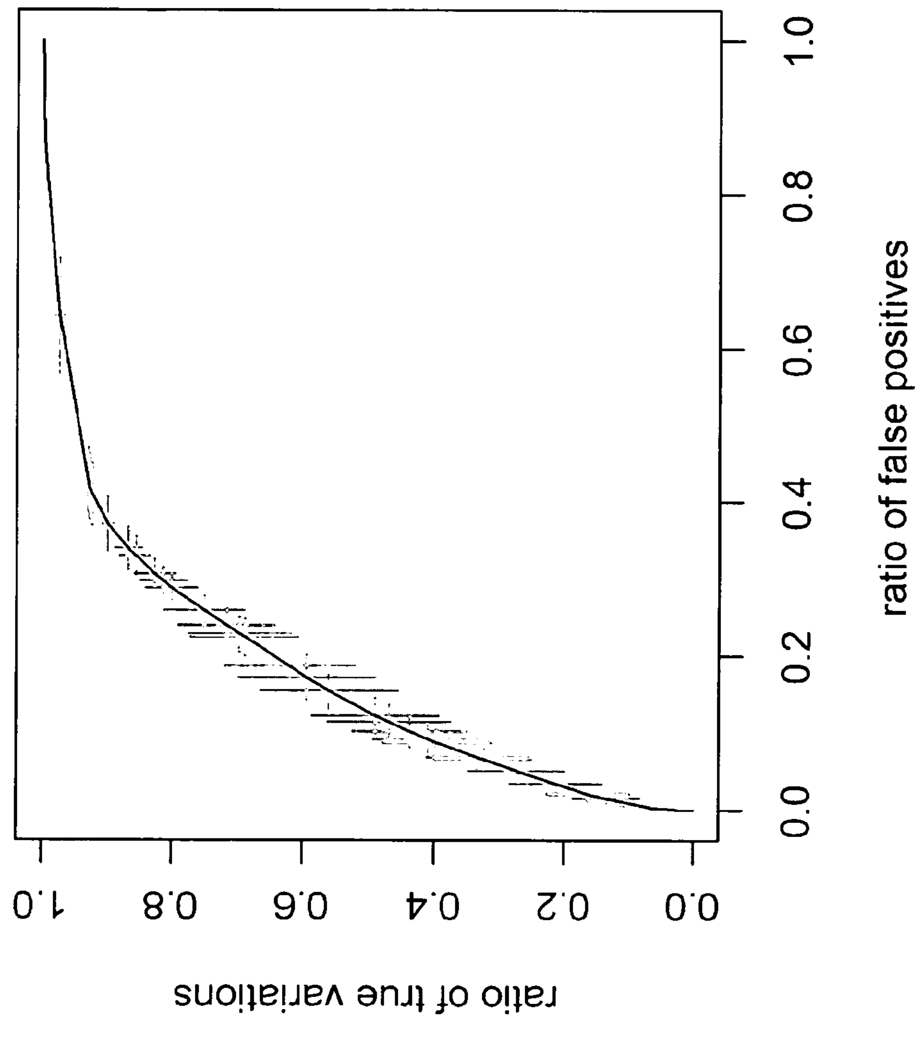

ratio of true variations
0.0
0.2
0.4
0.6
0.8
1.0
ratio of false positives
0.0
0.2
0.4
0.6
0.8
1.0

Fig. 19

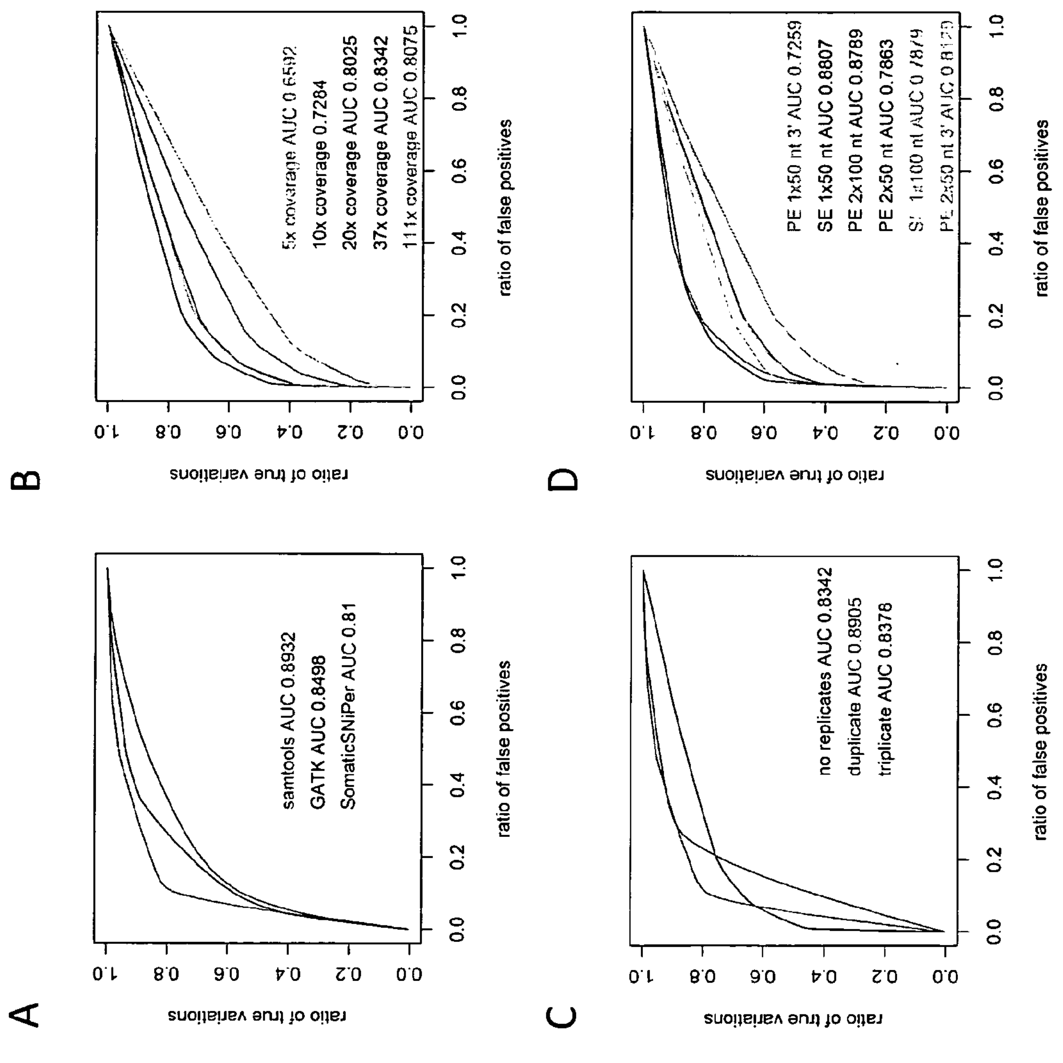

A
ratio of true variations
0.0
0.2
0.4
0.6
0.8
1.0
ratio of false positives
samtools AUC 0.8932
GATK AUC 0.8498
SomaticSNIPer AUC 0.81
B
ratio of true variations
ratio of false positives
5x coverage AUC 0.6592
10x coverage 0.7284
20x coverage AUC 0.8025
37x coverage AUC 0.8342
111x coverage AUC 0.8075
C
ratio of true variations
ratio of false positives
no replicates AUC 0.8342
duplicate AUC 0.8905
triplicate AUC 0.8376
D
ratio of true variations
ratio of false positives
PE 1x50 nt 3' AUC 0.7259
SE 1x50 nt AUC 0.8807
PE 2x100 nt AUC 0.8789
PE 2x50 nt AUC 0.7863
SE 1x100 nt AUC 0.7879
PE 2x50 nt 3' AUC 0.8179

Fig. 20

A

Chromosome	Position	Reference allele	Sample allele(s)	relative rank [%]	FDR [%]
8	110078987	G	R	3,2	0,6
1	59540714	G	S	7,6	0,7
5	124854313	G	K	7,8	0,7
10	59352802	C	M	8,3	0,7
16	36919828	A	M	9,8	0,7
2	144078227	C	Y	10,1	0,7
8	12834637	G	S	10,9	0,7
19	6121411	T	Y	11,4	0,7
1	58533360	A	M	12,1	0,7
15	98478052	A	R	13,6	0,7

B

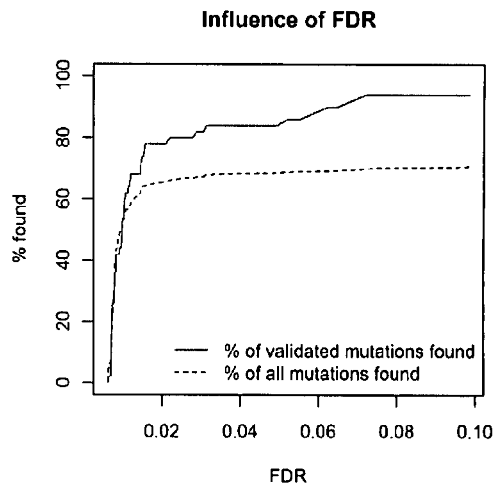
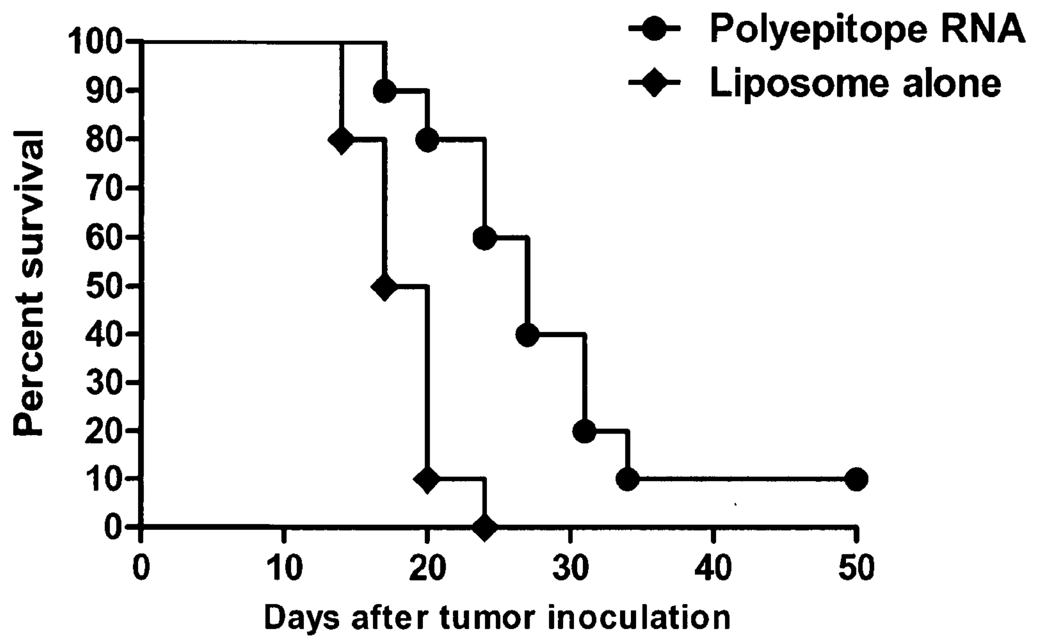
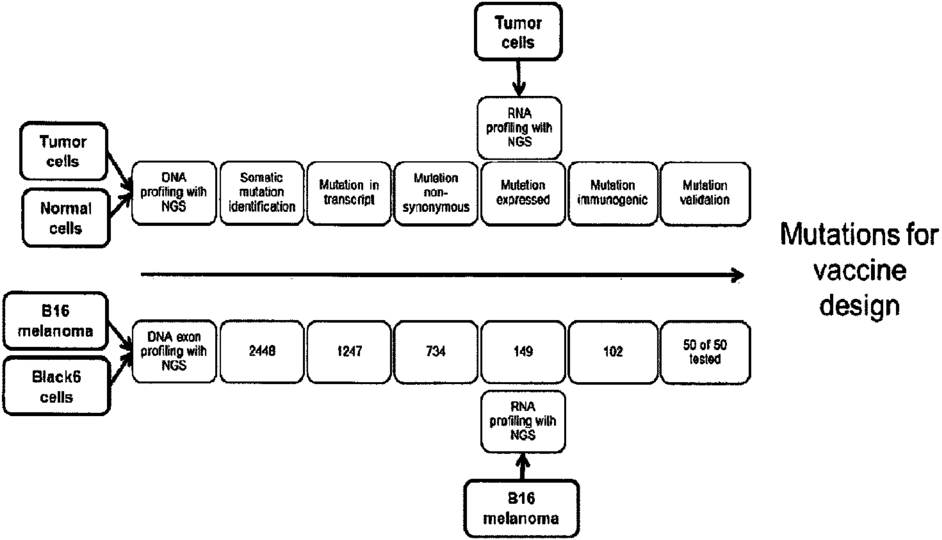


Fig. 21


Polyepitope RNA
Liposome alone
Percent survival
100
90
80
70
60
50
40
30
20
10
0
0
10
20
30
40
50
Days after tumor inoculation


Tumor cells
RNA profiling with NGS
Tumor cells
Normal cells
DNA profiling with NGS
Somatic mutation identification
Mutation in transcript
Mutation non-synonymous
Mutation expressed
Mutation immunogenic
Mutation validation
Mutations for vaccine design
B16 melanoma
Black6 cells
DNA exon profiling with NGS
2448
1247
734
149
102
50 of 50 tested
RNA profiling with NGS
B16 melanoma