nucleotide variant mutations but also neoantigens generated by insertions/deletion mutations that generate a frameshift peptide are important, the latter is expected to be particular immunogenic. Recently two different personalized vaccination approaches based either on RNA or on peptides have been evaluated in phase-I clinical studies. The data obtained shows that vaccination indeed can both expand pre-existing neoantigen-specific T cells and induce a broader repertoire of new T-cell specificity in cancer patients (Sahin et al., 2017). The main limitation of both approaches is the maximum number of neoantigens that are targeted by the vaccination. The upper limit for the peptide-based approach, based on their published data, is of twenty peptides and was not reached in all patients because in some cases peptides could not be synthesized. The described upper limit for the RNA-based approach is even lower, since they include only 10 mutations in each vaccine (Sahin et al., 2017).

The challenge for a cancer vaccine in curing cancer is to induce a diverse population of immune T cells capable of recognizing and eliminating as large a number of cancer cells as possible at once, to decrease the chance that cancer cells can "escape" the T cell response and are not being recognized by the immune response. Therefore, it is desirable that the vaccine encodes a large number of cancer specific antigens, i.e. neoantigens. This is particular relevant for a personalized genetic vaccine approach based on cancer specific neoantigens of an individual. In order to optimize the probability of success as many neoantigens as possible should be targeted by the vaccine. Moreover, experimental data support the notion that effective immunogenic neoantigens in patients cover a broad range of predicted affinities for the patient's MHC alleles (e.g. Gros et al., 2016). Most of the current prioritization methods instead apply an affinity threshold, for example the frequently used 500 nM limit, that may limit the selection of immunogenic neoantigens. There is therefore a need for a priorization method that avoids the limitations of current methods (e.g. exclusion due to low predicted affinity) and for a vaccination approach that allows for a personalized vaccine targeting a large and therefore broader and more complete set of neoantigens.

## Summary

In one example, the present disclosure provides a method for selecting cancer neoantigens for use in a personalized vaccine comprising the steps of:

(a) determining neoantigens in a sample of cancerous cells obtained from an individual, wherein each neoantigen

- is comprised within a coding sequence,

- comprises at least one mutation in the coding sequence resulting in a change of the encoded amino acid sequence that is not present in a sample of non-cancerous cells of said individual, and
- consists of 9 to 40, preferably 19 to 31, more preferably 23 to 25, most preferably 25 contiguous amino acids of the coding sequence in the sample of cancerous cells,

(b) determine for each neoantigen the mutation allele frequency of each of said mutations of step (a) within the coding sequence,

(c) determining the expression level of each coding sequence comprising at least one of said mutations,

(i) in said sample of cancerous cells, or

(ii) from an expression database of the same cancer type as the sample of cancerous cells,

(d) predicting the MHC class I binding affinity of the neoantigens, wherein

(I) the HLA class I alleles are determined from the sample of non-cancerous cells of said individual,

(II) for each HLA class I allele determined in (I) the MHC class I binding affinity of each fragment consisting of 8 to 15, preferably 9 to 10, more preferably 9, contiguous amino acids of the neoantigen is predicted, wherein each fragment is comprising at least one amino acid change caused by the mutation of step (a), and

(III) the fragment with the highest MHC class I binding affinity determines the MHC class I binding affinity of the neoantigen,

(e) ranking the neoantigens according to the values determined in steps (b) to (d) for each neoantigen from highest to lowest values, yielding a first, a second and a third list of ranks,

(f) calculating a rank sum from said first, second and third list of ranks and ordering the neoantigens by increasing rank sum, yielding a ranked list of neoantigens,

(g) selecting 30-240, preferably 40-80, more preferably 60, neoantigens from the ranked list of neoantigens obtained in (f) starting with the lowest rank.

In another example, the present disclosure provides a method for constructing a personalized vector encoding a combination of neoantigens as described herein for use as a vaccine, comprising the steps of:

(i) ordering the list of neoantigens in at least 10^5 -10^8, preferably 10^6 different combinations,

(ii) generating all possible pairs of neoantigen junction segments for each combination, wherein each junction segment comprises 15 adjoining contiguous amino acids on either side of the junction,

(iii) predicting the MHC class I and/or class II binding affinity for all epitopes in junction segments wherein only HLA alleles are tested that are present in the individual the vector is designed for, and

(iv) selecting the combination of neoantigens with the lowest number of junctional epitopes with an IC50 of ≤1500nM and wherein if multiple combinations have the same lowest number of junctional epitopes the combination first encountered is selected.

In another example, the present disclosure provides a collection of vectors encoding each a different set of neoantigens as described herein or the combination of neoantigens as described herein, wherein the collection comprises 2 to 4, preferably 2, vectors and preferably wherein the vector inserts encoding the portion of the list are of about equal size in number of amino acids.

In yet another example, the present disclosure provides a vector as described herein or a collection of vectors as described herein for use in cancer vaccination.

## List of Figures

In the following, the content of the figures comprised in this specification is described. In this context please also refer to the detailed description of the invention above and/or below.

**Figure 1:** Generation of neoantigens derived from a SNV: (A) generation of 25mer neoantigens with the mutation centered and flanked by 12 wt aa upstream and downstream, (B) generation of 25mer neoantigens including more than one mutation and (C) generation of a neoantigen shorter than a 25mer when the mutation is close to the end or start of the protein sequence**.**

## Claims

1. A method for selecting cancer neoantigens for use in a personalized vaccine comprising the steps of:

(a) determining neoantigens in a sample of cancerous cells obtained from an individual, wherein each neoantigen
- is comprised within a coding sequence,
- comprises at least one mutation in the coding sequence resulting in a change of the encoded amino acid sequence that is not present in a sample of non-cancerous cells of said individual, and
- consists of 9 to 40, preferably 19 to 31, more preferably 23 to 25, most preferably 25 contiguous amino acids of the coding sequence in the sample of cancerous cells,

(b) determine for each neoantigen the mutation allele frequency of each of said mutations of step (a) within the coding sequence,

(c) determining the expression level of each coding sequence comprising at least one of said mutations,
(i) in said sample of cancerous cells, or
(ii) from an expression database of the same cancer type as the sample of cancerous cells,

(d) predicting the MHC class I binding affinity of the neoantigens, wherein
(I) the HLA class I alleles are determined from the sample of non-cancerous cells of said individual,
(II) for each HLA class I allele determined in (I) the MHC class I binding affinity of each fragment consisting of 8 to 15, preferably 9 to 10, more preferably 9, contiguous amino acids of the neoantigen is predicted, wherein each fragment is comprising at least one amino acid change caused by the mutation of step (a), and
(III) the fragment with the highest MHC class I binding affinity determines the MHC class I binding affinity of the neoantigen,

(e) ranking the neoantigens according to the values determined in steps (b) to (d) for each neoantigen from highest to lowest values, yielding a first, a second and a third list of ranks,

(f) calculating a rank sum from said first, second and third list of ranks and ordering the neoantigens by increasing rank sum, yielding a ranked list of neoantigens,

wherein the rank sum is a weighted rank sum, wherein

- the number of neoantigens determined in step (a) is added to the rank value of each neoantigen:
    - in the third list of ranks for which the prediction of MHC class I binding affinity of step (d) resulted in an IC50 value higher than 1000 nM and/or
    - in the fourth list of ranks for which the prediction of MHC class II binding affinity of step (d') resulted in an IC50 value higher than 1000 nM;

and/or

- in case of step (c)(i) being performed by massively parallel transcriptome sequencing, the rank sum of step (f) is multiplied by a weighing factor (WF), wherein WF is
    - 1, if the number of mapped transcriptome reads for the mutation is >0,
    - 2, if the number of mapped transcriptome reads for the mutation is 0 and the number of mapped reads for the non-mutated sequence is 0 and the transcripts-per-million (TPM) value is at least 0.5,
    - 3, if the number of mapped transcriptome reads for the mutation is 0 and the number of mapped reads for the non-mutated sequence is >0 and the transcripts-per-million (TPM) value is at least 0.5,
    - 4, if the number of mapped transcriptome reads for the mutation is 0 and the number of mapped reads for the non-mutated sequence is 0 and the transcripts-per-million (TPM) value is < 0.5, or
    - 5, if the number of mapped transcriptome reads for the mutation is 0 and the number of mapped reads for the non-mutated sequence is >0 and the transcripts-per-million (TPM) value is < 0.5;

(g) selecting 30-240, preferably 40-80, more preferably 60, neoantigens from the ranked list of neoantigens obtained in (f) starting with the lowest rank.

2. The method according to claim 1, wherein steps (a) and (d)(I) are performed using massively parallel DNA sequencing of the samples and wherein the number of reads comprising the mutation at the chromosomal position of the identified mutation is:

- in the sample of cancerous cells at least 2, preferably at least 3,

- in the sample of non-cancerous cells is 2 or less, preferably 0.

3. The method according to claim 1, wherein the method comprises a step (d') in addition to or alternatively to step (d), wherein step (d') comprises:
    - determining the HLA class II alleles in the sample of non-cancerous cells of said individual,
    - predicting the MHC class II binding affinity of the neoantigen, wherein
        - for each HLA class II allele determined the MHC class II binding affinity for each fragment of 11 to 30, preferably 15, contiguous amino acids of the neoantigen is predicted, wherein each fragment is comprising at least one mutated amino acid generated by the mutation of step (a), and
        - the fragment with the highest MHC class II binding affinity determines the MHC class II binding affinity of the neoantigen;

    wherein the MHC class II binding affinity is ranked from highest to lowest MHC class II binding affinity, yielding a fourth list of ranks that is included in the rank sum of step (f).

4. The method of any one of claims 1 to 3, wherein the at least one mutation of step (a) is a single nucleotide variant (SNV) or an insertion/deletion mutation resulting in a frame-shift peptide (FSP).

5. The method according to claim 4, wherein the mutation is a SNV and the neoantigen has the total size defined in step (a) and consists of the amino acid caused by the mutation, flanked on each side by a number of adjoining contiguous amino acids, wherein the number on each side does not differ by more than one unless the coding sequence does not comprise a sufficient number of amino acids on either side, wherein the neoantigen has the total size defined in step (a).

6. The method according to claim 4, wherein the mutation results in a FSP and each single amino acid change caused by the mutation results in a neoantigen that has the total size defined in step (a) and consists of:
    (i) said single amino acid change caused by the mutation and 7 to 14, preferably 8, N-terminally adjoining contiguous amino acids, and

(ii) a number of contiguous amino acids adjoining the fragment of step (i) on either side, wherein the number of amino acids on either side differ by not more than one, unless the coding sequence does not comprise a sufficient number of amino acids on either side,

wherein the MHC class I binding affinity of step (d) and/or the MHC class II binding affinity of step (d') is predicted for the fragment of step (i).

7. The method according to any one of the preceding claims, wherein the mutation allele frequency of the neoantigen determined in step (b) in the sample of cancerous cells is at least 2%, preferably 5%, more preferably at least 10%.

8. The method according to any one of the preceding claims, wherein step (g) further comprises removing neoantigens from genes linked to autoimmune disease, and/or neoantigens with a Shannon entropy value for their amino acid sequence lower than 0.1 from said ranked list of neoantigens.

9. The method according to any one of the preceding claims, wherein the expression level of said coding genes in step (c)(i) is determined by massively parallel transcriptome sequencing and wherein the expression level determined in step (c) (i) uses a corrected Transcripts Per Kilobase Million (corrTPM) value calculated according to the following formula

$$corrTPM = TPM * \left(\frac{M + c}{M + W + c}\right)$$

wherein M is the number of reads spanning the location of the mutation of step (a) that comprise the mutation and W is the number of reads spanning the location of the mutation of step (a) without the mutation and TPM is the Transcripts Per Kilobase Million value of the gene comprising the mutation and the c is a constant larger than 0, preferably 0.1.

10. The method according to any one of the preceding claims, wherein step (g) comprises an alternative selection process, wherein the neoantigens are selected from the ranked list of neoantigens starting with the lowest rank until a set maximum size in total overall length in amino acids for all selected neoantigens is reached, wherein the maximum size

is between 1200 and 1800, preferably 1500 amino acids for each vector of a monovalent or multivalent vaccine; and optionally wherein two or more neoantigens are merged into one new neoantigen if they comprise overlapping amino acid sequence segments.

11. The method according to any one of the preceding claims, further comprising constructing a personalized vector encoding a combination of neoantigens according to any one of the preceding claims for use as a vaccine, comprising the steps of:

(i) ordering the list of neoantigens in at least 10^5-10^8, preferably 10^6 different combinations,

(ii) generating all possible pairs of neoantigen junction segments for each combination, wherein each junction segment comprises 15 adjoining contiguous amino acids on either side of the junction,

(iii) predicting the MHC class I and/or class II binding affinity for all epitopes in junction segments wherein only HLA alleles are tested that are present in the individual the vector is designed for, and

(iv) selecting the combination of neoantigens with the lowest number of junctional epitopes with an IC50 of ≤1500nM and wherein if multiple combinations have the same lowest number of junctional epitopes the combination first encountered is selected.