PLOS ONE

RESEARCH ARTICLE

HLA class I restricted epitopes prediction of common tumor antigens in white and East Asian populations: Implication on antigen selection for cancer vaccine design

Wei Hu[1,2], Meifang He[3]*, Liangping Li[2]*

1 Clinical Trials Unit, The First Affiliated Hospital of Sun Yat-sen University, Sun Yat-sen University, Guangzhou, China, **2** Institute of Clinical Oncology, The First Affiliated Hospital of Jinan University, Jinan University, Guangzhou, China, **3** Laboratory of General Surgery, The First Affiliated Hospital of Sun Yat-sen University, Sun Yat-sen University, Guangzhou, China

* liangping_li@jnu.edu.cn, liangping_li@yahoo.com(LL); hmf410@163.com(MH)

Abstract

Tumor antigens processed and presented by human leukocyte antigen (HLA) Class I alleles are important targets in tumor immunotherapies. Clinical trials showed that presence of CD8 + T cells specific to tumor associated antigens (TAAs) and tumor neoantigens is one of the main factors resulting in tumor regression. Affinity prediction of tumor antigen epitopes to HLA is an important reference index for peptide selection, which is highly individualized. In this study, we selected 6 CTAs (cancer-testis antigens) commonly used in cancer immuno-therapy and top 95 hot mutations from the Cancer Genome Atlas for analyzing potential epi-topes with high affinities to the common HLA class I molecules in white and East Asian population, respectively. The results showed that the overall difference in CTAs epitope pre-diction is small between the two populations. Meanwhile, there is a linear relationship between the CTAs peptide length and the relative overall epitope occurrence. However, the difference is bigger for epitopes prediction of missense mutations between the two popula-tions. It is worth noting that, both in the two populations, the single point mutations with the highest incidences have the lowest epitope occurrences while the mutations with the highest epitope occurrences are with low mutation incidences. This may be the result of long-term selection by the host immunosurveillance. Frameshift/inframe indel mutation neoantigens are between CTAs and spot mutation neoantigens in the relationship between peptide length and predicted epitope number. Our results help provide clues for tumor antigen and epitope selection in cancer vaccine design.

Introduction

T cells play a leading role in immunity to cancer in the majority of successful cancer immuno-therapies, including immune checkpoint inhibitor, adoptive T cell transfer and tumor vaccine [1–4]. CD8+ T cell is one of main effectors resulting in tumor regression. Immune monitoring

indicated that CD8+ T cell response targeting tumor antigens is an important index for immunotherapeutic effect.

T cell-mediated clinical trials need to predict the epitopes of tumor antigens with high HLA binding affinity *in silico*, for selecting peptide for cancer vaccination therapy or immune-monitoring of tumor specific T cells during the trial. The candidates in form of synthesized peptides or peptides expressing mRNAs were used to stimulate specific CD8+ T cells *in vitro* for T cell adoptive therapy or *in vivo* for tumor vaccine therapy [3, 5, 6]. Tumor antigens mainly classify into tumor associated antigens (TAAs) and tumor specific antigens (TSAs). TSAs comprise mutated neoantigens [7]. For safety and immunogenicity considerations, one kind of commonly used TAAs are cancer-testis antigens (CTAs) which are expressed exclusively in germ cells of the testis and fetal ovary but are silent in normal somatic cells [8–10].

The specific activation of CD8+ T cells is based on the antigenic epitope recognition by T cell receptor, which is restricted by human leukocyte antigen (HLA) [11–13]. All individuals express six HLA class I alleles and a population in an area has its own HLA allele frequency distribution. The affinity of a peptide to HLA determines whether the peptide can be presented. In this study, we selected 6 CTAs commonly used in cancer immunotherapy and 95 top hot mutations from the Cancer Genome Atlas to predict potential epitopes which may be presented by the common HLA class I molecules in white and East Asian population, respectively.

Materials and methods

HLA class I allele frequency ranking analysis

We collected HLA class I allele data from the Allele Frequency Net Database (AFND). The data collection criteria were as follows: 1). Alleles start from HLA-A*01:01 to HLA-C*18:03. 2). Choose two Ethnic Origins: Caucasoid or Oriental for white and East Asian population, respectively. 3). Select allele Level of resolution $>= 4$. Combine data from different populations, arrange the data in allele frequency from large to small and calculate the rate of HLA allele positive population using *Hardy-Weinberg equilibrium* [14], with the equation $p = 1-(1-q)^2$, where p is the rate of positive population, q is the HLA allele frequency.

Selection of CTAs

We selected 6 CTAs including KK-LC-1, MAGE-A1, MAGE-A4, NY-ESO-1, PRAME, SSX2, which were commonly used in tumor immunotherapy with high security record in clinical research, for Class I epitope prediction [15, 16].

Selection of top somatic mutations from TCGA

We collected the top 100 somatic mutations by incidence with amino acid sequence change, including missense variant, frameshift variant, inframe deletion and inframe insertion, from all kinds of cancers in TCGA; arranged the data and removed three stop codon gain mutations (chr9:g.21971121G>A, chr1:g.200857892delA, chr6:g.30653271delT), B2M mutation (chr15:g.44711583delCT), PAX2 3' Prime UTR mutation (chr10:g.100827567delC), and finally selected the rest 95 mutations for the following analysis (S1 Table).

Tumor epitopes prediction

We chose whole protein sequences of CTAs and mutated peptide sequences to predict potential epitopes binding to HLA class I molecules. Mutated amino acids of spot mutations were flanked on both sides by 10 additional normal amino acids while the frameshift mutations
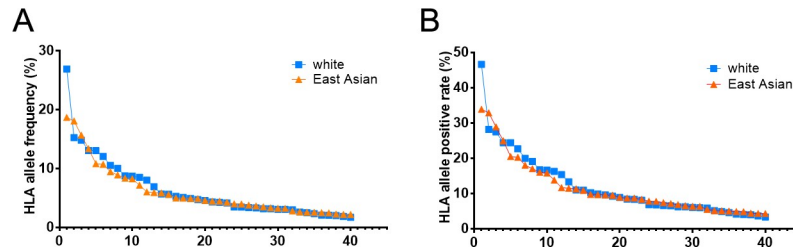
Fig 1. HLA allele frequency and positive rate in white and East Asian populations. The top 40 HLA alleles in white and East Asian populations were selected and displayed.

were flanked by 10 additional normal amino acids on upstream and the rest mutated sequence on downstream. NetMHCpan4.0 algorithm was selected for HLA class I molecule restricted epitope prediction. The epitope length was restricted to 8–11 mer. Set rank threshold for Strong Binding (SB) < = 0.5 and Weak Binding (WB) < = 2.0 according to the default values.

Results

To get HLA class I allele frequency of white and East Asian populations, total number of 1442312 white individuals from 97 population samples and 322437 East Asian individuals from 87 population samples were collected respectively. HLA allele information was rearranged according to relative criteria (Method). The distribution and HLA allele frequency are different in white and East Asian populations. In white population, the top commonest HLA-A*02:01 accounting for 26.9% is far more than the second commonest HLA-A*01:01 (15.2%), while the trend changes more slowly in East Asian population (Fig 1). We selected the HLA alleles with positive rate > 0.001% from each population for the following epitope analysis (S2 Table).

Due to the accuracy of the NetMHC prediction based on artificial neural networks, we selected only strong binding results (%Rank < = 0.5) for the next analysis. For each population we first calculated the SB epitope numbers of each HLA allele after epitope prediction of a specific antigen/peptide and weighted the data to HLA allele positive rate. Then summed over these data to get the average epitope number (AEN) of a specific antigen/peptide in the population (S3–S6 Tables).

In the part of epitope prediction of the 6 CTAs, both white and East Asian populations have the similar AENs after HLA allele positive rate weighting (Fig 2A). Moreover, the AENs of both populations have an excellent linear correlation with the peptide length of the antigens (Fig 2B).
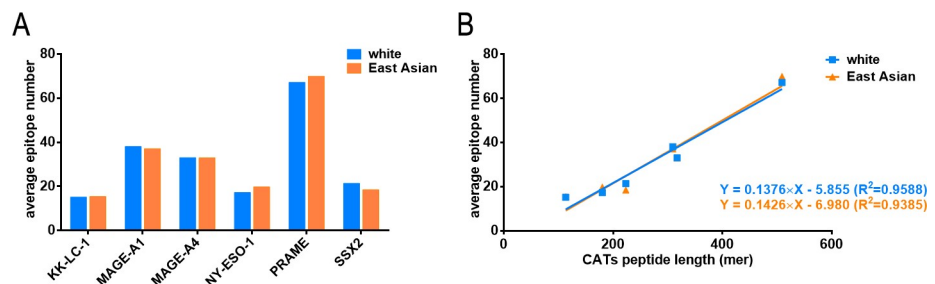

Fig 2. SB epitope prediction of CTAs in white and East Asian populations. A. Average epitope numbers of 6 CTAs. B. Linear analysis between average epitope numbers and antigen lengths. The linear equations and R square values are given in the figure.
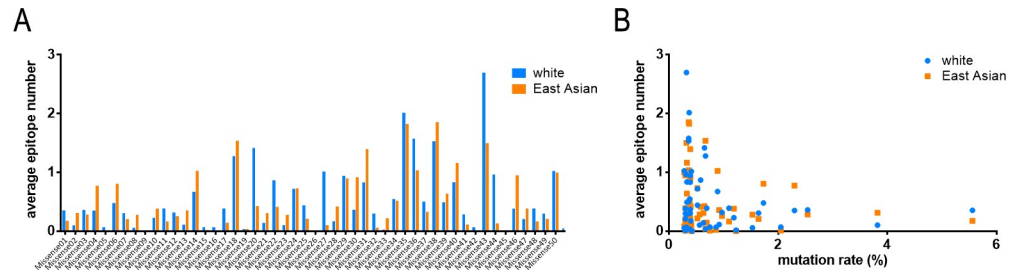
Fig 3. SB epitope prediction of single amino acid mutations in white and East Asian populations. A. Average epitope numbers of 50 missense mutations. B. Correlation analysis between average epitope numbers and mutation rates.

For the epitope prediction of tumor neoantigens, the AENs of missense mutations were quite different between the two populations (Fig 3A). For example, Missense20 (GTF2I_L424H), Missense27 (TP53_G245S), Missense44 (PPP2R1A_P179R). Diagram indicates that, in both populations, the three top missense mutations including BRAF_V600E, IDH1_R132H, PIK3CA_E545K (all over the rate of 2.5%) show low AENs, while the three highest AENs are all uncommon mutations including TP53_V157F (0.32%), ERBB2_S310F (0.37%), PIK3CA_H1047L (0.36%) (Fig 3B). The AENs of frameshift/inframe insertion mutations are similar between two populations (Fig 4A). In addition, AENs of different frameshift/inframe insertion mutations within one population are correlated with the lengths of mutated amino acids (Fig 4B), although the linear relation is not as obvious as in CTAs. Comparison of AENs fold difference of the two populations in three kinds of antigens (non-mutated and mutated) showed that there is a significant difference between CTAs and missense mutations but not the other two pairwise comparisons (Fig 5). These indicated that frameshift/inframe insertion mutations show similar characteristics of both CTAs and missense mutation.

Discussion

Specific immunotherapy of both adoptive T cell therapy and cancer vaccine involves in CD8+ CTL induction *in vitro* or *in vivo* by antigen stimulation, which is one of the key factors affecting immunotherapy effects. For the selection of optimal peptide pool (with high affinity), it needs to predict tumor antigen epitope accurately. NetMHC is one of the most commonly used epitope prediction algorithms developed by DTU Bioinformatics. The accuracy of the top predicted peptide is about 34% in version 4.0 of NetMHC [17]. There are only 81 HLA allele
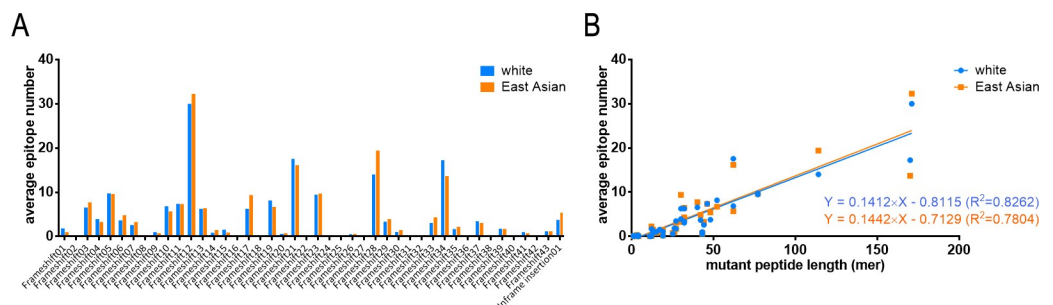

Fig 4. SB epitope prediction of frame related mutations in white and East Asian populations. A. Average epitope numbers of 43 frameshift mutations and 1 inframe insertion mutation. B. Linear analysis between average epitope numbers and mutant peptide lengths. The linear equations and R square values are given in the figure.
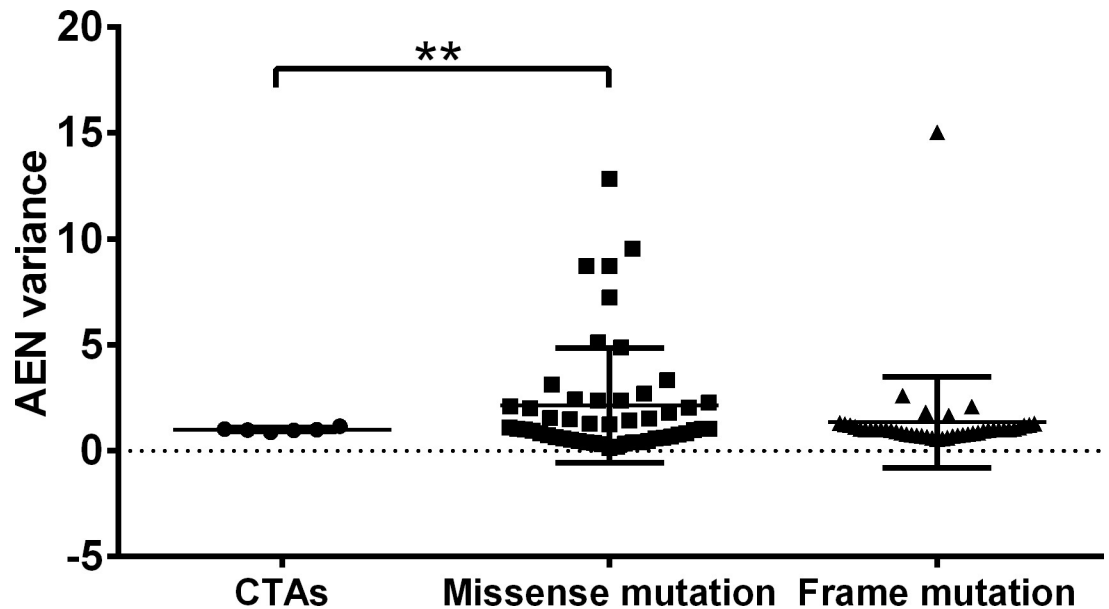
Fig 5. Comparison of average epitope number fold variances between white and East Asian populations in CTAs, neoantigens of missense mutations and frameshift/inframe insertion mutation. Mutations with AEN value 0 in both populations were set to have AEN variance values of 1. Frame mutations in the diagram includes frameshift and inframe insertion mutations. Unpaired t-test was used, $^{**}p<0.01$.

https://doi.org/10.1371/journal.pone.0229327.g005

options in this version, while NetMHCpan could cover most of the known HLA alleles. Therefore, in this study we chose NetMHCpan4.0 as the unified algorithm.

The HLA allele frequency varies widely among different races. This would lead to the same tumor antigen having different CD8+ T cellular immunogenicity in different populations. In this study, we predicted and compared the specific epitopes of CTAs and neoantigens in white and East Asian population, respectively. The results showed that there was a good linear relationship between AENs and the peptide lengths of the CTAs. Meanwhile, the AEN differences of CTAs were small between the two populations. Frameshift/inframe insertion mutation antigens showed the similar trends but the linearity was worse. Although with the same mutated length, the AENs of missense mutations were very different between each other in each population. Moreover, the AENs of many missense mutations vary greatly in two populations. It is worth noting that the single point mutations with the highest incidence have the lowest AEN while the mutations with the highest AEN scores are with low mutation incidence. This may be explained by the immune selective pressure and evolution. Hot mutations are often from cancer driver genes and need to be low immunogenic to avoid the host immunosurveillance. Therefore, the somatic mutation with high immunogenicity must be scarce at the population level. These mutations are often passenger mutations and are with high individualization. The neoantigen related immunotherapy researches so far indicate that the vast majority of antigen specific CD8+ T cells are targeting passenger mutations.

The effective tumor specific CD8+ T cell stimulation involves factors of HLA allele, tumor antigen epitope and TCR repertoire. These factors vary greatly among cancer patients. In terms of tumor antigen, mutation neoantigens have greater individual differences and less potential epitopes than TAA. Therefore, the valuable mutation neoantigens will be with high individualization, which reduces the versatility of neoantigen vaccine [7, 18]. In order to reduce costs and improve efficiency in clinical immunotherapy, one viable solution, we could design and produce common tumor antigen peptides or specific TCR-T cells in advance [19].

The result of this study shows that there are few neoantigens fulfilling the requirement to have both high immunogenicity and high incidence at the same time. However, CTAs could make up for the shortage of neoantigens [20]. CTAs have no mutation that reduced the personalization differences in population. Meanwhile, the long antigen sequence of CTAs ensures that they would always contain epitopes being presented by different HLA alleles. It is possible that the form of full-length CTAs-mRNA vaccine can be applied in different populations. This will greatly reduce the preparation time and cost in tumor immunotherapy [21].

Conclusion

This study predicted the epitopes of CTAs and neoantigens from the top mutations that were presented by the common HLA class I molecules in both white and East Asian populations. Epitopes analysis showed the immunogenicity of CTAs and neoantigens and their potential values in cancer vaccine design.

Supporting information

S1 Table. Mutation peptide sequences. 95 mutations selected for neoantigen epitope prediction.
(XLSX)

S2 Table. HLA allele analysis of white and East Asian. Population samples selected in HLA allele analysis, HLA allele frequencies and positive rates in white and East Asian populations.
(XLSX)

S3 Table. CTAs epitope prediction. SB epitope prediction of CTAs and average epitope number (AEN) analysis in white and East Asian populations.
(XLSX)

S4 Table. Missense mutations epitope prediction. SB epitope prediction of missense mutations and AEN analysis in white and East Asian populations.
(XLSX)

S5 Table. Frameshift mutations epitope prediction. SB epitope prediction of frameshift mutations and AEN analysis in white and East Asian populations.
(XLSX)

S6 Table. Inframe insertion mutations epitope prediction. SB epitope prediction of inframe insertion mutation and AEN analysis in white and East Asian populations.
(XLSX)

Acknowledgments

Author Contributions

Conceptualization: Wei Hu, Meifang He, Liangping Li.

Data curation: Wei Hu, Meifang He.

Formal analysis: Wei Hu.

Funding acquisition: Meifang He, Liangping Li.

Methodology: Wei Hu.

Resources: Meifang He.

Supervision: Liangping Li.

Writing – original draft: Wei Hu, Meifang He, Liangping Li.

Writing – review & editing: Wei Hu, Meifang He, Liangping Li.

References

1. Tumeh PC, Harview CL, Yearley JH, Shintaku IP, Taylor EJ, Robert L, et al. PD-1 blockade induces responses by inhibiting adaptive immune resistance. Nature. 2014; 515(7528):568–71. https://doi.org/10.1038/nature13954 PMID: 25428505; PubMed Central PMCID: PMC4246418.

2. Tran E, Robbins PF, Lu YC, Prickett TD, Gartner JJ, Jia L, et al. T-Cell Transfer Therapy Targeting Mutant KRAS in Cancer. N Engl J Med. 2016; 375(23):2255–62. https://doi.org/10.1056/NEJMoa1609279 PMID: 27959684; PubMed Central PMCID: PMC5178827.

3. Sahin U, Derhovanessian E, Miller M, Kloke BP, Simon P, Lower M, et al. Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. Nature. 2017; 547(7662):222–6. https://doi.org/10.1038/nature23003 PMID: 28678784.

4. Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. Science. 2015; 348(6230):124–8. https://doi.org/10.1126/science.aaa1348 PMID: 25765070; PubMed Central PMCID: PMC4993154.

5. Ott PA, Hu Z, Keskin DB, Shukla SA, Sun J, Bozym DJ, et al. An immunogenic personal neoantigen vaccine for patients with melanoma. Nature. 2017; 547(7662):217–21. https://doi.org/10.1038/nature22991 PMID: 28678778; PubMed Central PMCID: PMC5577644.

6. Stronen E, Toebes M, Kelderman S, van Buuren MM, Yang W, van Rooij N, et al. Targeting of cancer neoantigens with donor-derived T cell receptor repertoires. Science. 2016; 352(6291):1337–41. https://doi.org/10.1126/science.aaf2288 PMID: 27198675.

7. Yee C, Lizee GA. Personalized Therapy: Tumor Antigen Discovery for Adoptive Cellular Therapy. Cancer J. 2017; 23(2):144–8. https://doi.org/10.1097/PPO.0000000000000255 PMID: 28410303.

8. Thomas R, Al-Khadairi G, Roelands J, Hendrickx W, Dermime S, Bedognetti D, et al. NY-ESO-1 Based Immunotherapy of Cancer: Current Perspectives. Front Immunol. 2018; 9:947. https://doi.org/10.3389/fimmu.2018.00947 PMID: 29770138; PubMed Central PMCID: PMC5941317.

9. Schooten E, Di Maggio A, van Bergen En Henegouwen PMP, Kijanka MM. MAGE-A antigens as targets for cancer immunotherapy. Cancer Treat Rev. 2018; 67:54–62. https://doi.org/10.1016/j.ctrv.2018.04.009 PMID: 29763778.

10. Gordeeva O. Cancer-testis antigens: Unique cancer stem cell biomarkers and targets for cancer therapy. Semin Cancer Biol. 2018; 53:75–89. https://doi.org/10.1016/j.semcancer.2018.08.006 PMID: 30171980.

11. Verdegaal EM, de Miranda NF, Visser M, Harryvan T, van Buuren MM, Andersen RS, et al. Neoantigen landscape dynamics during human melanoma-T cell interactions. Nature. 2016; 536(7614):91–5. https://doi.org/10.1038/nature18945 PMID: 27350335.

12. Stevanovic S, Pasetto A, Helman SR, Gartner JJ, Prickett TD, Howie B, et al. Landscape of immunogenic tumor antigens in successful immunotherapy of virally induced epithelial cancer. Science. 2017; 356(6334):200–5. https://doi.org/10.1126/science.aak9510 PMID: 28408606.

13. Parkhurst M, Gros A, Pasetto A, Prickett T, Crystal JS, Robbins P, et al. Isolation of T-Cell Receptors Specifically Reactive with Mutated Tumor-Associated Antigens from Tumor-Infiltrating Lymphocytes Based on CD137 Expression. Clin Cancer Res. 2017; 23(10):2491–505. https://doi.org/10.1158/1078-0432.CCR-16-2680 PMID: 27827318.

14. Maiers M, Gragert L, Klitz W. High-resolution HLA alleles and haplotypes in the United States population. Hum Immunol. 2007; 68(9):779–88. https://doi.org/10.1016/j.humimm.2007.04.005 PMID: 17869653.

15. Zhang J, Wang L. The Emerging World of TCR-T Cell Trials Against Cancer: A Systematic Review. Technol Cancer Res Treat. 2019; 18:1533033819831068. https://doi.org/10.1177/1533033819831068 PMID: 30798772; PubMed Central PMCID: PMC6391541.

16. Gjerstorff MF, Andersen MH, Ditzel HJ. Oncogenic cancer/testis antigens: prime candidates for immunotherapy. Oncotarget. 2015; 6(18):15772–87. https://doi.org/10.18632/oncotarget.4694 PMID: 26158218; PubMed Central PMCID: PMC4599236.

17. Andreatta M, Nielsen M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. Bioinformatics. 2016; 32(4):511–7. https://doi.org/10.1093/bioinformatics/btv639 PMID: 26515819.

18. Srivastava PK. Neoepitopes of Cancers: Looking Back, Looking Ahead. Cancer Immunol Res. 2015; 3 (9):969–77. https://doi.org/10.1158/2326-6066.CIR-15-0134 PMID: 26342008.

19. Klebanoff CA, Wolchok JD. Shared cancer neoantigens: Making private matters public. J Exp Med. 2018; 215(1):5–7. https://doi.org/10.1084/jem.20172188 PMID: 29269561; PubMed Central PMCID: PMC5748868.

20. Hilf N, Kuttruff-Coqui S, Frenzel K, Bukur V, Stevanovic S, Gouttefangeas C, et al. Actively personalized vaccination trial for newly diagnosed glioblastoma. Nature. 2019; 565(7738):240–5. https://doi.org/10.1038/s41586-018-0810-y PMID: 30568303.

21. Forni G. Vaccines for tumor prevention: a pipe dream? J Infect Dev Ctries. 2015; 9(6):600–8. https://doi.org/10.3855/jidc.7201 PMID: 26142669.