HLA class I restricted epitopes prediction of common tumor antigens in Caucasoid and Oriental ethnic populations: implication on antigen selection for tumor specific CD8+ T cellular immunotherapy and cancer vaccine design

Wei Hu
Sun Yat-Sen University First Affiliated Hospital Department of Liver Surgery
He Meifang
Sun Yat-Sen University
Li Liangping (liangping_li@yahoo.com)

Research article

Keywords: Tumor antigen, Immunotherapy, CD8+ T cell, HLA allele frequency

Posted Date: June 13th, 2019

DOI: https://doi.org/10.21203/rs.2.10261/v1

Abstract

Background Tumor antigens processed and presented by human leukocyte antigen (HLA) Class I alleles are important targets in tumor immunotherapy. Clinical trials showed that CD8+ T cells specific to tumor associated antigens (TAAs) and tumor neoantigens is one of the main factors resulting tumor regression. Affinity prediction of tumor antigen epitopes to HLA is an important reference index for peptide selection which is highly individualized. Results In this study, we selected 6 CTAs (cancer-testis antigens) commonly used in cancer immunotherapy and top 95 hot mutations from the Cancer Genome Atlas for analyzing potential epitopes with high affinities to the common HLA class I molecules in Caucasoid and Oriental ethnic population respectively. The results showed that the overall difference of CTAs epitope prediction 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 populations. It's worth noting that, both in the two populations, the single point mutations with the highest incidences have the lowest epitope occurrence while the mutations with the highest epitope occurrence are with low mutation incidence. This may be the result of long-term selection by the host immunosurveillance. The characteristic of predicted epitopes of frameshift / inframe insertion mutations was approximately halfway between the other two antigens above. Conclusion Our results provide clues for tumor antigen and epitope selection in specific CD8+ T cellular immunotherapy and cancer therapeutic vaccine design.

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

Selection of top mutations from TCGA

We collected the top 100 mutations by incidence which cause 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 (Additional file 1: table S1).

Tumor epitopes prediction

We choose 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 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 Caucasoid and Oriental ethnic populations, total number of 1442312 Caucasian individuals from 97 population samples and 322437 Oriental individuals from 87 population samples were collected separately. HLA allele information was rearranged according to relative criteria (Method). The distribution and HLA allele frequency are different in Caucasoid and Oriental population. In Caucasoid 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 Oriental population (Fig.1). We selected the HLA alleles with positive rate > 0.001% from each population for the following epitope analysis (Additional file 2: table S2).

Due to the accuracy of the NetMHC prediction based on artificial neural networks, we select 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 (Additional file 3: table S3, Additional file 4: table S4, Additional file 5: table S5, Additional file 6: table S6).

In the part of epitope prediction of the 6 CTAs, both Caucasoid and Oriental population have the similar AEN after HLA allele positive rate weighting (Fig.2A). Moreover, the AEN of both populations have an excellent linear correlation with the peptide length of the antigen (Fig.2B).

For the epitope prediction of tumor neoantigen, the AEN 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 AEN, while the three highest AEN 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, AEN of different frameshift/inframe

insertion mutations within one population are correlated with the length of mutated amino acids (Fig. 4B), although the linear relation is not as obvious as in CTAs. Comparison of AEN 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 including CTAs-frameshift and inframe insertion mutations, frameshift/inframe insertion mutations and missense mutations (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 effect. 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 alleles option 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 same tumor antigen has different CD8+ T cellular immunogenicity in different population. In this study, we predicted and compared the specific epitopes of CTAs and neoantigens separately in Caucasoid and Oriental on the whole. The results showed that the overall difference of CTAs epitopes is small between the two populations. Frameshift/inframe insertion mutation antigens showed the similar trends: the longer the mutated sequence, the smaller difference of AENs between the two populations and the higher AEN it got. This results in a linear correlation between AEN and the peptide length of the antigen. However, as we had expected, AEN of missense mutations are different within and between the two populations. The prerequisite for immunogenicity is that the peptide could be presented by HLA allele. For the short peptides from CTAs, the epitope prediction is similar to missense mutations; however, CTAs with long enough protein sequences could always find out epitopes for different HLA alleles in both populations, which means that the CTAs could be presented in a wider population.

It’s 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 that hot mutations of cancer driver genes need to be low immunogenic to avoid the host immunosurveillance. Therefore, the mutation with high immunogenicity must be scarce at the population level. These mutations are often passenger mutations, which result from accumulated DNA damage of cancer cells 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 CD8 cell stimulation involves factors of HLA allele, epitope and TCR repertoire. The different people have different HLA alleles and TCR repertoire, the shorter of peptide sequence, more variable for epitope generation. With the individual variation of mutations, the valuable mutation epitopes will be with high individualization[7, 18].

In order to reduce costs and improve efficiency in clinical immunotherapy, one viable solution is to prepare the common tumor antigen vaccine or specific TCR-T cells in advance[19]. The result of this study shows that there is few neoantigen fulfilling the requirement to have both high immunogenicity and high incidence at the same time. However, TAAs could make up for the shortage of neoantigens[20]. TAAs have no mutation which reduced the personalization differences in population. Meanwhile, the long antigen sequence of TAAs ensure that there would always contain epitopes being presented by different HLA alleles. From this point of view as the results showed in this study, TAAs have advantages in providing abundant epitopes in TCR-T related immunotherapies both in Caucasoid and Oriental populations. It is possible that the form of full length TAAs-mRNA vaccine can be applied in different populations even without considering the individual HLA type. In additional, CTA-specific TCR-T cells restricted by the common HLA alleles can be manufactured in advance for T cellular adoptive therapy. 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 Caucasoid and Oriental populations. Epitopes analysis showed the immunogenicities of TAAs and neoantigens and their potential values in tumor immunotherapy.

Declarations

Additional files

Additional file 1: table S1. 95 mutations selected for neoantigen epitope prediction analysis

Additional file 2: table S2. Population samples selected in HLA allele analysis, HLA allele frequencies and positive rates in Caucasoid and Oriental populations

Additional file 3: table S3. SB epitope prediction of CTAs and average epitope number (AEN) analysis in Caucasoid and Oriental populations

Additional file4: table S4. SB epitope prediction of missense mutations and AEN analysis in Caucasoid and Oriental populations

Additional file 5: table S5. SB epitope prediction of frameshift mutations and AEN analysis in Caucasoid and Oriental populations

Additional file 6: table S6. SB epitope prediction of inframe insertion mutation and AEN analysis in Caucasoid and Oriental populations

Ethics approval and consent to participate

Not applicable.

Availability of data and materials

The cancer hot mutation relevant information can be downloaded at TCGA data portal (https://portal.gdc.cancer.gov) by Mutations/Consequence Type category.

The HLA allele frequency information can be download at the AFND (http://www.allelefrequencies.net) by HLA/HLA Allele Freq (Classical) category.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

Hu Wei, He Meifang and Li Liangping designed the study, performed data analysis and wrote the paper. All authors approved the manuscript.

Author details

1. The First Affiliated Hospital of Sun Yat-sen University, Department of Liver Surgery, Guangzhou, China

2. The First Affiliated Hospital of Jinan University, Institute of Clinical Oncology, Guangzhou, China

3. The First Affiliated Hospital of Sun Yat-sen University, Laboratory of General Surgery, Guangzhou, China

Acknowledgements

Thanks Zhou Qian from Clinical Trials Unit, the First Affiliated Hospital of Sun Yat-sen University for help in statistical analysis.

Funding

This study was supported by the National Natural Science Foundation of China (No. 81472824 and No. 81502464), the National Key R&D Program of China (No. 2016YFC1303404), the Guangdong Province Science and the Technology Project (No. 2015A020211009), the Guangzhou Science and Technology Project (No. 201604020133), and the Natural Science Foundation of Guangdong Province (No. 2016A030310186).

References

1. Tumeh PC, Harview CL, Yearley JH, Shintaku IP, Taylor EJ, Robert L, Chmielowski B, Spasic M, Henry G, Ciobanu V *et al*: PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature* 2014, 515(7528):568-571.

2. Tran E, Robbins PF, Lu YC, Prickett TD, Gartner JJ, Jia L, Pasetto A, Zheng Z, Ray S, Groh EM *et al*: T-Cell Transfer Therapy Targeting Mutant KRAS in Cancer. *N Engl J Med* 2016, 375(23):2255-2262.

3. Sahin U, Derhovanessian E, Miller M, Kloke BP, Simon P, Lower M, Bukur V, Tadmor AD, Luxemburger U, Schrors B *et al*: Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. *Nature* 2017, 547(7662):222-226.

4. Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, Lee W, Yuan J, Wong P, Ho TS *et al*: Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* 2015, 348(6230):124-128.

5. Ott PA, Hu Z, Keskin DB, Shukla SA, Sun J, Bozym DJ, Zhang W, Luoma A, Giobbie-Hurder A, Peter L *et al*: An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature* 2017, 547(7662):217-221.

6. Stronen E, Toebes M, Kelderman S, van Buuren MM, Yang W, van Rooij N, Donia M, Boschen ML, Lund-Johansen F, Olweus J *et al*: Targeting of cancer neoantigens with donor-derived T cell receptor repertoires. *Science* 2016, 352(6291):1337-1341.

7. Yee C, Lizee GA: Personalized Therapy: Tumor Antigen Discovery for Adoptive Cellular Therapy. *Cancer J* 2017, 23(2):144-148.

8. Thomas R, Al-Khadairi G, Roelands J, Hendrickx W, Dermime S, Bedognetti D, Decock J: NY-ESO-1 Based Immunotherapy of Cancer: Current Perspectives. *Front Immunol* 2018, 9:947.

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.

10. Gordeeva O: Cancer-testis antigens: Unique cancer stem cell biomarkers and targets for cancer therapy. *Semin Cancer Biol* 2018, 53:75-89.

11. Verdegaal EM, de Miranda NF, Visser M, Harryvan T, van Buuren MM, Andersen RS, Hadrup SR, van der Minne CE, Schotte R, Spits H *et al*: Neoantigen landscape dynamics during human melanoma-T cell interactions. *Nature* 2016, 536(7614):91-95.

12. Stevanovic S, Pasetto A, Helman SR, Gartner JJ, Prickett TD, Howie B, Robins HS, Robbins PF, Klebanoff CA, Rosenberg SA *et al*: Landscape of immunogenic tumor antigens in successful immunotherapy of virally induced epithelial cancer. *Science* 2017, 356(6334):200-205.

13. Parkhurst M, Gros A, Pasetto A, Prickett T, Crystal JS, Robbins P, Rosenberg SA: 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):2491-2505.

14. Maiers M, Gragert L, Klitz W: High-resolution HLA alleles and haplotypes in the United States population. *Hum Immunol* 2007, 68(9):779-788.

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.

16. Gjerstorff MF, Andersen MH, Ditzel HJ: Oncogenic cancer/testis antigens: prime candidates for immunotherapy. *Oncotarget* 2015, 6(18):15772-15787.

17. Andreatta M, Nielsen M: Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinformatics* 2016, 32(4):511-517.

18. Srivastava PK: Neoepitopes of Cancers: Looking Back, Looking Ahead. *Cancer Immunol Res* 2015, 3(9):969-977.

19. Klebanoff CA, Wolchok JD: Shared cancer neoantigens: Making private matters public. *J Exp Med* 2018, 215(1):5-7.

20. Hilf N, Kuttruff-Coqui S, Frenzel K, Bukur V, Stevanovic S, Gouttefangeas C, Platten M, Tabatabai G, Dutoit V, van der Burg SH *et al*: Actively personalized vaccination trial for newly diagnosed glioblastoma. *Nature* 2019, 565(7738):240-245.

21. Forni G: Vaccines for tumor prevention: a pipe dream? *J Infect Dev Ctries* 2015, 9(6):600-608.

Figures

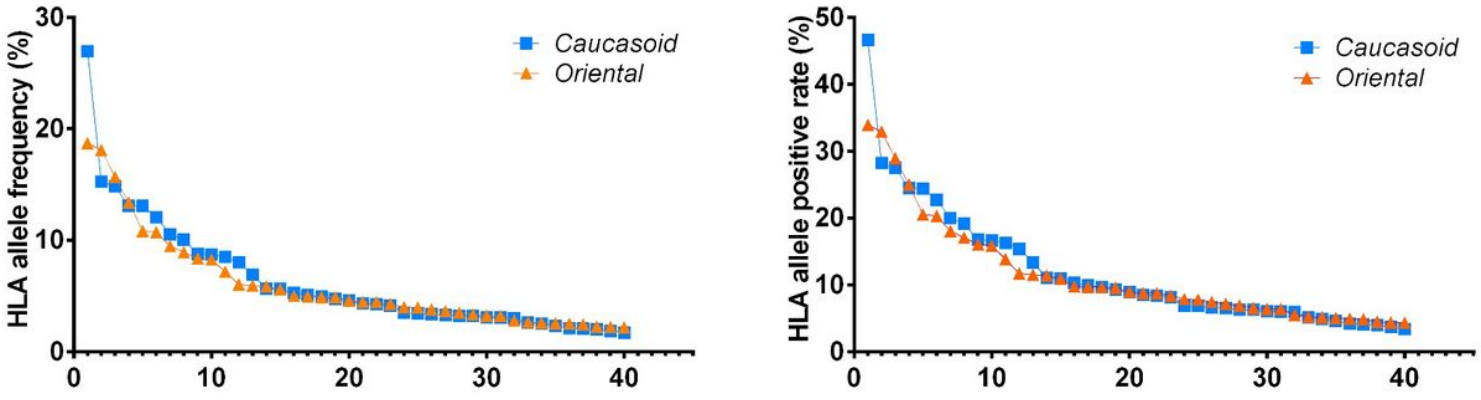

Figure 1

HLA allele frequency and positive rate in Caucasoid and Oriental population.

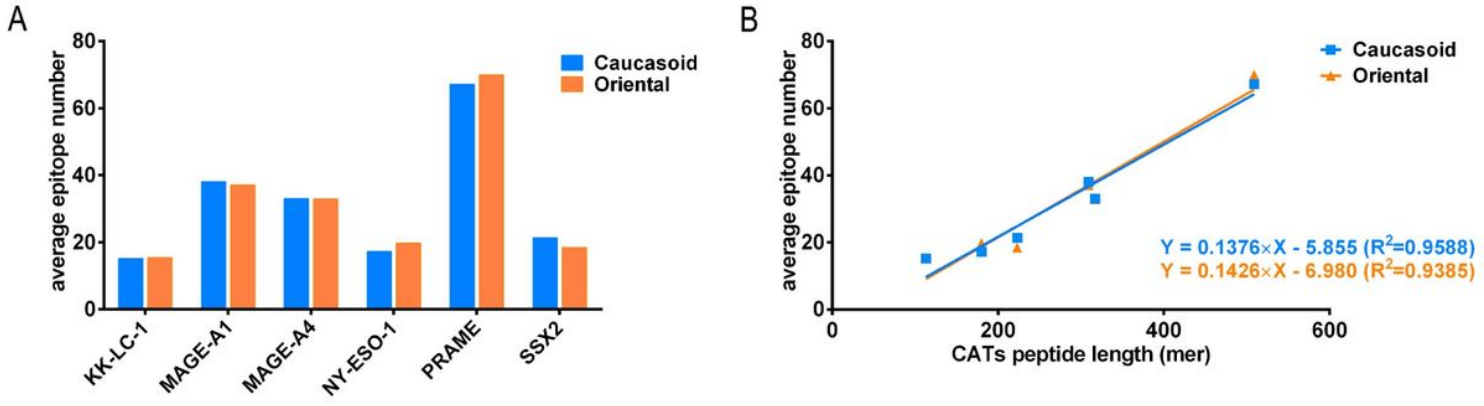

Figure 2

SB epitope prediction of CTAs in Caucasoid and Oriental population. A. Average epitope number of 6 CTAs. B. Linear analysis between average epitope number and antigen length. The linear equation and R square value are given in the figure.

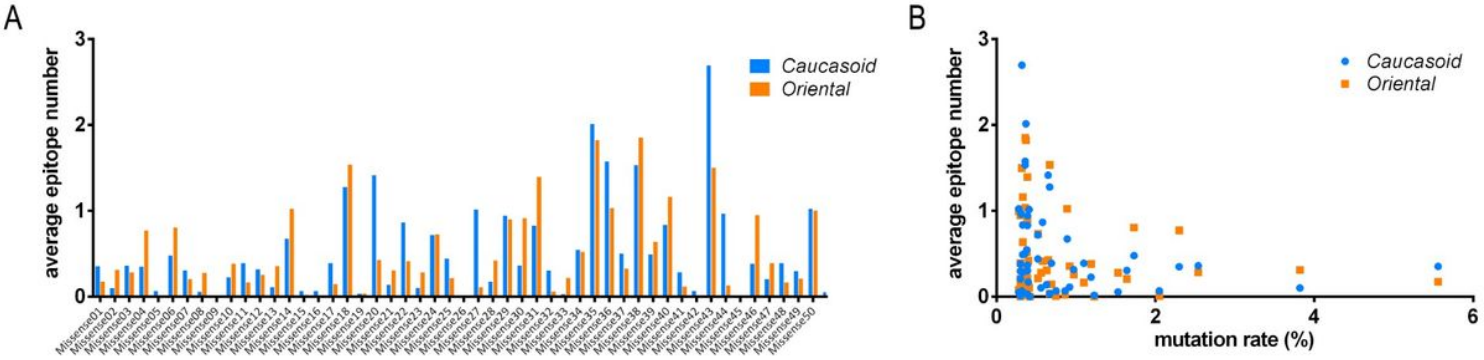

Figure 3

SB epitope prediction of single amino acid mutations in Caucasoid and Oriental population. A. Average epitope number of 50 missense mutations. B. Correlation analysis between average epitope number and mutation rate.

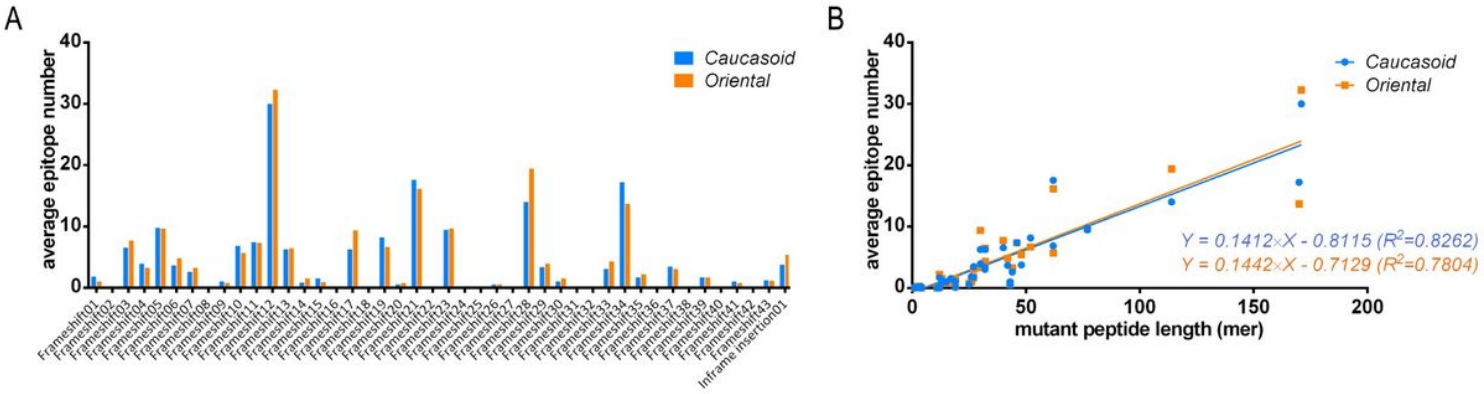

Figure 4

SB epitope prediction of frame related mutations in Caucasoid and Oriental population. A. Average epitope number of 43 frameshift mutations and 1 inframe insertion mutation. B. Linear analysis between average epitope number and mutant peptide length. The linear equation and R square value are given in the figure.

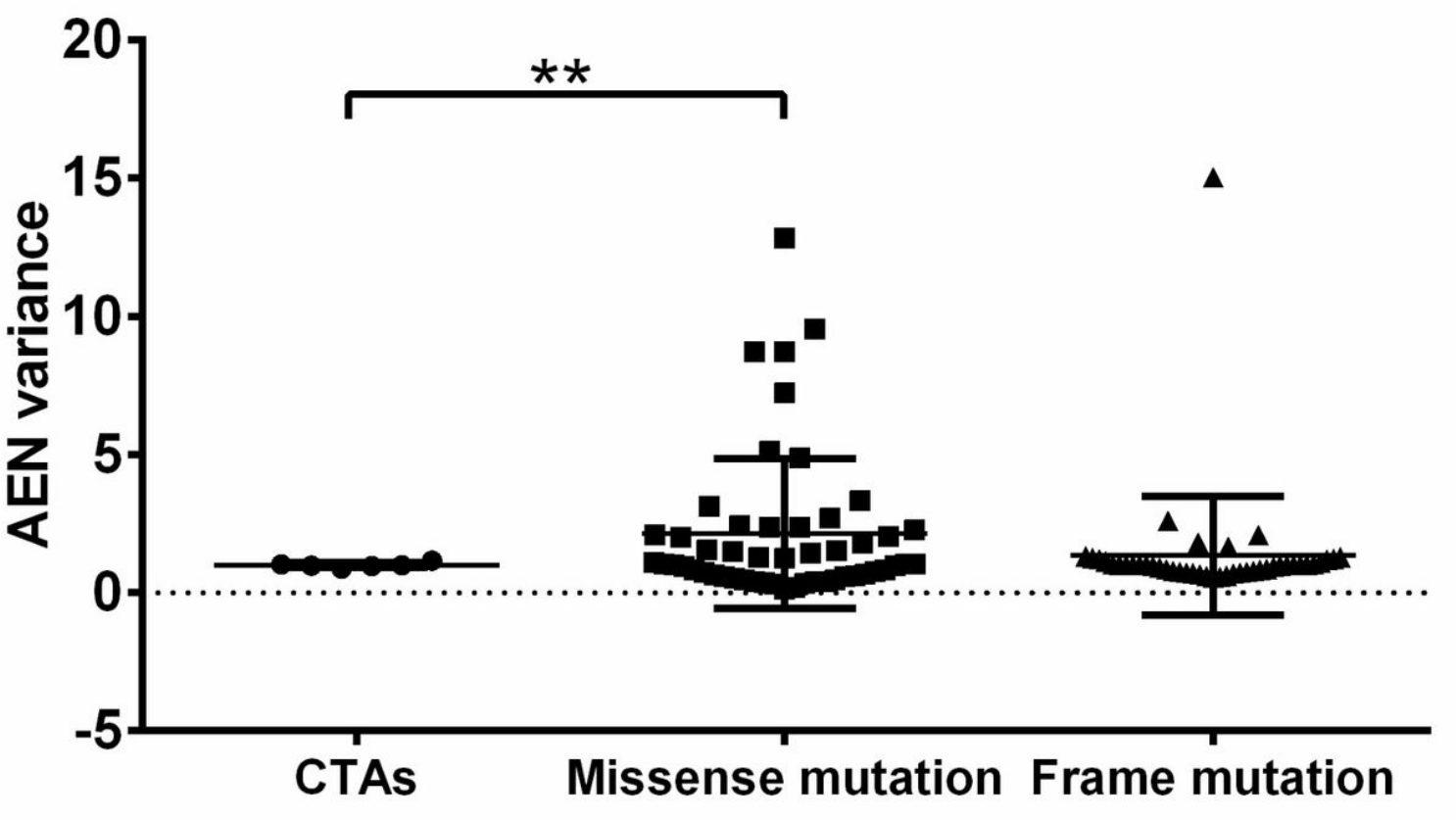

Figure 5

Comparison of average epitope number fold variances between Caucasoid and Oriental population 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 mutation in the diagram includes frameshift and inframe insertion mutations. Unpaired t-test was used, **p<0.01.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplement1.xlsx
- supplement2.xlsx
- supplement3.xlsx
- supplement4.xlsx
- supplement5.xlsx
- supplement6.xlsx