Identification of cryptic neoepitopes generated by chimeric transcripts in ovarian cancer

Vaishnavi M Dixit
National Centre for Cell Science, Savitribai Phule Pune University Campus
Arpita K Wagle
Department of technology , Savitribai Phule Pune University
Sharmila A Bapat (sabapat@nccs.res.in)
National Centre for Cell Science, Savitribai Phule Pune University Campus

Research Article

Keywords:

Posted Date: March 28th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1456664/v1

Abstract

Background: Considerable recent interest lies in the immunogenicity of non-canonical neoantigens generated by alternative ORFs emerging from non-synonymous mutations or nucleotide insertions and deletions, exonization of untranslated regions of the genome, non-coding transcripts, transposable elements, etc. Amongst these 'cryptic' epitopes, those derived from chimeric transcripts which harbor sequences derived from more than one gene remain undefined, since the translational potential and immunogenicity of these transcripts is relatively unexplored.

Methods: The present study was hence designed to predict and assess the neoepitopes generated by chimeric transcripts in ovarian cancer. Peptides identified in a customized targeted database were assessed for antigenicity. Further docking of peptide – MHC as well as peptide-MHC-TCR complexes provided proof-of-concept of antigenicity.

Results: We identified peptides uniquely generated by chimeric transcripts, expressed by reading-frame as well as sense/anti-sense isoforms. Further assessment of the antigenicity of these peptides identified a small number of putative neoepitopes that are likely to be presented to MHC-I in an allele-specific manner, and further recognized by specific TCRs.

Conclusions:

Such generation of chimeric molecules by relaxed rigidity of canonical transcription and translation may present opportunities in immunotherapy.

Introduction

The tractability of gene boundaries through gene and transcript fusions presents a wide horizon of transcript diversity, especially in aberrant cellular contexts such as cancer[1–3]. Chimeric transcripts (CTs) harboring sequences derived from more than one gene represents one such class of transcript diversity. The constancy of structural complexities of CTs across tumor types, arising from differential splicing and/or inclusion of regulatory, canonically untranslated sequences suggests a non-random mechanism of their origin[4–5]. While it is believed that a majority of such CTs are eliminated by cellular surveillance mechanisms including nonsense mediated decay, a few reportedly generate fusion proteins[6–7].

Current interests in cancer vaccines rely on discovery of effective tumor-specific neoepitopes[8–11], since they elicit T-cell responses that are not subject to host central tolerance, and also lower extent of autoimmune toxicities[12–13]. Conventionally, tumor-specific non-synonymous mutations 'reorganize' the ORFs of protein-coding genes and perturbcanonical protein sequences to generate neoepitopes that may be recognized and bound by specific T-cell receptors[14–16]. Personalized neoantigen discovery and validation remains a challenge since current algorithms rely on the binding affinity of putative neoantigens to HLA-I, while immunopeptidomics is complexed by the need for motif deconvolution in multi-allelic datasets[18–21]. Several cancers associated with a low mutational burden including acute

lymphoblastic leukemia, high-grade serous ovarian (HGSC), prostate and pancreatic cancer generate functional $CD8^+$ T cell responses suggesting the presence of neoantigens[22–23]. HGSC is considered as being immunologically 'cold' with a mutational landscape largely restricted to *TP53*, *BRCA1* and *BRCA2*[24]. Yet, a fraction of tumors presents with high tumor infiltrating lymphocytes, wherein a dominance of CD8 + cells correlate with improved prognosis[25]. This suggests a possibility of neoantigens being generated through other mechanisms.

On this background we hypothesized that chimeric transcripts identified in the TCGA ovarian cancer RNA sequencing datasets may be translated to generate novel chimeric, tumor-specific peptides, and a few of these presented as neoepitopes to CD8 + cells. We tested this by initially detecting their translational products in mass spectrometry datasets using a customized targeted chimeric transcript derived peptide database for identification of chimeric peptides. These peptides were further assessed for potential antigenicity *via* prediction of proteasome/ immunoproteasome processing, ferrying of degradation products to the endoplasmic reticulum by the transporter associated with antigen processing (TAP), recognition, binding to major histocompatibility complex class I (MHC-I) molecules for presentation and recognition by specific $CD8^+$ T-cell receptors (TCRs) that would mark the tumor cell for destruction. Identification of such chimeric epitopes highlights the cross-talks between relaxed transcription, translation and host immunity, which could be harnessed towards immunotherapy.

Methods

Identification of Chimeric Transcript Longest Read-derived peptides (cLRPs)

We identified several candidate chimeric transcripts (CTs) in TCGA RNA-Seq data from 160 OvCa samples applying a customized computational workflow on the Seven Bridges Cancer Genomics Cloud platform (SBgenomics; http://www.cancergenomicscloud.org). All spanning reads of each CT was consolidated to derive longest read (LR) CT sequences (ranged between 71-148 bp). To resolve the coding capabilities of CTs, we developed a customized chimeric LR peptide (cLRP) database comprising of *in silico*translation of all LR sequences in 6 reading frames(RF) along with ENSP protein sequences and randomly generated decoy peptides (Additional File 1;https://www.ensembl.org). This was appliedas a backend reference for probing ovarian tumor mass spectra datasets generated by the National Cancer Institute Clinical Proteomic Tumor Analysis Consortium (CPTAC; https://proteomics.cancer.gov/data-portal; Edwards et al. 2015), using the FenyoLab xtandem pipeline on the SBGenomics Cloud (Fig.S1a). This pipeline uses the X! Tandem search engine with pre-defined search parameters (Ruggles et al.2016; Ardeljan et al. 2019). Peptide Spectral Matches (PSMs) that qualified a confidence threshold were selected from outputs of this X!Tandem-based pipeline. Corresponding iTRAQ reporter ion intensities from spectral data (3 tumor samples and 1 site-specific pooled reference (labeled either with iTRAQ 117@PNNL/114@JHU) were extracted for computation of relative peptide abundances from sample enrichment ratios and summation of reporter ion intensities from different scans and bRPLC fractions; selecting for FDR values <0.01 (expect values 10^{-02}). Thence the order of magnitude of relative peptide

expression range was computed as log2 transformed total peptide abundance. Examination of the charge state of PSMs in terms of b- and y- ion fragmentation profiles was performed using Proteome Discoverer™ (Thermo Scientific™).

Prediction of antigenicity

Predictions of 9-11 residue degradation products of the tumor-specific cLRPs using the PCPS Proteasomal Cleavage Prediction Server (http://imed.med.ucm.es/Tools/pcps/;Gomez-Perosanz et al.2020) were followed by deployingTAPPred server for predicting binding affinity of peptides(IC50)towards the TAP transporter(http://crdd.osdd.net/raghava/tappred/; Raghava,2004) and generating TAP Scores. MHC Class-I alleles were selected for screening from two geographical regions *viz.*India (South Asia) and US (North America) from the HLA Allele Frequency Net Database (http://www.allelefrequencies.net;Gonzalez-Galarza et.al, 2020; A:B:C alleles - 245:404:120) as is outlined in detail in Additional File 2).Binding of peptides to MHC Class I was predictedusingNetmhcpan 4.1 that implements Neural Network Aligning (N-N Align) and is trained on Binding Affinity (BA) Data and Eluted Ligand (EL) datasets from mass spectrometry to present antigen to MHC molecules using concurrent motif deconvolution (process of associating ligand to MHC molecule), and rank peptides based on BA as a percentile score with respect to predication of top 100 peptides; hence lower BA rank corresponds to strong MHC I binders. Haplotype-based determination of the stability and affinity of potential neoantigens was determined by NetMHCstabpan1.0(NetMHCstabpan - 1.0 - Services - DTU Health Tech)that predicts stability and affinity of a peptide towards an allele; a threshold of 0.5% combined rank stability was set to identify peptides most likely to bind to MHC molecules with T-Half (Predicted Half Life) >2h and IC50 values < 100nM.

Derivation of Allele Harmonic Binding Rank (AHBR) and Peptide Harmonic Best Rank (PHBR)

PHBR represents the inverse average of a specific peptide expressed across patient samples, and was determined as follows –

$$PHBR = \frac{1}{(Average\ Relative\ Peptide\ Expression + x) * y}$$

where, x = number of all peptides from the corresponding cLRP detected in mass spectra from all tumor samples, y is a score for a predicted peptide being antigenic (score = 1 if the antigenic peptide is specifically detected in mass spectrometry data; score 2 if other peptides from the same cLRP are detected).

AHBR was calculated as inverse average of B.A Rank (Binding Affinity Rank) of all the MHC subtypes restricted by particular peptides i.e., ratio of summation of B.A Rank of restricted MHCs by respective peptide and total number restricted sub-alleles (a)

$$AHBR = \frac{1}{(Summation\ of\ Binding\ Affinity\ Rank\ /a)}$$

Peptide MHC Docking

Peptidestructures were designed inPepFold3D (RPBS Web Portal (univ-paris-diderot.fr);model 1 structure (by convention considered most stable) was visualized in PyMol 4.2 (PyMOL | pymol.org,The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC), and torsion angles edited using AutoDock 4.PDB structures of relevant and available HLA molecules were downloaded from IMGT (IMGT/3Dstructure-DB and IMGT/2Dstructure-DB Query page) and cleared of peptide ligands and artifacts from X-Ray crystallographic analysis using PyMol 4.2. These were submitted to CASTp [CASTp 3.0: Computed Atlas of Surface Topography of proteins (uic.edu)]for prediction of six peptide binding pockets within each molecule as below.

Table 1: MHC pocket (A-F) – associated residues derived from reported pdb structures

HLA Allele	PDB ID; References	MHC Pockets						References
		A	B	C	D	E	F	
HLA-A*11:01	2hn7 https://www.rcsb.org/structure/2HN7	Y7, Y171, Y159	E63	N66, Q155	W69, Y99			Blicher et.al,2019
HLA-B*27:09	1jgd https://www.rcsb.org/structure/1JGD	K1, Y159, K1, Y171	C67, E45, E63				L81, W147, Y123, D77, K70, H116	Hilling et al, 2004
HLA B*57:03	2hjl https://www.rcsb.org/structure/2HJL		Y9, E63, A24, N66, M45, M67				I94, I95, V97, Y99, V103, N114, Y116, R156	Kloverpris et al, 2011
HLA-B*15:01	6vb3 https://www.rcsb.org/structure/6VB3		E63, I66, S67, D70					Pavlos et al, 2017
HLA-A*24:02	2bck https://www.rcsb.org/structure/2BCK	E63	H70		K66	T73	Y84, T143	Cole et.al, 2006
HLA-B*58:01	5im7 https://www.rcsb.org/structure/5im7		Y9, E63, A24, N66, T45, M67, E46, N77					Kloverpris et al, 2011
HLA-B*27:05	5ib1 https://www.rcsb.org/structure/5ib1		E63, E45, R62				D116	Driller Et al,2019, Hilling et al, 2004

Grid box dimensions (size and coordinates) for peptide binding in the pdb structure of each MHC I molecule was defined in Auto dock 4 (Trott et al. 2010) applying the above pocket information as reference.The final pdbqt files were provided as the input for docking in AutodockVina, and outputs were visualized in PyMol and Discovery studio visualizer (BIOVIA, Dassault Systèmes, Discover Studio Visualizer, v21.1.0.20298 San Diego: Dassault Systèmes,2021). Successful docking outcomes presented the peptide centered within the defined grid.Ligand conformation and log file of affinity data (kCal/mol)

of the first pdb structure (lowest RMSD values) from nine conformations received, was considered as the best prediction. Complexes with affinity values (-9.00 to -6.00 kCal/mol) and polar bond length <3.00 A^o, electrostatic and hydrophobic bond length <4.00 A^o were considered stable complexes and their interactions studied with reference peptide:allele (R-pMHC) complexes derived through X-ray crystallography.

GibbsMotif analysis

Gibbs Cluster 2.0 [GibbsCluster-2.0 Server (dtu.dk)] was applied toinput test peptide sequences along with a set of validated restricting peptides for respective alleles obtained fromMHCBN: A comprehensive database of MHC binding and Non-binding peptides (osdd.net), VDJDb(cdr3.net) and published literature (Cole et.al 2006). Default MHC class I parameters were selected to identify clusters with highest Kullbach-Leiber distance (Andreatta et al. 2017)and probability scores to predict a binding motif for each allele.

Determination of Agretopicity Index andValidation of antigenicity of parental proteins

A modified agretopicity Index was derived to compare the binding affinities of antigenic epitopes with corresponding 8-11-mer amino acid sequences in the 6-frame *in silico*translated outputs of its transcript sequence.Shorter peptides generated in some cases due to the presence of stop codons in the alternative frames, were disregarded and only 9-mers were processed (NetMHCstabpan 1.0) for comparison of affinities with the same allele. Similarly, 6-frame protein isoforms of the specific parental transcripts generating the cLR of PSEN2-CABC1 transcript, *viz.* PSEN2-001 (ENST00000366783.3)andCABC1,ADCK3-004 (ENST00000366779.1; Ensembl Genome Browser,https://www.ensembl.org, Homo_sapiens - GRCh37)were predicted *in silico*, and each of the 12 generated proteins were processed through established pipeline for prediction of antigenicity as above.

Docking of p:MHC:TCR complexes

CDR3 sequences ofαandβchains with associated V and J genes of the available pMHC-TCR complexed with alleles HLA-A*11:01, HLA-A*24:02 and HLA-B*27:05 (Ladell, 2014) were derived from VDJ database (VDJdb :https://cdr3.net). This provided references of 2 TCR:viral (EBV) epitopes:HLA-A*11:01:2; 18 TCRs: human epitopes: HLA-A*24:02:4; and 1 TCR:viral (HIV) epitopes:HLA-B*27:05:4; (ImmuneScape VDJ Assembler,https://tcr2.cs.biu.ac.il/home;https://sysimm.ifrec.osaka-u.ac.jp/immune-scape/mhc1). 10 peptides restricting 3 alleles were modeled with respective complexed TCR, selecting TCR sequences where epitope of human origin was available; in case of epitopes of viral origin, TCR structures with minimum VDJdb acquisition Score of 1 were considered as structure of confidence. ERGOS II (https://biu.ac.il) was used to derive p:MHC:TCR binding scores for test peptide peptide:allele complex and CDR3 sequences ofαandβchain sequences with associated V and J genes. Autoencoder-based model along with VDJ Database as trainingset were provided as parameters. ImmuneScape VDJ chain assembler derived full length chain sequences of TCR-αβusing CDR3 and VJ genes were further provided for modeling in ImmuneScape Modeler along with epitope and HLA-Alleles to obtain PDB output, which in turn was provided as input to TCR 3D repertoire database (https://umd.edu) for derivation of Incident and

docking angles and TCR-CoM coordinates. Further, PyMol and Discovery Studio were used to visualize PDB structures for comparison of polar bond lengths and interacting residues with reference PDB structures obtained in VDJdb wherever available. while another complex structure was generated for HLA-A*11:01 towards evaluation of obtained TCR-p:MHC complexes.

Comparison with pre-reported neoantigens

Pre-validated neoantigens reported in Melanoma (Ott et.al,2020) and Glioblastoma (Keskin et.al, 2019) were processed through the same pipeline deployed in the present study. For themelanomadataset, areference peptide, DELEIKAYreported to restrict HLA-B*18:01(PDB ID:4XXChttps://www.rcsb.org/structure/4XXC)was selected for simulation of docking and comparison of interactions with predicted peptides.2 TCRs, *viz.* TRAV19*01-J3*01: TRBV20-1*01-J2-7*01; TRAV1-2*01-J32*01:TRBV18*01-J1-4*01 reported tocomplex with HLA-B*18:01 and epitopes of human origin (EEAAGIGIL, MEVDPIGHLY; VDJ assembler) were derived to execute the second level of molecular modeling was and identify similar interactions as those in respective references. Forglioblastoma, simulation of docking was performed and compared with reported reference, RRKWRRWHL for peptides predicted to restrict HLA-B*27:05 (https://www.rcsb.org/structure/5IB1) and further for secondary docking with TCRTRAV14/DV4/J21-TRBV6-5*01/J1-1*01reportedly complexed with HLA-B*27:05 and epitopes of Viral HIV origin (KRWIILGLNK; VDJ assembler) for comparison of interactions as those of reference.

Statistical Analysis

Two groups of patients were demarcated within the cohort based on their TcTP burden (lower and higher than median values of CTs and cLRPs), *viz.*Group 1 (n=51, low TcTP) and Group 2 (n=50, high TcTP).Kaplan Meier (K-M) plots were constructed using survival package in R and tested for significance by log-rank test ($p<0.05$). Differences between the K-M curves for the 2 groups were computed using survival probability as function of time and by inspecting the visual shape of plots, median overall survival (OS) and progression-free survival (PFS) in days. Between and within group analysis of variance (ANOVA) was performed for derived AHBR and PHBR values to segregate the variation of peptide distribution and allele restriction for two groups. Trends in distribution across G1 and G2 were analyzed through one-way ANOVA (Kruskal-Wallis H-test, performed using Microsoft Excel 2019 Office Analysis ToolPak), testing null hypothesis stating equal mean values of both groups which was rejected if calculated value (F-value) was observed to be greater than the critical chi-square value (F-crit) at $p<0.05$. Additional attributes (number of predicted antigenic peptides through NetMHCpan, number of these peptides detected in mass spectra, summation of relative abundance of cLRPs and MS detected antigenic peptides, allele restriction by NetMHCpan predicted antigenic peptides and mass spectrometry detected antigenic peptides, total AHBR-PHBR score, patient OS in days) were extracted and scaled based on their minimum to maximum values for visualization in MeV (Multiple Experiment Viewer v4.9).

Results

Translation of chimeric transcripts (CTs) in ovarian cancer generates tumor-specific (TS) peptides

A necessary first step was to identify peptides generated from chimeric transcript (CT) sequences around the fusionpoint identified in the TCGA ovarian cancer RNA-sequencing datasets. We consolidated all spanning reads of each CT to derive long read (LR) sequences for focused analyses rather than full-length transcript sequences since our working hypothesis was to evaluate chimeric peptides harboring residues of both parental proteins which were likely to be tumor-specific and perceived as non-self by the cellular machinery. Thus, 997 unique chimeric LR-derived peptides (cLRPs) were identified in ovarian tumor mass spectra datasets generated by the National Cancer Institute Clinical Proteomic Tumor Analysis Consortium[26] (CPTAC; https://proteomics.cancer.gov/data-portal) ;using a X! Tandem search engine-based pipeline and a targeted chimeric LR-derived peptide (cLRP) database[27–28] (Fig.S1a; Additional File1; Methods) on the Seven Bridges Cancer Genomics Cloud[29] (https://www.sevenbridges.com). Each tumor displayed a distinct profile of 28–132 cLRPs with a relative peptide expression range between 3.5-8 orders of magnitude (Fig. 1a-i). Since false-positive rates for PSM subgroups can vary (different peptide sizes, charge state of precursor ions, missed enzymatic cleavage, etc.), we affirmed chimerism by examining the charge state of PSMs in terms of b- and y- ion fragmentation profiles (Fig.S1b). While most peptides were transcribed from a single reading-frame (RF) without preferential usage of sense / antisense strands, RF isoforms of the same CT were also identified in some cases (Figs.S1c,S1d).

Tumor and tissue specificity was further inferred by probing expression of cLRPs in ovarian controls (OvCtrls; CPTAC data analyzed in parallel with tumor data), and human lymphoblastoid cell line data[30] (LBL; PXD001406). OvCtrls and LBLs displayed distinct cLRP profiles at frequencies of 29–49 and 0–9 /sample respectively, with corresponding relative peptide expression between 3.5–6.7 and 2-4.08 orders of magnitude. cLRP frequency as well as expression levels were lower in OvCtrls over primary tumors and even lower in LBL cultures (Figs. 1a-i,a-ii,a-iii). Peptide profiling designated 843 cLRPs to be Ovarian Tumor-Specific (TS), while remaining were Tumor-Associated (TA). Effectively, translation potential could be assigned to only 21.12% of the identified CTs as affirmed by the specificity imposed by ion fragmentation profiling. Such lower coding capability as compared to that of parental genes despite the enhanced peptide diversity *via* RF isoforms of a single CT may indicate either a non-coding regulatory nature of CTs, or elimination of newly formed proteins by cell surveillance and homeostatic mechanisms.

Tumor associated Chimeric Transcript-Protein (TcTP) burden may define MHC-I allele restriction by cLRPs and patient prognosis in a personalized manner

While oncogenic roles have been recently assigned to CTs[31], it is likely that several chimeric proteins may fail to achieve a stable conformation and thereby targeted for proteasomal degradation. Interestingly, recent findings suggest cryptic neoepitopes generated through proteasomal degradation bind to MHC class I molecules and are presented to CD8 + cells in eliciting anti-tumor responses[32–35]. To evaluate a

similar possible fate within the 844 tumor-specific cLRPs in our study, we first predicted their constitutive and immunoproteasome degradation products using the Proteasomal Cleavage Prediction Server followed by identifying good and moderate strong binding peptides to the transporter associated with antigen processing protein complex (TAP; binding scores 6–7, 7–8 and 8-9.5 respectively) on TAPPred server (Methods). Further presentation of these 526 9-mer peptides to MHC Class I molecules was screened using NetMHCPan 4.1 in a set of 769 high frequency restricted MHC Class-I alleles representing Indian and Caucasian populations derived from the HLA Allele Frequency Net Database[36] (Additional File 2). This identified 369 peptides that restrict 98.9% of selected MHC alleles with a Binding Affinity (BA) Rank </=0.5%; restriction being either between single or multiple peptide-alleles (Fig.S2; Tables S1, S2, S3). We also determined the Patient Harmonic Mean Best Rank (PHBR) and Allele Harmonic Binding Rank (AHBR) scores for these 369 putatively antigenic peptides, and correlated them with cLRP frequencies and relative expression in patients along with non-antigenic cLRP-derived peptides (Figs. 1b; Methods). This winnowed out a subset of 248 peptides (PHBR and AHBR < 0.3), whose wider distribution across patients and higher binding rank within the restricted HLA alleles indicated higher probability of their antigenicity (Fig. 1c; Table S4).

We further examined the association of CTs and cLRPs in terms of a tumor-associated Chimeric Transcript-Protein (TcTP) burden, with overall patient survival (OS) in the cohort. This identified 2 distinct groups; patients in Cluster1 with lower average OS presented with a lower TcTP burden (< 60 CTs, < 40 cLRPs), while Cluster 2 patients with significantly higher average OS presented with higher TcTP burden (> 65 CTs, > 55 cLRPs; Fig. 2a). As a corollary, differences in outcomes between patient groups in the TCGA cohort demarcated on the basis of median CT-cLRP values in the cohort as either Group 1 (lower TcTP burden) or Group 2 (higher TcTP burden) also indicated a chimerism-based survival advantage (Fig. 2b). Importantly, the variance between PHBR and AHBR within both groups was significant while only AHBR between the 2 groups was significant (Fig. 2c), emphasizing possible improved presentation and specificity in binding of cLRP to MHC-I alleles. Consolidation of all such parameters accentuated differences between the 2 groups of patients demarcated based on their TcTP burden (Fig. 2d-I). Lower TcTP (Group1) tumors were associated with fewer antigenic peptides and of a lower relative abundance, hence are likely to restrict fewer MHC-I alleles than Group 2 tumors with higher TcTP (Figs. 2d-II to 2d-VI). Most importantly, cLRP derived neoantigens as a determinant of patient allele(s) are suggested to influence prognosis; this allele-specificity of cLRPs could serve as protective factor in a personalized manner and effectively impart a survival advantage in at least a subset of ovarian cancer patients (Fig. 2d-VII). Overall, a certain threshold of TcTP burden may be essential to assign a CT-derived survival advantage, especially on the background of rare outliers in Group 1 that could imply other, unknown protective mechanisms.

restriction and patient prognosis in a personalized manner. a. Scatter plot representation of CT-cLRP distribution in tumors that clearly demarcates 2 clusters distinctly associated with overall survival (OS); b. KM survival plot of TcTP burden-based patient groups wherein Group 2 (higher TcTP burden) has a significant survival advantage over Group 1 (lower TcTP burden); c. Box notch plot of comparative AHBR-

PHBR scores (ANOVA) within and between Group 1 and Group 2 tumors. d. Heatmap representing the differential features of the 2 patient groups including - I:number of CTs (i) and cLRPs (ii), II: number of antigenic peptides predicted in NetMHCPan (i) or detected in MS (ii), III:Sum of relative peptide abundance of all cLRPs detected in MS, IV:Sum of relative peptide abundance of antigenic peptides detected in MS, V:alleles restricted by antigenic peptides predicted in NetMHCPan (i) or detected in MS (ii), VI:summation of PHBR (i) and AHBR (ii) scores of antigenic peptides, VII: Overall survival of patients; all values were mapped as per the scale bars

Positional amino acid binding preferences in cLRP-derived epitope-MHC complexes exhibit homology with binding pocket preferences of reference peptide:MHC structures

We increased the stringency of our prediction pipeline through haplotype-based determination of the stability and affinity of these potential neoantigens (NetMHCstabpan1.0; Methods), to identify a subset of 55 peptides restricting 117 HLA alleles at a high confidence and significant BA rank (p:MHC interactions; Thalf > 2; IC50 < 100nm; Table S5;Figs. 3a,3b;Fig.S3). HLA-C alleles were strikingly absent in this subset that represent cLRP derivatives most likely to harbor neoantigenic potential, including a highly stability ASCSVAWSW:HLA-B*57:26 complex (Thalf = 31.44h), while ATIRTVSSW:HLA-B*58:19 presented with highest affinity (IC50 = 3.12nM).

Further molecular modelling of the p:MHC interactions was performed using reported PDB allele structures (http://www.IMGT/3Dstructure-DB). 12 MHC allele structures were thus available within our list of restricted alleles, which interact with 27 peptides in 39 complexes, 27 of which display significant polar, electrostatic, and hydrophobic interactions (Figs.S4a,S4b). Within these, we performed Gibbs motif analyses for 7 complexes (for which binding pocket details were available in elucidated crystal p:MHC structures;Methods). This guided the assignment of positional residue specificity within cLR-p:MHC molecule interactions (Fig. 3c). Further comparisons *vis-à-vis* polar, electrostatic and hydrophobic interactions within these 7 complexes with reference to reference structures (RP:MHC) available for each allele revealed some identical interactions, amongst others involving interactions between a different peptide residue and the same MHC residue at a specific position (we termed these as MHC interactive residues; Table S6, Additional File 6 Video 1).

Thus, the high probability positional residues identified through Gibbs Cluster analysis anchor ATQGRSWRK in HLA-A*11:01 through interactions of A1 with T171 and E63 (pocket A, identical to reference), and K9 with Y99 (pocket D, MHC interactive residue); this complex is additionally stabilized through binding of T2 and Q3 with MHC interactive residues in pockets B and C respectively (Fig. 3d-i). Our observation of the KRMLASFSF:HLA-B*27:09 complex being stabilized through 3 polar and 1 electrostatic bonds between R2 and B pocket residues (E45,E63,T24, identical to reference) is supported by an earlier report indicating pocket B of HLA-B*27:09 to be sterically and electrostatically suited to bind to arginine[37]. This complex displays additional bonds between K1 and MHC interactive residues in pocket A that create an outward directed kink in the peptide and exposes remaining residues (Fig. 3d-ii). A high probability positional Y9 residue anchors RQRQKRIAY in pocket F of HLA B*15:01 identical to the

reference, along with additional bonds between R1 and MHC interactive residues in pockets A and B (Fig. 3d-iii). Pocket B of HLA-B*58:01 interacts with several residues of KQLLHSWKW including the high probability positional residues W9, S6 and K8; a predicted Y3-R97 salt bridge however may have a destabilizing effect (Fig. 3d-iv). The LAARPGPRW:HLA B*57:03 and IFWDIFCRF:HLA-A*24:02 models however did not compare favourably either with Gibbs analyses or comparative outputs with reference complexes (Figs.S4c-i,c-ii). The high probability positional R2 of RRTERAPRF:HLA-B*27:05 displayed interactions with MHC interactive residues in pockets E and F, with additional stabilization through interactions of Y3,P7 and F9 with MHC interactive residues (Fig.S4c-iii). Conclusively, commonalities identified between cLRP- and R- p:MHC structures *vis-à-vis* involvement and positional preferences of specific residues of peptides and MHC molecules, will contribute to their recognition as neoepitopes.

Antigenic epitopes present a low Agretopicity Index over corresponding sequence reading-frame / strand isoforms

We assessed the agretopicity index of each of the 7 predicted p:MHC complexes to compare the affinities of antigenic epitopes with their alternative (8–11)-mer reading-frame/strand isoforms (Mehods). All isoforms of the remaining peptides displayed lower affinity than the 7 test peptides (IC50 > 800-fold higher; Fig.S5a; Table S7), indicating the latter to be more robust binders to the respective alleles. We further extended this theme by evaluating the 6 reading-frame protein isoforms of full-length parental protein isoforms for antigenicity, as is exemplified in PSEN2-CABC1 that harbors the peptide ATQGRSWRK (Figs.S5b;5c;5d). This identified 6 peptides restricting HLA-A*11:01 (1 from PSEN2 and 5 from CABC1), all of which are derived from reading-frame / strand isoforms and none from the canonical parental proteins (Table S8). While all harbored the highest positional specificity binding residue K9 (predicted in Gibbs Motif analysis for HLA-A*11:01), none of these peptides except ATQGRSWRK were detected in mass spectrometry, suggesting that a single RF of this cLRP was translated. This also revealed ATQGRSWRK as not being a chimeric peptide, but derived from a frame-shift in PSEN2 transcript.

cLR-p:MHC:TCR complex stabilization involves dynamic interactions similar to reference structures

A next level of molecular modeling to examine interactions involved in recognition of cLR-p:MHC complexes by α and β chains of TCR was performed, which is the defining prelude to MHC-I mediated T-cell cytotoxicity. cLR-p:MHC:TCR complexes were generated considering the reported complementary CDR3 chain sequences available for 3 alleles (HLA-A*24:02:4, HLA-A*11:01:2, HLA-B*27:05:4). Comparison of these models with reference epitope interactions displayed relatively similar ERGO II scores across all complexes, with those for ATQGRSWRK:HLA-A*11:01 and KIINPIIRK:HLA-A*11:01 being indicated at a higher confidence (Fig.S6a; Table S9). Interactions of ATQGRSWRK:HLA-A*11:01 with 2 reported TCRs were examined and compared for stability and affinity with those of the 2 reference peptides (one each for MHC and TCR binding) *viz.* AIMPARFYPK:HLA-A*11:01[38] (RP:MHC) and HLA-A*11:01-TCR: IVTDFSVIK [39] (RP-TCR). ATQGRSWRK:HLA-A*11:01:TRAV21*01/J50*01-

TRBV6.6*01/J2.3*01 complex displayed a strong binding score with 4 polar interactions similar to those in the RP:TCR (R5-S98; peptide:TCR-β Figs. 4a-i,4a-ii) and RP:MHC (A1-Y7, A1-Y159 in A pocket; T2-E63 in B pocket; Fig. 4a-iii), which were further stabilized with *de novo* interactions including K9-D116 (MHC-F pocket; Fig. 4a-iv; Table S10). Alignment of crossing and incident angles (61.368° and 11.3246° respectively) with a smaller θ (x-axis angle of projection of the TCR:CoM–MHC:CoM vector; -33.41°) led to a diagonal placement of TCR on the cLR-p:MHC structure and facilitates a R5-S98 interaction (Fig. 4b,Fig.S6b, Additional File 7 Video 2). Visualization of polar, hydrophobic and electrostatic clouds indicated evasion of steric hinderance; specifically, presentation of R5 towards the outside facing plane (surface) of the MHC groove along with a lateral shift of its amino terminal within the same pocket makes it available for interaction with the TCR-β chain (Fig.S6c-i,6c-ii, left panels). TCR recognition and R5-S98 interaction also achieves a robust anchoring of the peptide deeper into the hydrophobic MHC pocket (Fig.S6c-i,56c-ii, right panels) enhancing stabilization of entire complex.

On the other hand, no interactions between peptide and TCR residues were evident in the ATQGRSWRK:HLA-A*11:01: TRAV35*01/J49*01- TRBV11-2*01/J1-2*01 complex, although most of the polar, electrostatic and hydrophobic p:MHC interactions were retained and geometrical alignment of crossing and incident angles were similar to those within the ATQGRSWRK:HLA-A*11:01:TRAV21*01/J50*01-TRBV6.6*01/J2.3*01 complex (

Table 11; Fig.S6d). This is likely due to the steric hindrance generated by an altered TCR:CoM–MHC:CoM vector (θ: -143.15°), which in turn results in a positional shift of S6 deeper into the MHC complex that collaterally decreases the possibility of R5 interacting with TCR–β. Comparative modeling by superimposing ATQGRSWRK:HLA-A*11:01 with the 2 reported TCRs further emphasized the planar shift and steric hinderance reducing the chances of p:TCR interactions within the A*11:01: TRAV35*01/J49*01- TRBV11-2*01/J1-2*01 complex (Fig. 4c). Effectively, this highlights prioritization of neoepitopes through its recognition and binding to the TCR by, (i) determining the orientation and conformation of a firmly anchored peptide in the MHC molecule with specific exposed residues, and (ii) extent of perturbations within the p:MHC-I complex due to the proximity of a TCR through steric hinderance.

Similar assessment of earlier reported immunizing mutated neoantigens affirms prediction analyses

We further compared our evaluation with reported, tumor-specific mutated immunizing neoantigens in melanoma[40] and glioblastoma[41]. Processing of the reported immunizing mutated peptides in these studies through our stability- and affinity-based prediction identified 6 and 24 antigenic peptides in melanoma and glioblastoma respectively (Figs. 5a-i,5b-i; Table S11;Figs.S7a;S7b;S7c). Molecular modeling of some of these, for example NEVSEVTVF-HLA-A*18:01 (melanoma) revealed polar interactions (V8-R97;F9-W147) similar to those in reference p:MHC complex (Fig. 5a-ii; Table S12). Following TCR recognition and binding, the NEVSEVTEF:HLA-B*18:01:TRAV1-2*01-J32*01:BV18*01-J1-4*01 complex displayed a polar p:MHC interaction (E2-S24) identical with the reference (DELEIKAY:HLA-

B*18:01), and unique p:TCR-α (polar:S4-G92) and p:TCR-β (electrostatic:E8-K51;Fig. 5a-iii) interactions. A second complex NEVSEVTEF: HLA-B*18:01: TRAV19*01-J3*01:BV20-1*01-J2-7*01 was characterized with similar polar p:MHC interaction (E8-W147) and unique p:TCR-α (polar:S4-R96) and p:TCR-β (electrostatic:E8-R99) interactions as compared with reference complexes.(Fig. 5b-iv).

The glioblastoma peptides AAHRARYFW and VSAAHRARY restricting HLA-B*27:05 display an overlap with the reported peptide ARYFWGNLA, which presents 3 polar p:MHC interactions (A2-E63, A2-R62, A2-Y159;Fig.S7d-i), and complexes with TRAV14/DV4/J21-TRBV6-5/J1-1 through three similar polar, and single electrostatic and hydrophobic p:MHC interactions (A1-Y171, A2-E63, A2-Y159, W9-D77, A1-W167 respectively) and exclusive polar (R4-P28, R6-N98) and hydrophobic (A5-Y52) interactions with TCR-α (Fig.S7d-ii,iii; Table S12). VSAAHRARY displayed 6 similar polar p:MHC interactions (V1-Y171,V1-Y159,S2-E63,S2-R62,Y9-D77,Y9-W147) and a single hydrophobic interaction (V1-W167) as that of reported peptide, and interacts with TCR TRAV14/DV4/J21-TRBV6-5/J1-1 through five similar polar pMHC (V1-Y171,V1-Y7,V1-Y159,S2-E63,A3-Y99), and exclusive polar interactions in TCR-β (A7-G98) and TCR-α (H5-Y52), and two exclusive electrostatic interactions TCR-α (R6-E95, R8-E30; Figs. 5b-ii;5b-iii;5b-iv). A few more peptides in both tumor datasets were indicated to be antigenic through Gibbs cluster motif analysis and docking (data not shown). Taken together, our prediction of neoepitopes generated from cLRPs represents a robust approach suggesting chimeric transcript derived proteins and their reading frame isoforms to present cryptic epitopes in tumors.

Discussion

Genomic sequencing followed by evaluation of immunogenicity is a recognized neoantigen prediction approach towards personalized cancer therapy, especially in tumors with a high mutation burden[42]. An underlying assumption is that while most intragenic mutations would not alter gene expression and translation capabilities, they generate a tumor-specific protein modified, which could harbour a neoepitope capable of evoking host immunity. Chimeric transcripts on the other hand, being considered a fall-out of aberrant transcriptional and/or splicing machinery in a cell, often may not be translated and have not been evaluated as a source of neoantigens. The mass spectrometry-based identification of peptides derived from chimeric reads using a customized database was hence an essential prelude in our antigenicity prediction, and assigned protein-coding potential to around a fifth of the chimeras produced, a large majority of which were tumor-specific. While we began with the premise of screening only the chimeric peptides with residues of both partners, a realization was that chimera generation may also mediate frame-shift of either / both parental partners besides using of the anti-sense strand, to generate additional neoantigens. A further study towards derivation of full-length chimeric transcripts and mass-spectrometry based identification of their translation products to explore for neoantigenecity would widen the horizons of the present study.

Our assessment of the frequency of chimeric transcripts and peptides vis-a-vis a Chimeric Transcript-Protein (TcTP) burden indicated influences on overall patient survival in a personalized manner through MHC-I allele-specific restriction by some peptides. Modelling the latter at a molecular level revealed the

complex dynamics of peptide recognition and presentation, which is challenging since a few thousand human MHC-I alleles are known, and individuals can express as many as six distinct alleles, each with different epitope affinities. We hope that since our predictions were monoallelic, validation may widen the horizons of neoantigenicity to enhance the repertoire of peptide recognition. Allele – based patient stratification may thus identify individuals with an immune checkpoint blockade (ICB) benefit, as is reported in non-small-cell lung cancer and melanoma[43]. At the other end of the personalization spectrum, patients expressing HLA-I molecules with promiscuous peptide binding capabilities are indicated to display a significantly worse prognosis after ICB[44]. The specificity in recognition assigned through our modelling of the TCR interactions shifts the focus to peptide-TCR interactions as a definitive complementation to MHC-I based predictions and thereby strengthens the neoepitope prediction pipeline.

Conclusion

The field of chimeric neoepitopes is thus nascent as yet, and will require several further approaches as in case of other tumor-specific neoepitopes, to realize its full potential. Attrition of host responses through *in situ* systems influences of an immunosuppressive microenvironment (T-regs, cytokine balance, etc), inactivation of MHC-I and loss of antigen presentation, T-cell apoptosis/ exhaustion / anergy, altered immune checkpoints, etc. that define personalized features besides a patient's alleles will continue to present challenges[45]. Considered broadly, our findings including studying the conformation and dynamics of peptide-MHC-TCR interactions to qualify a *bona fide* neoepitope from chimeric transcript-derived peptides also contribute to understanding why only some of the several predicted neoantigens are likely to elicit immunogenicity; this will be useful in developing novel, immunogenicity based personalized therapies.

Abbreviations

HGSC - High-grade serous ovarian tumors
ORF – Open Reading Frame
CTs - Chimeric transcripts
MHC-I - Major Histocompatibility Complex class I
TCR-T-cell receptor
TS-Tumor-Specific
LR- Long Read
cLRPs- Chimeric LR-derived Peptides
CPTAC- Clinical Proteomic Tumor Analysis Consortium
RF- Reading Frame
PSMs-Peptide Spectral Matches
OvCtrls- Ovarian Controls
LBL- Lymphoblastoid
TA- Tumor-Associated

AFND- Allele Frequency Net Database
BA Rank – Binding Affinity Rank
EL -Eluted Ligand
PHBR-Patient Harmonic Mean Best Rank
AHBR-Allele Harmonic Binding Rank
TOA- Tumor Ovarian control associated
TLA- Tumor-LBL associated
TcTP- Tumor-associated Chimeric Transcript-Protein
OS- Overall Survival
PFS- Progression Free Survival
pMHC- peptide MHC
RP
MHC- Reference Peptide MHC
cLR-p
MHC- Chimeric Longest Read peptide-MHC
cLR-p
MHC TCR- Chimeric Longest Read peptide-MHC TCR
RP-TCR Reference Peptide-TCR
TCR
CoM– TCR Centre of Mass
MHC
CoM- MHC Centre of Mass
PCPS- Proteasomal Cleavage Prediction Server
VDJdb -VDJ database
KM-Kaplan Meier
ANOVA- Analysis of Variance

Declarations

Ethical Approval and Consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The materials that support the conclusion of this review have been included within the article.

Competing interests

The authors declare no competing interests.

Funding

This work was supported by the Tata Innovation Fellowship of the Department of Biotechnology, Government of India (BT/HRD/35/01/04/2017), Fulbright-Nehru grants of USIEF (Award No. 2055/F-NAPE/2015), ICMR-DHR lnternational Fellowship for Senior Bio-medical Scientists (INDO/FRC/452(S-60y2019-20-lHD). VMD was granted project research fellowship by Department of Biotechnology, (India BT/HRD/35/01/04/2017).

Author Contributions –

SAB conceptualized and planned the study design, performed and curated proteomics analysis, VMD performed, analysed and interpreted experimental data and wrote the manuscript draft AW analysed proteomics data using Proteome Discoverer™ and generated b-y ion plots. SAB supervised and interpreted experimental data and edited the manuscript draft to its final form.

Acknowledgements –

Research was supported by the Tata Innovation Fellowship of the epartment of Biotechnology, Government of India (BT/HRD/35/01/04/2017), Fulbright-Nehru grants of USIEF (Award No. 2055/F-NAPE/2015), ICMR-DHR lnternational Fellowship for Senior Bio-medical Scientists (INDO/FRC/452(S-60;2019-20-lHD). VMD was granted project research fellowship by Department of Biotechnology, (India BT/HRD/35/01/04/2017).Data used in this publication were generated by the Clinical Proteomic Tumor Analysis Consortium (NCI/NIH) and is acknowledged. We extend our gratitude to Prof. David Fenyö and Dr. McKerrow, for sharing and explaining workflows of the FenyoLab xtandem pipeline on the SBGenomics Cloud. The collaborative grant with Seven Bridges Genomics ($10,000 of compute and storage credits is deeply appreciated). The Seven Bridges Cancer Research Data Commons Cloud Resource has been funded in whole or in part with Federal funds from the National Cancer Institute, National Institutes of Health, Contract No. HHSN261201400008C and ID/IQ Agreement No. 17X146 under Contract No. HHSN261201500003I and 75N91019D00024Technical assistance from Mr. Avinash Mali and support from the NCCS Proteomics facility are gratefully acknowledged. All figures in this manuscript were created using BioRender.com

Authors' information

Affiliations:

Ovarian Cancer Program, National Centre for Cell Science, Savitribai Phule Pune University Campus, Ganeshkhind, Pune 411007, India

Authors - Vaishnavi M Dixit, Sharmila A Bapat (Corresponding Author)

Department of Technology, Savitribai Phule Pune University, Ganeshkhind, Pune 411007, India.

Author - Arpita K Wagle

References

1. Magnuson B, Bedi K, Ljungman M. Genome stability versus transcript diversity. *DNA Repair*. 2016;44:81–86. doi:10.1016/j.dnarep.2016.05.010
2. Parris TZ. Pan-cancer analyses of human nuclear receptors reveal transcriptome diversity and prognostic value across cancer types. *Scientific Reports*. 2020;10(1). doi:10.1038/s41598-020-58842-6
3. Cheng S, Li Z, Gao R, Xing B, Gao Y, Yang Y, Qin S, Zhang L, Ouyang H, Du P, Jiang L, Zhang B, Yang Y, Wang X, Ren X, Bei JX, Hu X, Bu Z, Ji J, Zhang Z. A pan-cancer single-cell transcriptional atlas of tumor infiltrating myeloid cells. Cell. 2021 Feb 4;184(3):792–809.e23. doi: 10.1016/j.cell.2021.01.010
4. Gorohovski A, Tagore S, Palande V, Malka A, Raviv-Shay D, Frenkel-Morgenstern M. ChiTaRS-3.1-the enhanced chimeric transcripts and RNA-seq database matched with protein-protein interactions. *Nucleic Acids Research*. 2017;45(D1):D790-D795. doi:10.1093/nar/gkw1127
5. Mukherjee S, Detroja R, Balamurali D, et al. Computational analysis of sense-Antisense chimeric transcripts reveals their potential regulatory features and the landscape of expression in human cells. *NAR Genomics and Bioinformatics*. 2021;3(3). doi:10.1093/nargab/lqab074
6. Jividen K, Li H. Chimeric RNAs generated by intergenic splicing in normal and cancer cells. *Genes Chromosomes and Cancer*. 2014;53(12):963–971. doi:10.1002/gcc.22207
7. Urbanski LM, Leclair N, Anczuków O. Alternative-splicing defects in cancer: Splicing regulators and their downstream targets, guiding the way to novel cancer therapeutics. Wiley Interdiscip Rev RNA. 2018 Jul;9(4):e1476. doi: 10.1002/wrna.1476.
8. Cafri G, Gartner JJ, Zaks T, et al. mRNA vaccine–induced neoantigen-specific T cell immunity in patients with gastrointestinal cancer. *Journal of Clinical Investigation*. 2020;130(11):5976–5988. doi:10.1172/JCI134915
9. Brohl AS, Sindiri S, Wei JS, Milewski D, Chou HC, Song YK, Wen X, Kumar J, Reardon HV, Mudunuri US, Collins JR, Nagaraj S, Gangalapudi V, Tyagi M, Zhu YJ, Masih KE, Yohe ME, Shern JF, Qi Y, Guha U, Catchpoole D, Orentas RJ, Kuznetsov IB, Llosa NJ, Ligon JA, Turpin BK, Leino DG, Iwata S, Andrulis IL, Wunder JS, Toledo SRC, Meltzer PS, Lau C, Teicher BA, Magnan H, Ladanyi M, Khan J. Immuno-transcriptomic profiling of extracranial pediatric solid malignancies. Cell Rep. 2021 Nov 23;37(8):110047. doi: 10.1016/j.celrep.2021.110047. PMID: 34818552; PMCID: PMC8642810.
10. Verdon DJ, Jenkins MR. Identification and targeting of mutant peptide neoantigens in cancer immunotherapy. *Cancers*. 2021;13(16). doi:10.3390/cancers13164245
11. Roesler AS, Anderson KS. Beyond Sequencing: Prioritizing and Delivering Neoantigens for Cancer Vaccines. Methods Mol Biol. 2022;2410:649–670. doi: 10.1007/978-1-0716-1884-4_35. PMID: 34914074.

12. McGranahan N, Furness AJ, Rosenthal R, Ramskov S, Lyngaa R, Saini SK, Jamal-Hanjani M, Wilson GA, Birkbak NJ, Hiley CT, Watkins TB, Shafi S, Murugaesu N, Mitter R, Akarca AU, Linares J, Marafioti T, Henry JY, Van Allen EM, Miao D, Schilling B, Schadendorf D, Garraway LA, Makarov V, Rizvi NA, Snyder A, Hellmann MD, Merghoub T, Wolchok JD, Shukla SA, Wu CJ, Peggs KS, Chan TA, Hadrup SR, Quezada SA, Swanton C. Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. Science, 351–6280:1463–9 (2016)
13. Jhunjhunwala S, Hammer C, Delamarre L. Antigen presentation in cancer: insights into tumour immunogenicity and immune evasion. *Nature Reviews Cancer*. 2021;21(5):298–312. doi:10.1038/s41568-021-00339-z
14. Parkhurst M, Gros A, Pasetto A, et al. Isolation of T-cell receptors specifically reactive with mutated tumor-associated antigens from tumor-infiltrating lymphocytes based on CD137 expression. Clinical Cancer Research. 2017;23(10):2491–2505. doi:10.1158/1078-0432.CCR-16-2680
15. Veatch JR, Lee SM, Fitzgibbon M, Chow IT, Jesernig B, Schmitt T, Kong YY, Kargl J, Houghton AM, Thompson JA, McIntosh M, Kwok WW, Riddell SR. Tumor-infiltrating BRAFV600E-specific CD4 + T cells correlated with complete clinical response in melanoma. J Clin Invest, 128(4):1563–1568 (2018)
16. Joglekar A v., Li G. T cell antigen discovery. *Nature Methods*. 2021;18(8):873–880. doi:10.1038/s41592-020-0867-z
17. Ghorani E, Rosenthal R, McGranahan N, et al. Differential binding affinity of mutated peptides for MHC class I is a predictor of survival in advanced lung cancer and melanoma. *Annals of Oncology*. 2018;29(1):271–279. doi:10.1093/annonc/mdx687
18. Wilson EA, Anderson KS. Lost in the crowd: identifying targetable MHC class I neoepitopes for cancer immunotherapy. *Expert Review of Proteomics*. 2018;15(12):1065–1077. doi:10.1080/14789450.2018.1545578
19. Alvarez B, Barra C, Nielsen M, Andreatta M. Computational Tools for the Identification and Interpretation of Sequence Motifs in Immunopeptidomes. *Proteomics*. 2018;18(12). doi:10.1002/pmic.201700252
20. Freudenmann LK, Marcu A, Stevanović S. Mapping the tumour human leukocyte antigen (HLA) ligandome by mass spectrometry. *Immunology*. 2018;154(3):331–345. doi:10.1111/imm.12936
21. Hilf N, Kuttruff-Coqui S, Frenzel K, Bukur V, Stevanović S, Gouttefangeas C, Platten M, Tabatabai G, Dutoit V, van der Burg SH, Thor Straten P, Martínez-Ricarte F, Ponsati B, Okada H, Lassen U, Admon A, Ottensmeier CH, Ulges A, Kreiter S, von Deimling A, Skardelly M, Migliorini D, Kroep JR, Idorn M, Rodon J, Piró J, Poulsen HS, Shraibman B, McCann K, Mendrzyk R, Löwer M, Stieglbauer M, Britten CM, Capper D, Welters MJP, Sahuquillo J, Kiesel K, Derhovanessian E, Rusch E, Bunse L, Song C, Heesch S, Wagner C, Kemmer-Brück A, Ludwig J, Castle JC, Schoor O, Tadmor AD, Green E, Fritsche J, Meyer M, Pawlowski N, Dorner S, Hoffgaard F, Rössler B, Maurer D, Weinschenk T, Reinhardt C, Huber C, Rammensee HG, Singh-Jasuja H, Sahin U, Dietrich PY, Wick W. Actively personalized vaccination

trial for newly diagnosed glioblastoma. Nature. 2019 Jan;565(7738):240–245. doi: 10.1038/s41586-018-0810-y. Epub 2018 Dec 19. Erratum in: Nature. 2019 Feb;566(7745):E13. PMID: 30568303.

22. Zamora AE, Crawford JC, Allen EK, et al. Pediatric patients with acute lymphoblastic leukemia generate abundant and functional neoantigen-specific CD8 + T cell responses. *Science Translational Medicine*. 2019;11(498). doi:10.1126/scitranslmed.aat8549
23. Yang W, Lee KW, Srivastava RM, Kuo F, Krishna C, Chowell D, Makarov V, Hoen D, Dalin MG, Wexler L, Ghossein R, Katabi N, Nadeem Z, Cohen MA, Tian SK, Robine N, Arora K, Geiger H, Agius P, Bouvier N, Huberman K, Vanness K, Havel JJ, Sims JS, Samstein RM, Mandal R, Tepe J, Ganly I, Ho AL, Riaz N, Wong RJ, Shukla N, Chan TA, Morris LGT. Immunogenic neoantigens derived from gene fusions stimulate T cell responses. Nat Med. 2019 May;25(5):767–775. doi: 10.1038/s41591-019-0434-
24. Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. Nature. 2011 Jun 29;474(7353):609 – 15. doi: 10.1038/nature10166. Erratum in: Nature. 2012 Oct 11;490(7419):298. PMID: 21720365; PMCID: PMC3163504.
25. Shih IM, Wang Y, Wang TL. The Origin of Ovarian Cancer Species and Precancerous Landscape. *American Journal of Pathology*. 2021;191(1):26–39. doi:10.1016/j.ajpath.2020.09.006
26. Edwards NJ, Oberti M, Thangudu RR, et al. The CPTAC data portal: A resource for cancer proteomics research. *Journal of Proteome Research*. 2015;14(6):2707–2713. doi:10.1021/pr501254j
27. Ruggles K v., Tang Z, Wang X, et al. An analysis of the sensitivity of proteogenomic mapping of somatic mutations and novel splicing events in cancer. *Molecular and Cellular Proteomics*. 2016;15(3):1060–1071. doi:10.1074/mcp.M115.056226
28. Ardeljan D, Wang X, Oghbaie M, et al. LINE-1 ORF2p expression is nearly imperceptible in human cancers. *Mobile DNA*. 2019;11(1). doi:10.1186/s13100-019-0191-2
29. Lau JW, Lehnert E, Sethi A, et al. The cancer genomics cloud: Collaborative, reproducible, and democratized - A new paradigm in large-scale computational research. *Cancer Research*. 2017;77(21):e3-e6. doi:10.1158/0008-5472.CAN-17-0387
30. Battle A, Khan Z, Wang SH, et al. Impact of regulatory variation from RNA to protein. *Science*. 2015;347(6222):664–667. doi:10.1126/science.1260793
31. Shi X, Singh S, Lin E, Li H. Chimeric RNAs in cancer. In: *Advances in Clinical Chemistry*. Vol 100. Academic Press Inc.; 2021:1–35. doi:10.1016/bs.acc.2020.04.001
32. Karandikar SH, Sidney J, Sette A, Selby MJ, Korman AJ, Srivastava PK. Identification of epitopes in ovalbumin that provide insights for cancer neoepitopes. *JCI Insight*. 2019;4(8). doi:10.1172/jci.insight.127882
33. Erhard F, Dölken L, Schilling B, Schlosser A. Identification of the cryptic HLA-I immunopeptidome. *Cancer Immunology Research*. 2020;8(8):1018–1026. doi:10.1158/2326-6066.CIR-19-0886
34. Roudko V, Bozkus CC, Orfanelli T, McClain CB, Carr C, O'Donnell T, Chakraborty L, Samstein R, Huang KL, Blank SV, Greenbaum B, Bhardwaj N. Shared Immunogenic Poly-Epitope Frameshift Mutations in Microsatellite Unstable Tumors. Cell. 2020 Dec 10;183(6):1634–1649.e17. doi: 10.1016/j.cell.2020.11.004. Epub 2020 Nov 30. PMID: 33259803; PMCID: PMC8025604.

35. Javitt A, Shmueli MD, Kramer MP, Kolodziejczyk AA, Cohen IJ, Kamer I, Litchfield K, Bab-Dinitz E, Zadok O, Neiens V, Ulman A, Radomir L, Wolf-Levy H, Eisenberg-Lerner A, Kacen A, Alon M, Rêgo AT, Stacher-Priehse E, Lindner M, Koch I, Bar J, Swanton C, Samuels Y, Levin Y, da Fonseca PCA, Elinav E, Friedman N, Meiners S, Merbl Y. The proteasome regulator PSME4 drives immune evasion and abrogates anti-tumor immunity in NSCLC. bioRxiv 2021. 10.24.464690 ; doi: https://doi.org/10.1101/2021.10.24.464690
36. Gonzalez-Galarza FF, McCabe A, Santos EJM dos, et al. Allele frequency net database (AFND) 2020 update: Gold-standard data classification, open access genotype data and new query tools. *Nucleic Acids Research*. 2020;48(D1):D783-D788. doi:10.1093/nar/gkz1029
37. Hillig RC, Hülsmeyer M, Saenger W, et al. Thermodynamic and Structural Analysis of Peptide- and Allele-dependent Properties of Two HLA-B27 Subtypes Exhibiting Differential Disease Association. *Journal of Biological Chemistry*. 2004;279(1):652–663. doi:10.1074/jbc.M307457200
38. Blicher T, Kastrup JS, Pedersen LØ, Buus S, Gajhede M. Structure of HLA-A*1101 in complex with a hepatitis B peptide homologue. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*. 2006;62(12):1179–1184. doi:10.1107/S1744309106044228
39. Zhang QJ, Gavioli R, Klein G, Masuccit MG. *An HLA-A11-Specific Motif in Nonamer Peptides Derived from Viral and Cellular Proteins*. Vol 90.; 1993.
40. Ott PA, Hu Z, Keskin DB, Shukla SA, Sun J, Bozym DJ, Zhang W, Luoma A, Giobbie-Hurder A, Peter L, Chen C, Olive O, Carter TA, Li S, Lieb DJ, Eisenhaure T, Gjini E, Stevens J, Lane WJ, Javeri I, Nellaiappan K, Salazar AM, Daley H, Seaman M, Buchbinder EI, Yoon CH, Harden M, Lennon N, Gabriel S, Rodig SJ, Barouch DH, Aster JC, Getz G, Wucherpfennig K, Neuberg D, Ritz J, Lander ES, Fritsch EF, Hacohen N, Wu CJ. An immunogenic personal neoantigen vaccine for patients with melanoma. Nature. 2017 Jul 13;547(7662):217–221. doi: 10.1038/nature22991. Epub 2017 Jul 5. Erratum in: Nature. 2018 Mar 14;555(7696):402. PMID: 28678778; PMCID: PMC5577644.
41. Keskin DB, Anandappa AJ, Sun J, Tirosh I, Mathewson ND, Li S, Oliveira G, Giobbie-Hurder A, Felt K, Gjini E, Shukla SA, Hu Z, Li L, Le PM, Allesøe RL, Richman AR, Kowalczyk MS, Abdelrahman S, Geduldig JE, Charbonneau S, Pelton K, Iorgulescu JB, Elagina L, Zhang W, Olive O, McCluskey C, Olsen LR, Stevens J, Lane WJ, Salazar AM, Daley H, Wen PY, Chiocca EA, Harden M, Lennon NJ, Gabriel S, Getz G, Lander ES, Regev A, Ritz J, Neuberg D, Rodig SJ, Ligon KL, Suvà ML, Wucherpfennig KW, Hacohen N, Fritsch EF, Livak KJ, Ott PA, Wu CJ, Reardon DA. Neoantigen vaccine generates intratumoral T cell responses in phase Ib glioblastoma trial. Nature. 2019 Jan;565(7738):234–239. doi: 10.1038/s41586-018-0792-9. Epub 2018 Dec 19. PMID: 30568305; PMCID: PMC6546179.
42. Hu Z, Leet DE, Allesøe RL, Oliveira G, Li S, Luoma AM, Liu J, Forman J, Huang T, Iorgulescu JB, Holden R, Sarkizova S, Gohil SH, Redd RA, Sun J, Elagina L, Giobbie-Hurder A, Zhang W, Peter L, Ciantra Z, Rodig S, Olive O, Shetty K, Pyrdol J, Uduman M, Lee PC, Bachireddy P, Buchbinder EI, Yoon CH, Neuberg D, Pentelute BL, Hacohen N, Livak KJ, Shukla SA, Olsen LR, Barouch DH, Wucherpfennig KW, Fritsch EF, Keskin DB, Wu CJ, Ott PA. Personal neoantigen vaccines induce persistent memory T

cell responses and epitope spreading in patients with melanoma. Nat Med. 2021 Mar;27(3):515–525. doi: 10.1038/s41591-020-01206-4.

43. Cummings AL, Gukasyan J, Lu HY, Grogan T, Sunga G, Fares CM, Hornstein N, Zaretsky J, Carroll J, Bachrach B, Akingbemi WO, Li D, Noor Z, Lisberg A, Goldman JW, Elashoff D, Bui AAT, Ribas A, Dubinett SM, Rossetti M, Garon EB. Mutational landscape influences immunotherapy outcomes among patients with non-small-cell lung cancer with human leukocyte antigen supertype B44. Nat Cancer. 2020 Dec;1(12):1167–1175. doi: 10.1038/s43018-020-00140-1.
44. Manczinger M, Koncz B, Balogh GM, Papp BT, Asztalos L, Kemény L, Papp B, Pál C. Negative trade-off between neoantigen repertoire breadth and the specificity of HLA-I molecules shapes antitumor immunity. Nat Cancer. 2021 Sep;2(9):950–961. doi: 10.1038/s43018-021-00226-4.
45. Oliveira G, Stromhaug K, Klaeger S, Kula T, Frederick DT, Le PM, Forman J, Huang T, Li S, Zhang W, Xu Q, Cieri N, Clauser KR, Shukla SA, Neuberg D, Justesen S, MacBeath G, Carr SA, Fritsch EF, Hacohen N, Sade-Feldman M, Livak KJ, Boland GM, Ott PA, Keskin DB, Wu CJ. Phenotype, specificity and avidity of antitumour CD8<background-color:#CCCCFF;supvertical-align:sub;>+</background-color:#CCCCFF;supvertical-align:sub;> T cells in melanoma. Nature. 2021 Aug;596(7870):119–125. doi: 10.1038/s41586-021-03704-y.

Figures

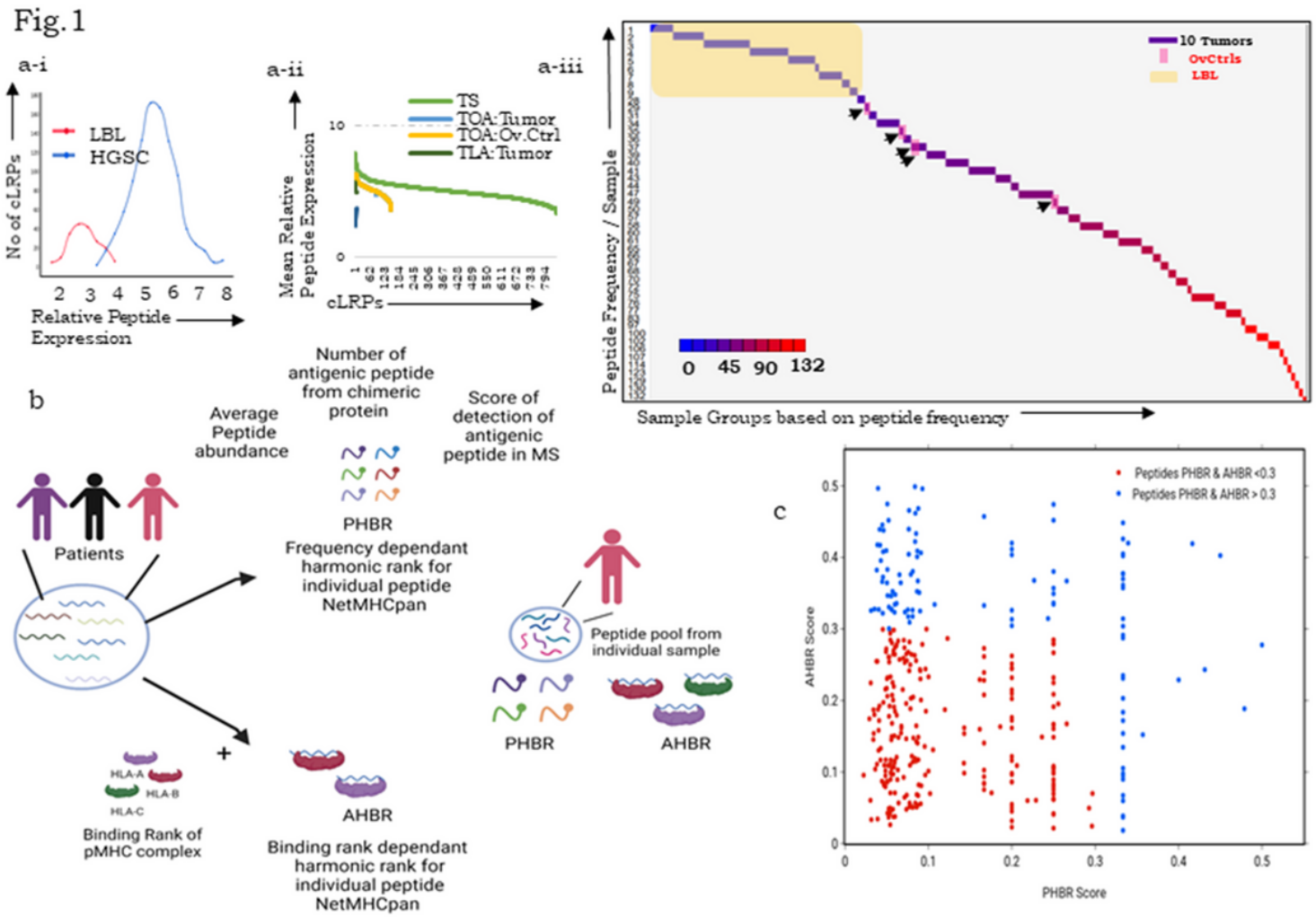

Figure 1

Identification and characterization of chimeric longest read-derived peptides (cLRPs) in OvCa. a-i. Distribution of relative peptide expression in high-grade serous ovarian tumor (HGSC) and lymphoblastoid (LBL) mass spectral data; a-ii. Relative peptide expression of cLRPs including 843 tumor specific (TS), 149 tumor- ovarian control associated (TOA), 4 tumor-LBL associated (TLA); a-iii. cLRP frequency per sample (range 1-132 represented as 48 groups on Y-axis), lymphoblastoid (LBL,yellow highlight) and ovarian control (OvCtrl, pink highlight; arrow) associated peptides are expressed at lower frequencies than those in ovarian tumors (squares, each square representing 10 tumors), blue to red color scale indicates peptide frequency; b. Schematic indicating determination of Patient Harmonic Mean Best Rank (PHBR) and Allele Harmonic Binding Rank (AHBR); c. Scatter plot of 369 peptides with PHBR & AHBR <0.3 (red; 248 high-confidence petides) and others with PHBR & AHBR > 0.3 (blue)

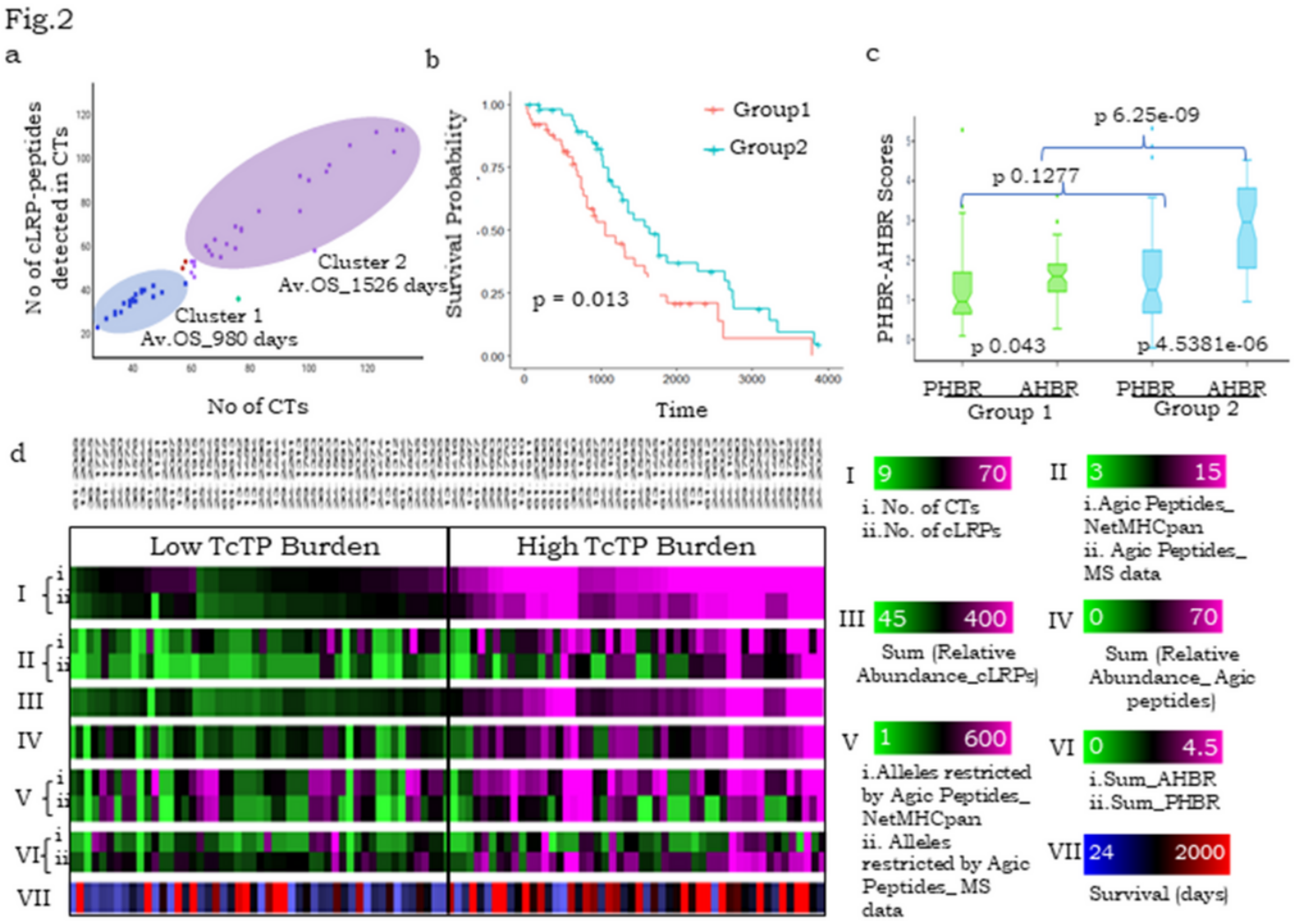

Figure 2

Tumor associated Chimeric Transcript-Protein (TcTP) burden may define allele

restriction and patient prognosis in a personalized manner. a. Scatter plot representation of CT-cLRP distribution in tumors that clearly demarcates 2 clusters distinctly associated with overall survival (OS); b. KM survival plot of TcTP burden-based patient groups wherein Group 2 (higher TcTP burden) has a significant survival advantage over Group 1 (lower TcTP burden); c. Box notch plot of comparative AHBR-PHBR scores (ANOVA) within and between Group 1 and Group 2 tumors. d. Heatmap representing the differential features of the 2 patient groups including - I:number of CTs (i) and cLRPs (ii), II: number of antigenic peptides predicted in NetMHCPan (i) or detected in MS (ii), III:Sum of relative peptide abundance of all cLRPs detected in MS, IV:Sum of relative peptide abundance of antigenic peptides detected in MS, V:alleles restricted by antigenic peptides predicted in NetMHCPan (i) or detected in MS (ii), VI:summation of PHBR (i) and AHBR (ii) scores of antigenic peptides, VII: Overall survival of patients; all values were mapped as per the scale bars

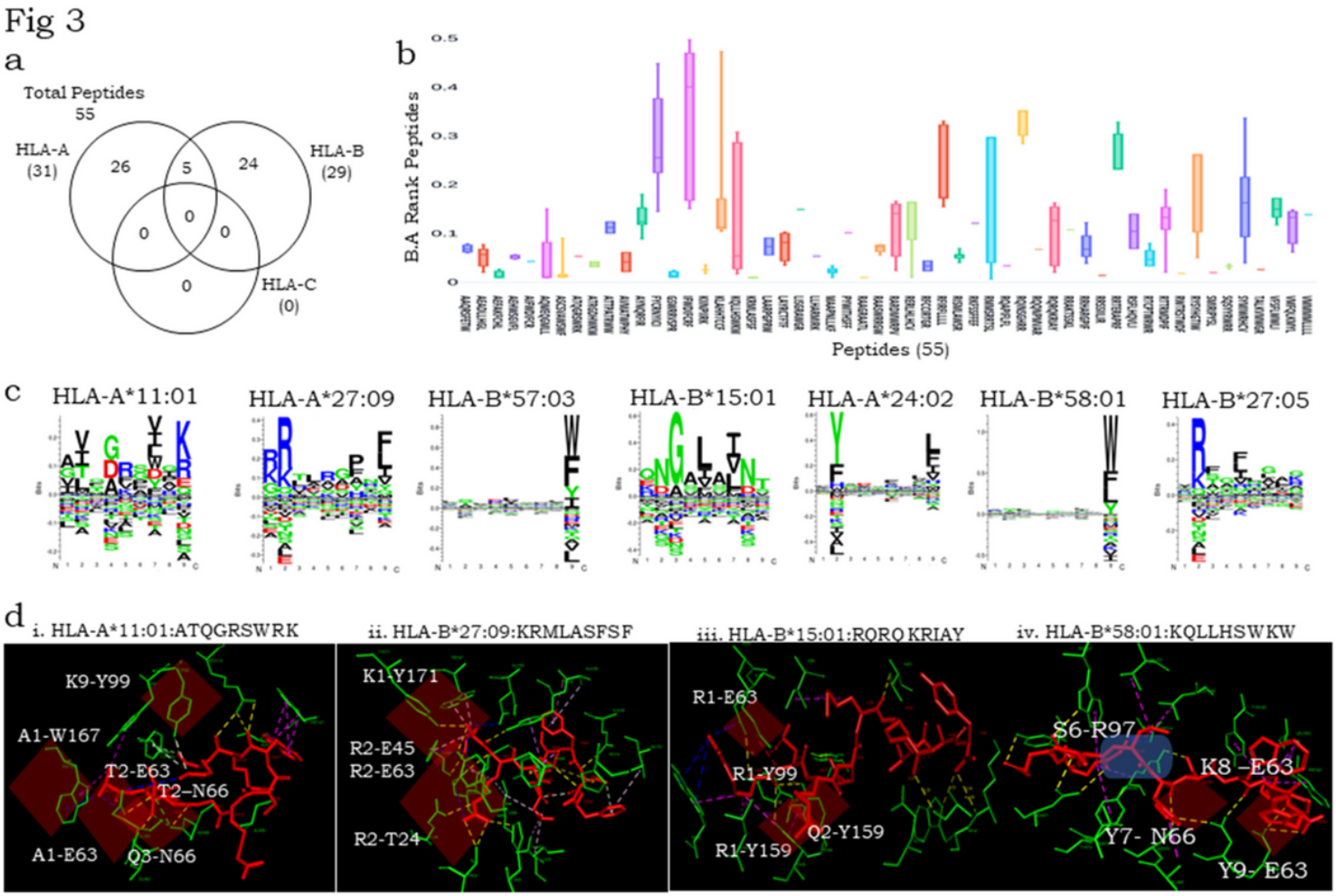

Figure 3

a. Haplotype-based allele restriction (NetMHCstabpan) identified 55 high confidence antigenic peptides; b. Binding Affinity (BA) Rank distribution of these 55 peptides; c.Gibbs Cluster motif generated for 7 MHC molecules HLA-A*11:01; HLA-A*27:09; HLA-B*57:03; HLA-B*15:01; HLA-A*24:02; HLA-B*58:01; HLA-B*27:05 predicting positional significance of specific amino acid residues based on reported p:MHC structures; d. Predicted p:MHC interactions modelled in Discovery studio where peptide residues (red) interact with MHC residues (green) through polar (yellow), electrostatic (blue) and hydrophobic (pink) bonds, interactions are highlighted in a red cloud and salt bridge in a blue cloud, some of these interactions were also indicated in models generated in PyMOL (Table S6).

Figure 4

Molecular modeling of ATQGRSWRK:HLA-A*11:01: TRAV21*01/J50*01-TRBV6.6*01/J2.3*01 complex (peptide-red, MHC-green, TCR-pink, polar and electrostatic interactions - green and orange bond respectively) a. Comparison of interactions with the reference models, i. 3 similar polar p:MHC interactions (A1-159, A1-Y7, T2-E63) as in RP:MHC, ii. R5-S98 polar interaction similar to RP-TCR; iii. *De*

novo K9-D116 electrostatic interaction; b. Geometric parameters in the cLR-p:MHC:TCR complex where Φ=7.43° (angle between TCR:CoM and p:MHC:CoM vector), r=29.1 (vector connecting TCR:CoM and p:MHC:CoM) and Θ=-33.41 (x-axis angle of projection of TCR:CoM–MHC:CoM vector); c. Comparative modeling indicating the steric hinderance for R5-S98 interactions observed by superimposing ATQGRSWRK (Cyan):HLA-A*11:01 with the 2 reported TCRs v*iz.* TRAV21*01/J50*01-TRBV6.6*01/J2.3*01 (α chain in blue, β chain in red) and TRAV35*01/J49*01-TRBV11-2*01/J1-2*01 (α chain in pink, β chain in yellow; Additional File 6,7 Videos 1,2).

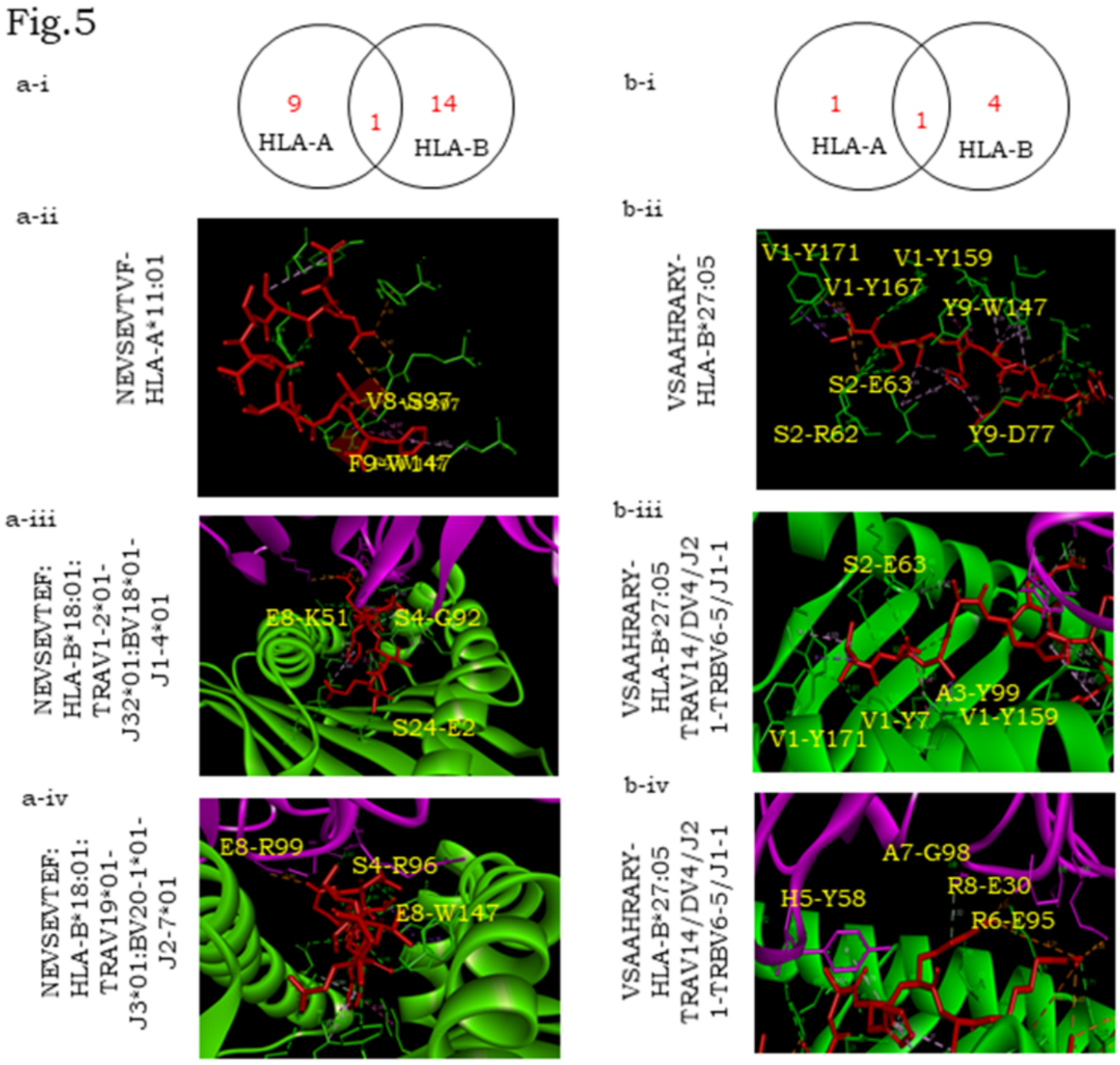

Figure 5

a-i. Venn-diagram representing HLA distribution analyses (NetMHCstabpan 1.0 to predict high stability-affinity peptides) of known immunizing mutated peptides in melanoma; a-ii.Molecular modeling of NEVSEVTVF-HLA-A*11:01 complex (peptide-red, MHC-green) indicating two significant similar interactions; a-iii. Molecular modeling of NEVSEVTEF: HLA-B*18:01: TRAV1-2*01-J32*01:BV18*01-J1-4*01 (peptide-red, MHC-green, TCR-pink) indicating one single interaction each of peptide with MHC (polar:E2-S24), TCR-α (polar:S4-G92), TCR- β (electrostatic:E8-K51); a-iv. Molecular modeling of NEVSEVTEF: HLA-B*18:01: TRAV19*01-J3*01:BV20-1*01-J2-7*01 (peptide-red, MHC-green, TCR-pink) indicating one single interaction each of peptide with MHC (polar:E8-W147), TCR-α (polar:S4-R96), TCR- β (electrostatic:E8-R99); b-i. Venn-diagram representing HLA distribution analyses (NetMHCstabpan 1.0 to predict high stability-affinity peptides) of known immunizing mutated peptides in glioblastoma; b-ii. Representative VSAAHRARY-HLA-B*27:05 complex where peptide (red) interacts with MHC (green) through six similar polar (V1-Y171, V1-Y159, S2-E63, S2-R62, Y9-D77, Y9-W147) and single hydrophobic (V1-W167) bonds; b-iii. Molecular modeling of complex VSAAHRARY-HLA-B*27:05 TRAV14/DV4/J21-TRBV6-5/J1-1 (peptide-red, MHC-green, TCR-pink) indicating peptide interactions with MHC (5 similar polar:V1-Y171, V1-Y7, V1-Y159, S2-E63, A3-Y99), TCR-α (2 electrostatic:R6-E95, R8-E30) and TCR-β (2 polar: A7-G98, TCR-α H5-Y52); b-iv. Molecular modeling of VSAAHRARY-HLA-B*27:05 TRAV14/DV4/J21-TRBV6-5*01/J1-1*01 (peptide-red, TCR-pink) where peptide interacts with TCR-α (2 electrostatic: R6-E95, R8-E30) and TCR-β (2 polar: A7-G98, TCR-α H5-Y52)

Supplementary Files

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

- AdditionalFile1cLRPhumanproteomeDecoy.fasta
- AdditionalFiles23.docx
- AdditionalFile4.csv
- AdditionalFile5.pdf
- AdditionalFile6.mp4
- AdditionalFile7.mp4