Mauriello *et al.*
Journal of Translational Medicine (2025) 23:558
https://doi.org/10.1186/s12967-025-06478-3

Journal of Translational Medicine

RESEARCH Open Access

Shared neoantigens' atlas for off-the-shelf cancer vaccine development

Angela Mauriello[1†], Beatrice Cavalluzzo[1†], Concetta Ragone[1], Maria Tagliamonte[1*] and Luigi Buonaguro[1*]

Abstract

Background We have recently described that the most prevalent 100 mutations identified in human cancers, both single nucleotide variations (SNVs) and InDels, generate a handful number of shared mutated neoantigens (SNV and InDel-NeoAgs) in association with 5 HLA-A and 7 B haplotypes.

Methods In the present study, we expanded such analysis to 50 haplotypes in the three MHC class I loci (10 HLA-A, 27 HLA-B and 13 HLA-C), including all the mutated proteins identified in at least 5% of cancer patients.

Results Overall, the extended analysis identified 15 SNV-NeoAgs and 55 InDel-NeoAgs with a significant affinity improvement over the corresponding wt (DAI > 10). These targetable shared NeoAgs are prevalently derived from $PIK3CA_{H1047R}$ (6/15 SNV-NeoAgs) and $LARP4B_{T163Hfs}$ (30/55 InDel-NeoAgs). From the HLA perspective, the HLA-A*33:03 is associated with the largest number of SNV-NeoAgs (4/15 NeoAgs) and the HLA-B*58:01 is associated with the largest number of InDel-NeoAgs (16/55 NeoAgs).

According to the distribution of each HLA haplotype in at least 10% of the regional populations, therapeutic cancer vaccines based on mutated shared SNV and InDel-NeoAgs, might be developed for COAD, STAD and UCEC cancers, with a global coverage, and for PAAD and UVM, with a regional coverage.

Conclusions This represents the first in-depth analysis for the identification of a specific repertoire of shared mutated NeoAgs, most of which never reported before. Such shared SNV and InDel-NeoAgs are indispensable for the development of "off-the-shelf" cancer vaccines targeting a relevant percentage of cancers in a significant percentage of cancer patients worldwide.

Keywords Mutations, Shared neoantigens, Off-the-shelf cancer vaccines, T cell immunity

Introduction

Driver mutations in specific proteins confer a growth advantage on the cellular survival or proliferation, possibly leading to transformation and cancer development [1–5]. Single nucleotide variants (SNVs), occurring at the first or second position of the codon, may result into a single amino acid substitution in the protein sequence (nonsynonymous change or a missense variant). Short insertions or deletions (InDels), if not a multiple of three bp, will result in a frameshift and a complete change in the reading frame of the downstream sequence of the gene. This will alter the product of translation, potentially leading to targeted decay of the alternative mRNA [6–8].

The resulting SNV or InDel mutation may fall in a peptide generated by the proteasome and loaded onto major histocompatibility complex (MHC) class I molecules by the transporter associated with antigen processing (TAP). When the peptide carrying the mutation is compatible with the cellular HLA allele, the stable peptide-MHC

†Angela Mauriello and Beatrice Cavalluzzo have contributed equally to this work.

*Correspondence:
Maria Tagliamonte
m.tagliamonte@istitutotumori.na.it
Luigi Buonaguro
l.buonaguro@istitutotumori.na.it
[1] Innovative Immunological Models Unit, Istituto Nazionale Tumori IRCCS – "Fond G. Pascale", Via Mariano Semmola, 52, Naples, Italy

(pMHC) complex moves to the cell surface and cancer cells will present tumor-specific mutated antigens ("neoantigens"). They are scrutinized by the immune system and, if identified as non-self, become targets for immune-mediated destruction [9].

Consequently, tumor-specific mutated neoantigens represent the most specific and potent non-self immunogens to be used for developing cancer vaccines [10, 11]. Indeed, they would elicit a T cell immune response that can exclusively target the tumor while sparing healthy tissue [12]. To this aim, the identification of tumor-specific mutated neoantigens shared among patients with the same tumor or different type of tumors would represent the "holy grail" of cancer immunotherapy. Such neoantigens would be valuable tools for off-the-shelf vaccines, but only if they are shared among a substantial number of patients and presented by common HLA alleles. Unfortunately, this does not seem to be the case in the driver mutations found in all solid tumors at The Cancer Genome Atlas (TCGA) [13].

A list of about 20 putative shared mutated neoantigens, out of more than 1 million screened nonsynonymous missense mutations, have been previously predicted in at least 5% of patients in one or more cancer types [14, 15]. However, the selection parameters used to define such shared neoantigens raise few doubts. By definition, a "neoantigen" is a mutated epitope that, for a given HLA allele, is a strong binder and the wt nonmutated counterpart is not a binder. Alternatively, the ratio of MHC binding affinity between the mutant and normal peptide, namely differential agretopicity index (DAI), should be > ten [16]. Only neoantigens identified according to such a definition have been shown to correlate with intratumoral T-cell responses and predict patient survival [14]. Indeed, given the T cell receptor degeneration, if the efficiency of antigen presentation is very similar (e.g. DAI < 10), the mutated neoantigens would not be seen as non-self and suffer from the same immunological tolerance of the self nonmutated counterpart [17].

Along the same path of searching for shared mutated neoantigens, our group has recently performed a prediction analysis on the most frequent 100 mutations reported at the TCGA, which collectively occur in 56.65% of all cancer cases [18]. Moreover, these include the driver missense and frameshift mutations found in more than 5% of patients affected by a specific cancer or shared by more than 5% of cancers. Among others, were analyzed the $BRAF_{V600E}$ (found in > 40% of melanoma and > 60% of thyroid ca), $KRAS_{G12D}$ (found in > 30% of pancreas ca), $IDH1_{R132H}$ (found in > 35% of brain ca) and $GNA11_{Q209L}$ (found in > 40% of uveal melanoma). The neoantigen prediction was performed taking into consideration the most frequent 12 HLA-A (5) and B (7) alleles and selecting only the mutated peptides predicted to have very strong affinity (< 100 nM) while the corresponding non-mutated wt peptide show very low (DAI > 10) or no affinity to the same allele. Based on such stringent parameters, the results returned only 10 predicted neoantigens from 7 missense mutations (SNV-NeoAgs) and 9 predicted neoantigens from 6 frameshift mutations (InDel-NeoAgs). Of these, only the $GNA11_{Q209L}$ FRMVDVGG**L** SNV-NeoAg may have a relevant application as off-the-shelf vaccine in > 40% of uveal melanoma (UVM) cases when positive for the HLA-B*27:05 or 39:01.

The aim of the present study was to expand such a prediction analysis to a much broader number of alleles, including HLA-A, B and C haplotypes. A total of 50 alleles (10 HLA-A, 27 HLA-B and 13 HLA-C) were considered to predict neoantigens from the SNV and InDel mutations found in more than 5% of patients affected by a specific cancer. The findings show that, based on the distribution of specific HLA haplotypes present in at least 10% of regional populations, therapeutic cancer vaccines targeting shared SNV and InDel-derived neoantigens could be developed with global coverage for colon adenocarcinoma (COAD), stomach adenocarcinoma (STAD), and uterine corpus endometrial carcinoma (UCEC), and with regional applicability for pancreatic adenocarcinoma (PAAD) and uveal melanoma (UVM).

Materials and methods

Selection of cancer mutations from TCGA

The TCGA was interrogated for the selection of cancer mutations. The top 100 most recurrent somatic mutations across all solid tumors in the TCGA database were selected for the analysis. Collectively, they are identified and reported in 51.8% of all cancer cases at the TCGA database. Each mutation was assessed to confirm that they are identified in more than 5% or shared by more than 5% of cancer cases.

Prediction of mutated neoantigens (NeoAgs)

Each of the wild-type (wt) proteins were downloaded from the UniProt database (https://www.uniprot.org). The amino acid sequences were manually modified, introducing the described mutation (substitution or insertion/deletion). The paired wt and mutated sequences from each protein were analyzed using the NetMHCpan4.1 algorithm (https://services.healthtech.dtu.dk/service.php?NetMHCpan-4.1) to predict the best epitopes. All the newly derived sequences from the premature stop codons were included in the analysis, even if very short. Peptides of 8–12 amino acidic residues with binding level defined as "SB", "WB" to the 50 most frequent HLA-A, B and C alleles were selected. For subsequent analysis, epitopes with an affinity value < 100 nM

were arbitrarily defined as strong binders (SB); < 200 nM as weak binders (WB) and > 200 nM as poor binders (PB). Differential agretopicity index (DAI) was used to evaluate the strength of NeoAgs derived from mutations and only epitopes with DAI > 10 (Affinity nonmutated/Affinity mutated) were selected.

Evaluation of novel identified NeoAgs

The mutated NeoAgs identified as SB with a DAI > 10 were submitted to the Immune Epitope Database & Tools (www.iedb.org) and literature search to verify whether the predicted epitopes have been already described and validated in literature. The setting parameters were for exact sequence matching search.

Statistical analysis

The statistical analyses were performed with GraphPad software (Version 6.01). Comparison between individual data points were performed with the two-sided Student's t-test and ANOVA, as appropriate. Normally distributed data were represented as mean ± S.E.M. Two-way ANOVA and Bonferroni post-hoc analysis were used to examine the significance of differences among groups. All P values were two-tailed and considered significant if less than 0.05.

Results

Strategy for prediction of mutated antigens from shared mutations

Cancer mutations from the TCGA were selected if identified in more than 5% of patients affected by a specific cancer or shared by more than 5% of cancers. The most frequently identified 100 mutations have such characteristics. For each protein carrying such mutations, the corresponding wild-type sequence was manually curated to introduce either the missense point mutations from single nucleotide variations (SNV) or the new sequence from the alternative open reading frame resulting from InDel mutations.

In order to predict mutated antigens derived from SNVs, a 23mer peptide was extracted for each protein, centered on the mutated residue (from −11 to + 11), and overlapping 8–12aa peptides were designed with the mutated residue at each of the positions.

Alternatively, to predict mutated antigens from the InDel mutations, a peptide from each protein was based on a sequence starting at position −11 from the mutated aminoacid residue and including the entire new downstream sequence derived from the alternative open reading frame. The length of the latter mutated peptides ranged from 4 to 62 aa, according to the position of the newly generated stop codon along the shifted reading frame (Suppl. Figure 1).

For both types of mutations, wt and mutated peptides from each protein were subjected to the same prediction analysis, in order to assess the affinity to the selected 50 HLA-A, B and C alleles.

Selection of the HLA alleles

The NetMHCpan 4.1 algorithm was interrogated to predict antigens in the mutated peptides associated to the most prevalent MHC-I HLA alleles, with a global coverage in all Continents. To such aim, 50 alleles, 10 HLA-A, 27 HLA-B and 13 HLA-C alleles were selected. Among these, the A*02:01 and 24:02 are the most prevalent across the Continents; the A*11:01 is highly prevalent in South and South-East Asia as well as in Oceania; the A*31:01 is specifically represented only in South America; the A*33:03 is specifically represented only across Asia; the A*34:01 is specifically represented only in Oceania (Fig. 1A). The most prevalent B alleles across the Continents were selected, including those represented only in specific Continents, such as the B*51:01 which is the most frequent in South Asia and highly frequent in South-East Asia; the B*54:01 and the B*58:01 which are highly frequent or the most frequent in South-East Asia, respectively (Fig. 1B). Similarly, the most prevalent C alleles across the Continents were selected, including the C*04:03 which is the most frequent in South America and highly frequent in Oceania; the C*08:01 which is specifically frequent in North-East and South-East Asia as well as in Oceania; and the C*15:02 which is specific to South-Asia and Oceania (Fig. 1C).

Prediction of mutated antigens associated with HLA-A haplotypes

Predicted mutated antigens (mut-Ags) associated with HLA-A haplotypes are identified only in 11 and 7 proteins, characterized by SNVs (missense mutations) and InDel (frameshift mutations), respectively. Considering the total number from both SNV and InDel mutations, 78% are scored poor binders (> 200 nM affinity—PB), 10.5% are scored weak (100–200 nM—WB) and 11.7% strong (< 100 nM—SB) binders, on average (Suppl. Table 1A). Moreover, the majority of the strongest binder antigens (75.8%) are predicted to bind more than a single haplotype (shared) (Suppl. Table 1).

Regarding the SNVs, the number and the score of the predicted mut-Ags greatly varied among the 11 proteins. In particular, the largest number of WB and SB SNV mut-Ags are predicted in the five SNVs identified in KRAS (8 WBs and 3 SBs), however the four SNVs identified in PIK3CA provided the largest number of SBs (nr. 6). Overall, the individual SNVs generating the largest number of SBs (nr. 3) are the $PIK3CA_{H1047R}$ and $TP53_{R175H}$. On the contrary, the individual SNVs identified in PTEN,

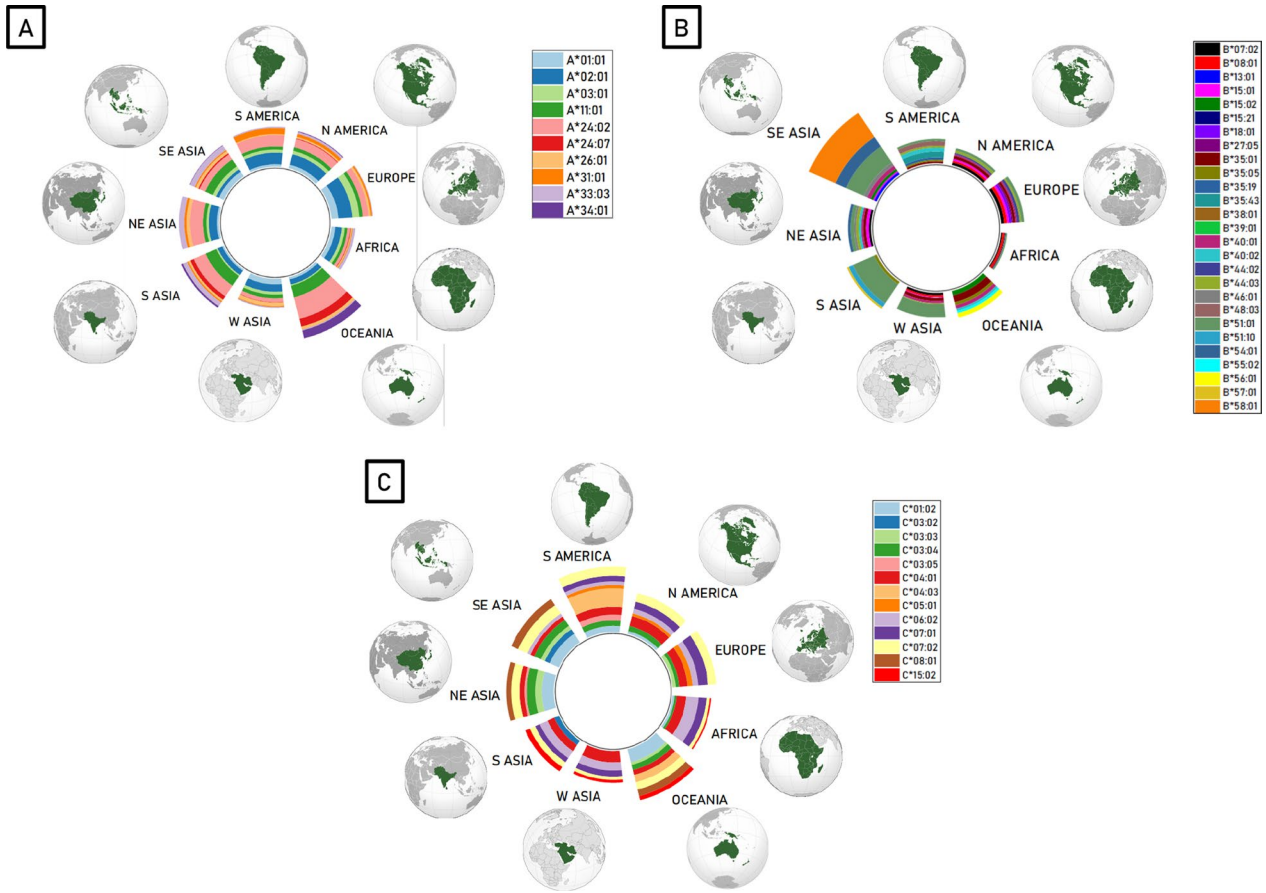

Fig. 1 HLA haplotypes' prevalence. The prevalence of the most common HLA-I haplotypes in sub continental regions for HLA-A (**A**), HLA-B (**B**) and HLA-C (**C**) alleles

IDH1, FGFR3 and the $KRAS_{G13D}$ do not generate any WB nor SB (Fig. 2A and Suppl. Table 1A).

Similar to SNVs, the number and the score of the predicted mut-Ags derive from the InDels (InDel mut-Ags) greatly varied among the 7 proteins. In particular,

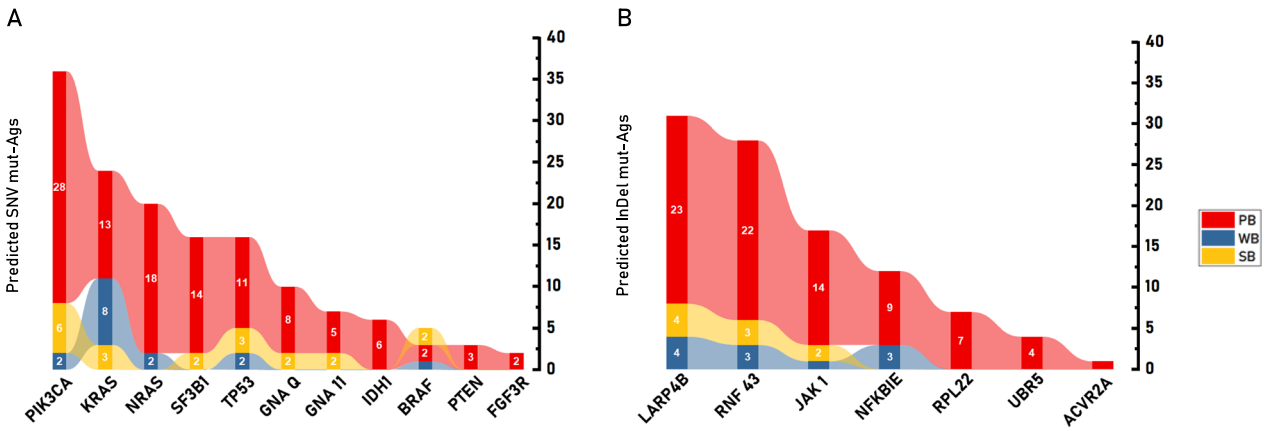

Fig. 2 Predicted shared mut-Ags for HLA-A. The diagrams show the number of putative mut-Ags derived from SNV (**A**) and InDels (**B**). The antigens are grouped by the affinity to the HLA molecules (*PB* poor binders, *WB* weak binders, *SB* strong binders)

the largest number of WB and SB InDel mut-Ags are predicted in the LARP4B$_{T163Hfs_47}$ (4 WBs and 4 SBs), of which 2 WBs and only 1 SB are shared. In parallel, the same LARP4B InDel mutation generates also the largest number of PBs (23 in total). Interestingly, the RNF43$_{G659Vfs_41}$ and the JAK1$_{K860Nfs_16}$ generate less WBs and SBs (6 and 3, respectively) but more shared SBs (2 each).On the contrary, the InDels identified in RPL22, UBR5 and ACVR2A do not generate any top scoring neoantigen (Fig. 2B and Suppl. Table 1B).

HLA-A haplotypes presenting the predicted mutated neoantigens

The stratification of the predicted mut-Ags based on the HLA-A haplotypes, showed a significant variability for both SNVs and InDel mutations.

Considering the SNV mut-Ags, the largest number are predicted to be associated with the A*34:01 (nr. 27), the other haplotypes are associated with 14 mut-Ags, on average, except the A*02:01 which is associated to only 3 neoantigens. However, selecting only those with the highest predicted affinity (100–200 nM and < 100 nM) (nr. 35), none of these are associated with the A*34:01. Indeed, these are predicted in association only with A*01:01 (nr. 2), A*03:01 (nr. 6), A*11:01 (nr. 12), A*31:01 (nr. 7) and A*33:03 (nr. 8). Furthermore, taking into consideration the top scoring prediction (< 100 nM and SBs) (nr. 15) the associated haplotypes are A*03:01 (nr. 3), A*11:01 (nr. 3), A*31:01 (nr. 3) and A*33:03 (nr. 6) (Fig. 3A; Suppl. Figure 2).

Considering the InDel mut-Ags, the largest number are predicted to be associated with the A*31:01 (nr. 21), the other haplotypes are associated with 9 mut-Ags, on average. In this case, even selecting those with the highest predicted affinity (100–200 nM and < 100 nM) (nr. 20), the A*31:01 remains the haplotype associated with the largest number of mut-Ags (nr. 9). On the contrary, the A*01:01, A*02:01, A*26:01 and A*34:01 do not show any predicted mut-Ags. Furthermore, taking into consideration the top scoring prediction (< 100 nM and SBs) (nr. 4) the associated haplotypes are A*03:01 (nr. 1), A*24:02 (nr. 1) and A*31:01 (nr. 2) (Fig. 3B; Suppl. Figure 3).

Affinity map of individual predicted mutated mut-Ags to HLA-A haplotypes

The affinity to HLA-A haplotypes of each predicted SNV mut-Ags and InDel mut-Ags is shown, including those with single and multiple affinities (Tables 1 and 2).

Considering the SNV mut-Ags, the average affinity to the individual haplotypes is 3502.3 nM, ranging from 8513.05 nM (A*01:01) to 330.19 nM (A*31:01). In detail, the average affinity to A*03:01, 11:01, 31:01 and 33:03 is significantly higher than to the remaining haplotypes ($p < 0.01$) and only mut-Ags associated with these four haplotypes are SBs. Indeed 4 mut-Ags for each of the A*03:01, 11:01, 31:01 and 8 for A*33:03 are predicted.

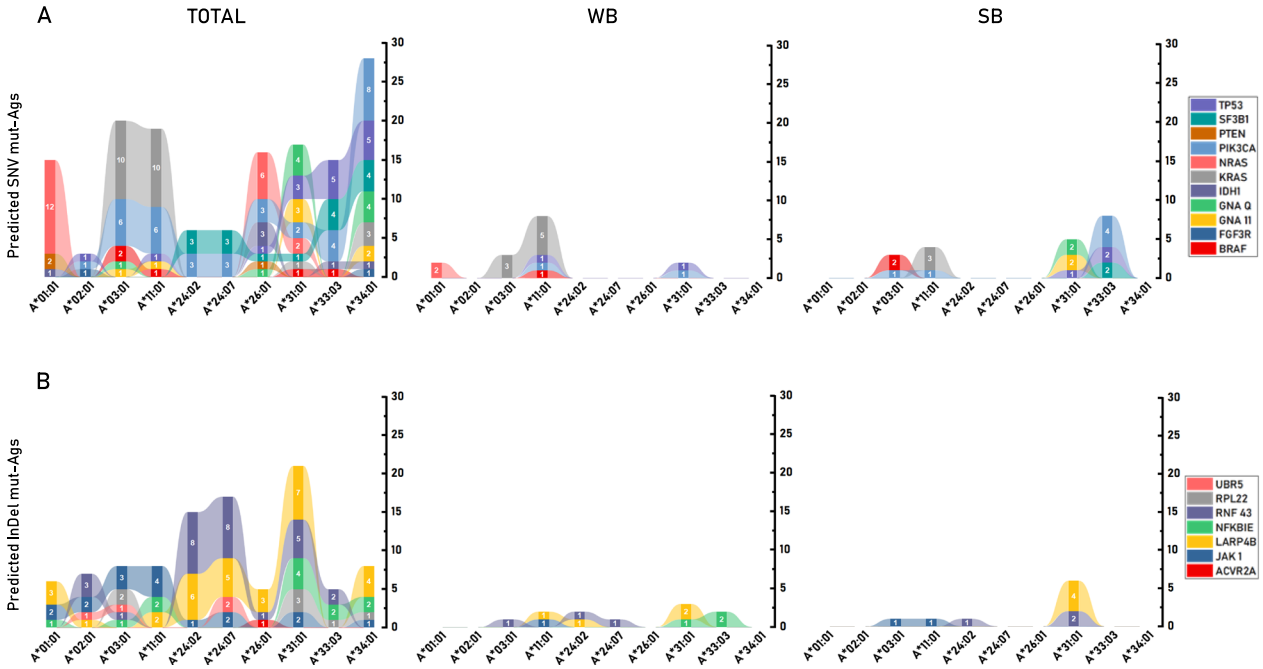

Fig. 3 Stratification of shared mut-Ags based on HLA-A haplotypes. The plots indicate the number of SNV mut-Ags (**A**) and InDel mut-Ags (**B**) for each of the selected HLA-A haplotypes, grouped by the level of binding (WB; SB)

Table 1 Affinity map of SNV mut-Ags predicted to bind HLA-A haplotypes

PROTEIN	PEPTIDE	HLA-A									
		01:01	02:01	03:01	11:01	24:02	24:07	26:01	31:01	33:03	34:01
BRAF	DFGLATEKSR									440	
	FGLATMKSR								501		
	KIGDFGLATEK			68							
	KIGDFGLATMK			48	108						
FGF3R	DVLERCPHR										2572
	YTLDVLERC		1748								
GNA 11	DVGGLRSER										1026
	IFRMVDVGGLR								95		
	VDVGGLRSER								417		
	RMVDVGGLR			475	1006				24		966
GNA Q	DVGGPRSERR										3198
	IFRMVDVGGPR			85							
	VDVGGPRSER			532							
	DVGGPRSER							14020			699
	MVDVGGPRSER								633		3100
	RMVDVGGPR			453					20		582
IDH1	IIGHHAYGDQY	11705									
	IIIGHHAY							11275			
	PIIIGHHAY							8317			
	WVKPIIIGH							7167		2879	1659
KRAS	EYKLVVVGAR								347		
	VVGACGVGK			215	73						
	VVGADGVGK			1105	172						
	VVGAGDVGK			1911	314						
	VVGARGVGK			135	107						
	VVGAVGVGK			134	39						3397
	VVVGACGVGK			266	122						
	VVVGADGVGK			590	194						
	VVVGAGDVGK			592	213						
	VVVGARGVGK			209	162						1566
NRAS	AGKEEYSAMR								427		
	AGREEYSAMR								365		
	DTAGKEEYSAM							673			
	DTAGREEYSAM							563			
	ILDTAGKEEY	157									
	ILDTAGKEEYS	8910									
	ILDTAGREEY	143									
	ILDTAGREEYS	8321									
	KEEYSAMRDQY	20082									
	LDTAGKEEY	15847									
	LDTAGREEY	12404									

Table 1 (continued)

PROTEIN	PEPTIDE	HLA-A									
		01:01	02:01	03:01	11:01	24:02	24:07	26:01	31:01	33:03	34:01
	REEYSAMRDQY	17860									
	VVVGAVGVGK			186	69						903
	DILDTAGKEEY	3622						3493			
	DILDTAGREEY	3401						3181			
	DTAGKEEY	12880						9253			
	DTAGREEY	10765						5967			
PIK3CA	AMESEITKI		2108								
	DARHGGWTTK										1830
	EITKITEQEK										2206
	EQAMESEITK										2578
	ETRQLCDLRLF							4680			
	YFMKQMNDAR									51	
	AISTRDPLSK			54	99						
	EFFDETRQL					12070	11397	11697			4878
	ETRQLCDLR									47	1192
	EYFMKQMNDAR								533	64	
	FMKQMNDAR								138	68	
	ISTRDPLSK			1444	333						
	ITKITEQEK			2661	913						
	KAISTRDPLSK			470	256						
	QAMESEITK			2668	137						2924
	REEFFDETRQL					12042	9650				
	RQLCDLRLF					496	1371				
	STRDPLSK			2154	1880			5176			1558
PTEN	GTGVMICAY	1069									
	QTGVMICAY	530						2675			
SF3B1	DEYVHNTTAR									432	
	EYVCNTTAR									62	
	YVHNTTARA										584
	EYVCNTTARAF					2056	4460				
	EYVHNTTAR								901	58	1864
	EYVHNTTARAF					3259	7124				
	VHNTTARAF					10698	10902				
	YVHNTTAR									893	1714
	YVHNTTARAF							362			300
TP53	EVVRHCPHH							3022			2738
	EVVRHCPHHER									88	1224
	SCMGGMNQR								409	312	2597
	SSCMGGMNQR				171				114	268	1400
	VVRHCPHHER								31	79	2398
	HMTEVVRHC		4543								
	NSSCMGGMNQR									405	

Table 2 Affinity map of InDel mut-Ags predicted to bind HLA-A haplotypes

PROTEIN	PEPTIDE	HLA-A									
		01:01	02:01	03:01	11:01	24:02	24:07	26:01	31:01	33:03	34:01
ACVR2A	EVVVHKKRGLF							2706			
JAK1	VSEKNQQLKW	13147									
	IVSEKNQQL		3228				20573				8403
	IVSEKNQQLK			525	240						
	QLKWTPHILK			32	89				378		
	QQLKWTPHI		183			2040	1718				
	QQLKWTPHILK			692	258				1238		
	VSEKNQQLK	6913			741						
LARP4B	HWNSAYLGR								176		
	KHWNSAYLGR								110		
	LVTCILYHR								17		
	RWLTSTTSR								33		
	SAYLGRTLL										2156
	SQRWLTSTTSR								99		
	TLLVTCILY	5714									
	VTCILYHR								216		
	VTCILYHRW					3275					
	YLGRTLLV		257								
	AYLGRTLL					5867	5502				
	AYLGRTLLV					157	209				
	EVLKKHWNSAY							257			1310
	HWNSAYLGRTL					2482	2963				
	LYHRWIVTSM					651	630				
	RTLLVTCILY	912			138						
	SAYLGRTLLV					653	647				
	STTSRSSAL							4458			1308
	TTSRSSALM	3026						1023			517
	VTSMCQSQR				436				77		
NFKBIE	ALTSSPRTETR								789		
	QQLEALTSSPR								780		
	RTETRWSTW	9726									
	LTSSPRTETR			3120	1081				106	141	1643
	TSSPRTETR				1708				267	159	1561
RNF 43	ALGPRMQLC		3868								
	FPITPPVW					10758					
	GVPPSPPLA		6908								
	ITPPVWHI						6075				
	MQLCTQLAR								44		
	PITPPVWHI						11294				

Table 2 (continued)

PROTEIN	PEPTIDE	HLA-A									
		01:01	02:01	03:01	11:01	24:02	24:07	26:01	31:01	33:03	34:01
	PVWHILGPQR									489	
	ARFFPITPPVW					1874	3021				
	FFPITPPVW					176	452				
	FFPITPPVWHI					1403	1611				
	ITPPVWHIL					405	569	10968			
	RFFPITPPV		261			1041	808		412		
	RFFPITPPVW					86	192				
	RFFPITPPVWH					3701			566		
	RMQLCTQLAR			130					11		
	VWHILGPQR								302	604	
RPL22	KKLVVKGGKK			384							
	KKLVVKGGKKR								2896		
	KLVVKGGKKR								964		
	VVKGGKKRSK			4730							
	LVVKGGKKR								1894	1415	2076
UBR5	KLRVQNQGH			2905							
	KLRVQNQGHLL		4118								
	RVQNQGHLL						4526				
	VQNQGHLLM						7682				

Looking at the same results from the protein perspective, the average affinity of mut-Ags to all haplotypes is 2501.76 nM, ranging from 7166.87 (IDH1) to 233.07 (BRAF) (Table 1 and Fig. 4A and B). None of the proteins showed mut-Ags with an overall average affinity significantly different from all other proteins. However, SBs are predicted for all the proteins with the exception of FGF3R, IDH1, PTEN and NRAS. In particular, 2 mut-Ag for each of BRAF, GNA11, GNAQ, SF3B1, 3 for KRAS and TP53 as well as 6 for PIK3CA (Fig. 2 and Suppl. Table 1A).

Considering the InDel mut-Ags, the average affinity to the individual haplotypes is 2510.3 nM, ranging from 6573.08 nM (A*01:01) to 541.58 nM (A*31:01). In detail, also for these antigens the average affinity to A*03:01, 11:01, 31:01 and 33:03 is higher than to the remaining haplotypes, although it did not reach the statistical significance. SBs are predicted only for these haplotypes, namely 1 each of the A*03:01, 11:01, 24:02 and 6 for A*31:01. Looking at the same results from the protein perspective, the average affinity of mut-Ags to all haplotypes is 2676.42 nM, ranging from 4807.62 (UBR5) to 1460.54 (LARP4B) (Table 2 and Fig. 4C and D). None of the proteins showed mut-Ags with an overall average affinity significantly different from all other proteins. However, SBs are predicted only for the proteins JAK1 (nr. 2), LARP4B (nr. 4) and RNF 43 (Nr. 3). None is predicted for the ACVR2A, NFKBIE, RPL22 and UBR5 proteins (Fig. 2B and Suppl. Table 1B).

Targetable predicted mutated neoantigens linked to HLA-A haplotypes

In order to verify whether the predicted SNV and InDel mut-Ags are real neoantigens and may be actual targets for shared off-the-shelf cancer immunotherapy strategies, the differential agretopicity index (DAI) for each WB and SB antigens was calculated. In particular, a ratio > 10 of MHC affinity of the mutant peptide:the nonmutated counterpart is considered as meaningful [16–20].

According to such analysis, only 5 predicted SNV-NeoAgs and 15 InDel-NeoAgs are targetable because they show either a DAI > 10 or they are de novo protein sequences which do not have a match in the wt corresponding sequence (Table 3). Regarding the SNV-NeoAgs, 4 of the 5 (80%) are SB; 3 are associated with the HLA-A*33:03 ($PIK3CA_{H1047R}$ YFMKQMNDAR, EYFMKQMNDAR and FMKQMNDAR). The $PIK3CA_{E542K}$ AISTRDPLSK peptide binds to both A*03:01 and 11:01 (Table 3).

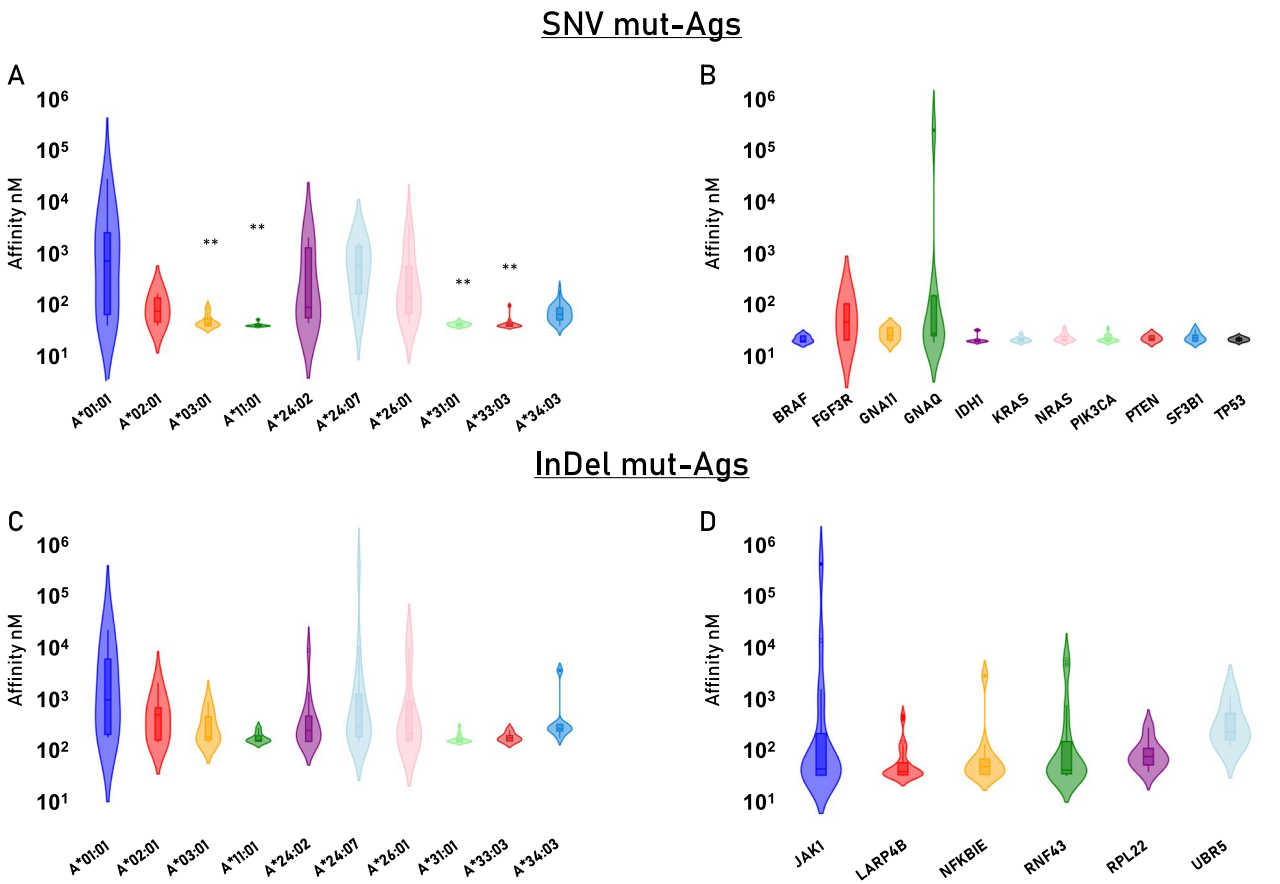

Fig. 4 Affinity of shared SNV and InDel mut-Ags to HLA-A haplotypes. The affinity to HLA-A haplotypes of all predicted mut-Ags (PB, WB and SB) are plotted. Values are shown according to each haplotype (**A** and **C**) or mutated protein from which the mut-Ags are derived (**B** and **D**). *nM* nanomolarity

Concerning the InDel-NeoAgs, 8 of 15 (53.3%) are SB and are associated to the HLA-A*03:01, 11:01, 24:02 and 31:01. While the first three are associated with one SB each, the 31:01 is associated with 6 of them (75%). 50% of them (4 out of 8) are derived from the LARP4B$_{T163Hfs}$ protein and are all associated with the B*31:01 haplotype. Finally, the majority of the targetable SB InDel-Neo-Ags are associated with a single haplotype ad only the JAK1$_{K860Nfs}$ QLKWTPHILK peptide is associated with two haplotypes (HLA-A* 03:01 and 11:01) (Table 3).

Prediction of mut-Ags associated with HLA-B haplotypes

Predicted mut-Ags associated with HLA-B haplotypes are identified in the same 11 proteins, characterized by SNVs, and in 7 proteins characterized by InDel. Considering the total number of both types of mut-Ags, 86.1% are scored PB, 4.7% are scored WBs and 9.1% SBs, on average. Moreover, the majority of WB and SBs (61.9%) are predicted to bind more than a single haplotype (Suppl. Table 2A and B).

Also for the HLA-B, The number and the score of the predicted SNV mut-Ags greatly varied among the 11 proteins. The largest number of WB and SB SNV mut-Ags are predicted in the 4 SNVs identified in PIK3CA (9 WBs and 3 SBs); however the two SNVs identified in SF3B1 provided the largest number of SBs (nr. 5). Overall, the SNVs generating the largest number of such mut-Ags (nr. 3) are the SF3B1$_{R625C}$, GNA11$_{Q209L}$ and FGFR3$_{S249C}$. On the contrary, the SNVs identified in TP53, NRAS, KRAS, IDH1 did not generate any SB (Fig. 5A, Suppl. Table 2A).

Similar to SNVs, the number and the score of the predicted InDel mut-Ags greatly varied among the 7 proteins. In particular, the largest number of WB and SB InDel mut-Ags are predicted in the LARP4B$_{T163Hfs_47}$ (4 WBs and 13 SBs). On the contrary, the InDels identified in RPL22 and ACVR2A did not generate any WB and SBs (Fig. 5B, Suppl. Table 2B).

Table 3 WB and SB targetable Neo-Ags binding to HLA-A haplotypes

	PROTEIN	PEPTIDE	HLA-A				DAI	
			03:01	11:01	31:01	33:03		
SNV - NeoAgs	$PIK3CA_{H1047R}$	YFMKQMNDAR				51	105	
		EYFMKQMNDAR				64	130	
		FMKQMNDAR			137	68	88	170
	$PIK3CA_{E542K}$	AISTRDPLSK	53	99			467	281
	$PIK3CA_{E545K}$	QAMESEITK		136			232	

	PROTEIN	PEPTIDE	HLA-A							DAI	
			02:01	03:01	11:01	24:02	24:07	31:01	33:03		
InDel - NeoAgs	$JAK1_{K860Nfs}$	QLKWTPHILK		31	89			378,3		n/a	n/a
		QQLKWTPHI	182			206	1717,9			n/a	n/a
	$LARP4B_{T163Hfs}$	HWNSAYLGR						176		161	n/a
		AYLGRTLLV				156				n/a	n/a
		LVTCILYHR						17		n/a	n/a
		RWLTSTTSR						32		n/a	n/a
		RTLLVTCILY			138					n/a	n/a
		SQRWLTSTTSR						98		n/a	n/a
		VTSMCQSQR						76		n/a	n/a
	$NFKBIE_{Y254Sfs}$	LTSSPRTETR						105	142	342	272
		TSSPRTETR							159	255	n/a
	RNF 43 $_{G659Vfs}$	MQLCTQLAR						43		n/a	n/a
		FFPITPPVW				175				n/a	n/a
		RFFPITPPVW				86	191			n/a	n/a
		RMQLCTQLAR		129				10		n/a	n/a

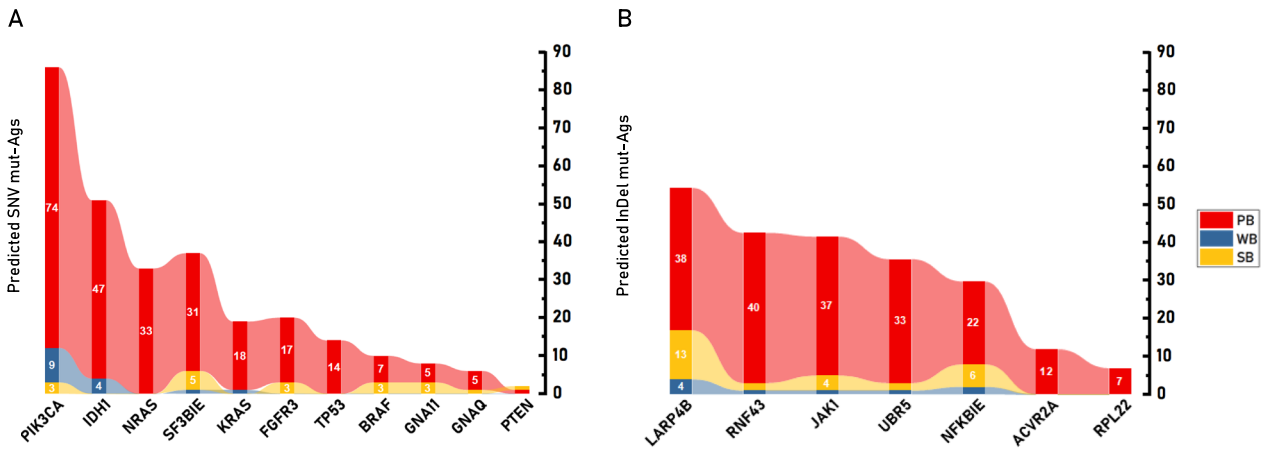

Fig. 5 Predicted shared mut-Ags for HLA-B. The diagrams show the number of putative mut-Ags derived from SNV (A) and InDels (B). The antigens are grouped by the affinity to the HLA molecules (PB = poor binders; WB = weak binders; SB = strong binders)

HLA-B haplotypes presenting the predicted shared mut-Ags

The stratification of the predicted mutated mut-Ags based on the HLA-B haplotypes, showed a significant variability for both SNVs and InDel mutations.

Considering the mutated mut-Ags derived from SNVs, the largest number are predicted to be associated with the B*18:01 (nr. 18), the other haplotypes are associated with 10.3 mut-Ags, on average, except the B*13:01 and B*51:01 which are associated to only 3 mut-Ags. However, selecting only WB and SBs (nr. 34), only 2 are associated with the B*18:01 and they are almost evenly distributed among several B haplotypes. Furthermore, taking into consideration the SBs only (nr. 19), the B*27:05 haplotype is associated with largest number (nr. 4) (Fig. 6A and Suppl. Figure 4).

Considering the InDel mut-Ags, the largest number are predicted to be associated with the B*58:01 (nr. 18), the other haplotypes are associated with 7 mut-Ags, on average, except the B*15:02, B*35:01 and B*35:05 which are associated to only 1 neoantigen. In this case, even selecting the WB and SBs (nr. 32), the B*58:01 remains the haplotype associated with the largest number of mut-Ags (nr. 15). Such finding is further confirmed also taking into consideration the SBs only, indeed 11/23 are associated with the B*58:01 haplotypes. On the contrary, most of the B haplotypes (20/27) do not show any predicted SB (Fig. 6B and Suppl. Figure 5).

Affinity map of individual predicted mutated mut-Ags to HLA-B haplotypes

The affinity to HLA-B haplotypes of each predicted mut-Ags from SNVs and InDels is shown, including those with single and multiple affinities (Tables 4 and 5).

Considering the SNV mut-Ags, the average affinity to the individual haplotypes is 4890.5 nM, ranging from 15780.5 nM (B*46:01) to 618.5 nM (B*15:01). In detail, the average affinity to B*15:01 and 27:05 is significantly higher than to most of the remaining haplotypes ($p < 0.01$) and most of the SBs are found associated with these two haplotypes (4 each out of 16 total) (Table 4). Looking at the same results from the protein perspective, the average affinity of mut-Ags to all haplotypes is 4131.3 nM, ranging from 7088.05 (NRAS) to 935.67 (PTEN) (Table 4 and Fig. 7A and B). None of the proteins showed mut-Ags with an overall average affinity significantly different from all other proteins. However, SBs are predicted for all the proteins with the exception of IDH1, KRAS, NRAS and TP53. In particular, 1 mut-Ag for GNAQ and FGF3R, 2 for PTEN and GNA11, 3 each for BRAF and, PIK3CA, 4 for SF3B1 (Fig. 5A and Supplementary Table 2A).

Considering the neoantigens from InDels, the average affinity to the individual haplotypes is 6027.8 nM, ranging from 18,499.7 nM (B*46:01) to 825.9 nM (B*15:01). In detail, also for the InDel mut-Ags the average affinity to B*15:01 and 58:01 is higher than to the remaining haplotypes, although it did not reach the statistical

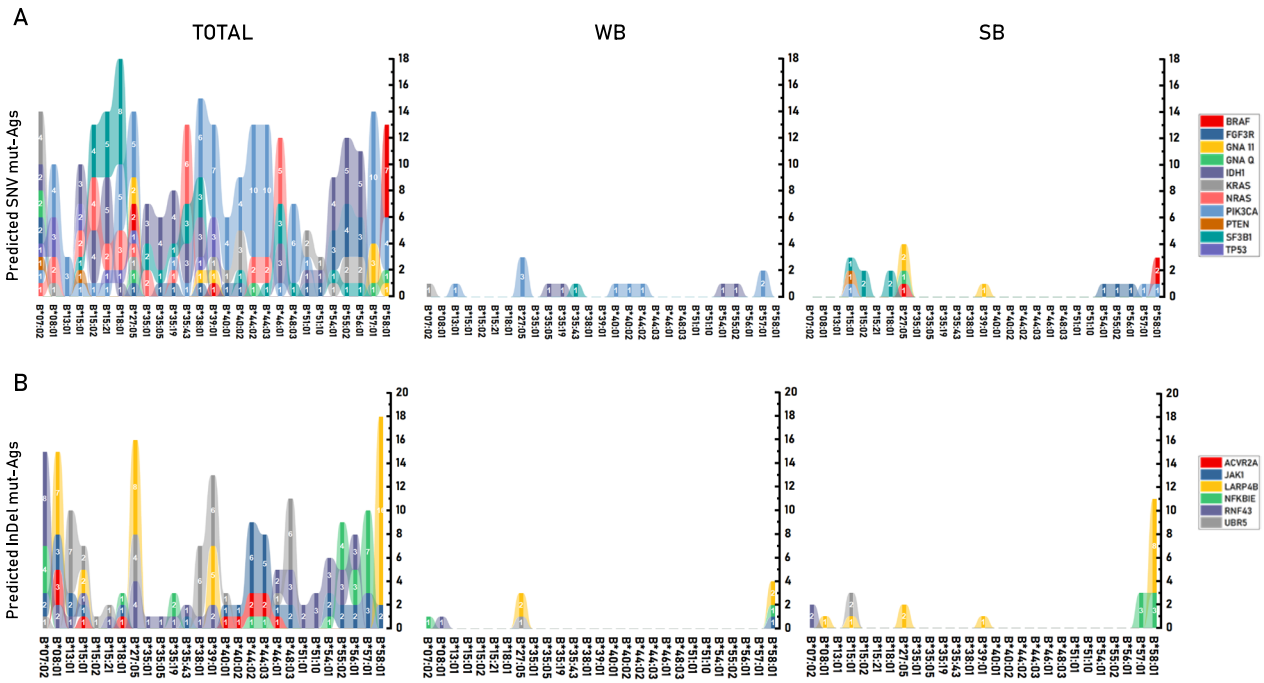

Fig. 6 Stratification of shared mut-Ags based on HLA-B haplotypes. The plots indicate the number of SNV mut-Ags (**A**) and InDel mut-Ags (**B**) for each of the selected HLA-B haplotypes, grouped by the level of binding (WB; SB)

Table 4 Affinity map of SNV mut-Ags predicted to bind HLA-B haplotypes

PROTEIN	PEPTIDE	HLA-B																										
		07:02	08:01	13:01	15:01	15:02	15:21	18:01	27:05	35:01	35:05	35:19	35:43	38:01	39:01	40:01	40:02	44:02	44:03	46:01	48:03	51:01	51:10	54:01	55:02	56:01	57:01	58:01
BRAF	EKSRWSGSHQF								446						11971													
	MKSRWSGSHQF								42																			758
	FGLATEKSRW																											92
	FGLATMKSRW																											67
	GLATEKSRW																											2205
	GLATMKSRW																											1034
	LATEKSRW																											1460
	LATMKSRW																											768
FGFR3	CPHRPILQA	561									2334	2797										4215	2400	19	41	88		
	ERCPHRPIL								2721					10175														
	ERCPHRPILQA																							5088	4461	8264		
	LERCPHRPIL															803	266											
	RCPHRPILQA																							1014	436	1162		
	CPHRPILQAG																								1673			
	CPHRPILQAGL	945																										
GNA11	FRMVDVGGL								99					1399	92													
	GGLRSERRKW																										1843	2989
	FRMVDVGGLR								65																			
	GLRSERRKW																										5248	
	VGGLRSERRKW																										5241	
GNAQ	GPRSERRKW	3661																22075										
	VGGPRSERRKW																										5069	6284
	FRMVDVGGPR								83																			
	GPRSERRKWI	515																										
IDH1	GHHAYGDQY						7909							8064														
	HHAYGDQY						11495							13160														
	IIIGHHAY				759	412	5976			2832	2452	2482	3152							12437								
	KPIIIGHHA	758									6037	5729										13868	7468	164	112	249		
	KPIIIGHHAY	1716			276	633	8807	6180		290	197	138	1308							19193					3694	8097		
	PIIIGHHA																							3875	5647	10909		
	PIIIGHHAY				1529	555	7329			3008	3435	2607	5454							22049								
	VKPIIIGHHA																							224	436	1007		
	WVKPIIIGHHA																							622	1062	2436		
	ARGVGKSAL								2242						2617													
	HHAYGDQYRA													7750														
	VKPIIIGHHAY					2781																						
KRAS	DVGKSALTI																					6576	10000					
	GARGVGKSA	2550																							6829	11572		
	GAVGVGKSA																							8152	8710	12401		
	TEYKLVVVGAV															2553	504											
	ADGVGKSAL																7919											
	AGDVGKSAL	6717																										
	AVGVGKSAL	855																										
	DGVGKSAL		4851																									
	DGVGKSALTI																					6154						
	GARGVGKSAL	166																										
	GDVGKSAL																8943											
NRAS	AGKEEYSAM		3564										6830							14514								
	AGREEYSAM	1381	3402										4959							12157								
	DILDTAGKEEY					3510				8933																		
	DILDTAGREEY					3143				8383		9210																
	ILDTAGKEEY				1910	2282	13224													26102								
	ILDTAGREEY				833	1299	9917													22535								
	KEEYSAMRDQY							3358										1923	1532									
	REEYSAMRDQY							3103										1540	1229									
	TAGKEEYSAM												1830							15832								
	DTAGKEEY												18562															
	DTAGREEY												13579															
	GREEYSAMR								1290																			
	LDTAGREEY							10698																				
	TAGREEYSAM												1341															
PIK3CA	ARHGGWTTKM								149					7486	5241													
	EEFFDETRQL							2986								897	1118	1016	725		20448							
	EEFFDETRQLC																	12614	10862									
	EFFDETRQL		14677				11462						13144	19203	12420													
	FFDETRQL		4359												10586													
	KITEQEKDFLW																	648	1235								175	74
	KQEKDFLW			4256																							4195	
	KQMNDARHGGW																	2991	4749								2580	1181
	LSEITKQEKDF																	10190	13479									
	QMNDARHGGW																										2694	1677
	REEFFDETRQL			4434	143											157	143	5285	5733		4944							

significance. In particular, 12 out 22 SBs are predicted for the 58:01 haplotype (Tables 5). Looking at the same results from the protein perspective, the average affinity of mut-Ags to all haplotypes is 4776.42 nM, ranging from 6901.62 (RNF43) to 2833.9 (NFKBIE) (Table 5 and Fig. 7C and D). None of the proteins showed mut-Ags with an overall average affinity significantly different from all other proteins. InDel-NeoAgs SBs are predicted for all the proteins, except ACVR2A and RPL22,

Table 4 (continued)

PROTEIN	PEPTIDE	HLA-B																										
		07:02	08:01	13:01	15:01	15:02	15:21	18:01	27:05	35:01	35:05	35:19	35:43	38:01	39:01	40:01	40:02	44:02	44:03	46:01	48:03	51:01	51:10	54:01	55:02	56:01	57:01	58:01
	RHGGWTTKM													837	951						10425							
	SEITKQEKDF							11887									3608	184	295									
	SEITKQEKDFL																1686	4565	6500		18697							
	STRDPLSKI	3057																		19444		16491					5486	
	TKQEKDFLW													5161				9320	9678									
	RQLCDLRLF			104	47																1695							
	TRDPLSKI													15058	16051													
	TRQLCDLRL								354					2381	1054													
	ARHGGWTTK								152																			
	DETRQLCDL							1804																				
	DLRLFQPF					2184																						
	DPLSKITEQ											10962																
	EEFFDETRQ							7568																				
	EITKQEKDFLW																										195	
	FFDETRQLC		10590																									
	FFDETRQLCDL		6446																									
	ITKQEKDFLW																										14	
	ITKQEKDFLWS																										475	
	KQEKDFLWSH																		9308									
	LSKITEQEKDF																										2992	
	MESEITKI															1907												
	MESEITKIT															2381												
	MNDARHGGW																											1697
	NDARHGGW																	8784										
	NDARHGGWTTK								2989																			
	RHGGWTTKMDW																										1258	
	RQLCDLRL																				4510							
	TKQEKDFLWSH							3753																				
	TRDPLSKIT														9445													
	TRQLCDLRLF								195																			
PTEN	AGKGGTGVM	2762																										
	GQTGVMICAY				22																							
	GQTGVMICAY				22																							
SF3BIE	DEYVHNTTA							85									1859		9088									
	EYVHNTTARAF						6013							7032														
	VCNTTARAF					3376	12312			5152										19912								
	VHNTTARAF					4340	3751	6665					10413	948							15640							
	YVCNTTARAF				23	67	1479						586							4020								
	YVHNTTARA																							1354	3704	5874		
	YVHNTTARAF					27	502			224	487	203	172	817						1171								
	DEYVCNTT							617																				
	DEYVCNTTA							88																				
	DEYVCNTTAR							2755																				
	DEYVHNTTARA							3166																				
	DEYVHNTT							425																				
TP53	NQRPILTII		3363		649										4893													
	QHMTEVVRH						19552							20993														
	QRPILTIITL	419							277						979													
	GMNQRPIL		2178																									
	NQRPILTI		5672																									
	SQHMTEVVRH				1209																							
	NQRPILTIITL														3381													
	TEVVRHCPH							1323																				
	VRHCPHHER								482																			

and the largest number (nr. 13) is predicted for LARP4B (Fig. 5B and Supplementary Table 2B).

Targetable predicted mutated neoantigens in HLA-B haplotype

The differential agretopicity index (DAI) was calculated also for each WB and SB mut-Ags linked to B haplotypes. According to such analysis, 7 predicted SNV mut-Ags and 30 InDel mut-Ags can be considered as NeoAgs and are targetable because they show either a DAI > 10 or they are de novo protein sequences which do not have a match in the wt corresponding sequence (Table 6). Concerning the SNV-NeoAgs, 4 of the 7 are SB (57.1%) and are associated with HLA-B*15:01 (PIK3CA$_{R88Q}$ RQLCDLRLF and PTEN$_{R130Q}$ GQTGVMICAY), 15:02 (SF3B1$_{R625C}$ YVCNTTARAF), 27:05 and 39:01 (GNA11$_{Q209L}$ FRMVDVGGL) haplotypes. Therefore, only the latter is associated with > 1 haplotype (Table 6).

Regarding the InDel-NeoAgs, 22 of the 30 (73.3%) are SBs and are associated to 9 of 27 HLA-B haplotypes analyzed in the study. Most of such haplotypes are associated with one or two SB InDel-NeoAgs, while the 58:01 is associated with 12 of them (54.5%). Interestingly, most of the SB InDel-NeoAgs (13 out of 27) are derived from the LARP4B$_{T163Hfs}$ protein and 8 of them are associated with the B*58:01 haplotype. Finally, the majority of the targetable SB InDel-NeoAgs are associated with a single haplotype ad only the JAK1$_{K860Nfs}$ SEKNQQLKW and VSEKNQQLKW, NFKBIE$_{Y254Sfs}$ LTSSPRTETRW, RTETRWSTW, TSSPRTETRW peptides are associated with two haplotypes (Table 6).

Table 5 Affinity map of InDel mut-Ags predicted to bind HLA-B haplotypes

PROTEIN	PEPTIDE	HLA-B																										
		07:02	08:01	13:01	15:01	15:02	15:21	18:01	27:05	35:01	35:05	35:19	35:43	38:01	39:01	40:01	40:02	44:02	44:03	46:01	48:03	51:01	51:10	54:01	55:02	56:01	57:01	58:01
ACVR2A	QEVVVHKKR							12831										12995	12517									
	QEVVVHKKRGL															4666	2228	3494	3787									
	VVHKKRGLF		1041		1037															19423								
	VVHKKRGL		641																									
	VVVHKKRGL		5682																									
JAK1	EKNQQLKW																	6705	10299									
	IVSEKNQQL	7714	4131	7898									13413							21969	14105							
	IVSEKNQQLKW																	2716	4423								281	176
	QLKWTPHIL		936		3135																							
	SEKNQQLKW			10394				7560								19680	10805	35	61								13542	11688
	SEKNQQLKWT																	1284	2304									
	SEKNQQLKWTP																	6181	9004									
	TPHILKSA																							1178	982	2509		
	TPHILKSAS	775																						1003	753	1530		
	VSEKNQQLKW																	503	965								67	44
	VSEKNQQLKWT																	15328									3398	3270
	DIVSEKNQQL		11029																									
	NQQLKWTPHIL																				8876							
LARPB4	DPREVLKKH							19646			13008	15155	17341															
	HRWIVTSM								152						1657													
	KHWNSAYL													2508	4393													
	KKHWNSAYL														2210						2076							
	LKKHWNSAY				640	1602	6502																					
	QEDPREVLKKH																	19185	18657									
	REVLKKHW																	933	1888									
	SAYLGRTLL		1228												2235													
	VLKKHWNSAY				32	224														9944								
	YHRWIVTSM		2397												72													
	EDPREVLKKHW																	15562										
	EVLKKHWNSAY					1076																						
	GRTLLVTCI								152																			
	GRTLLVTCIL								70																			
	GRTLLVTCILY								424																			
	HRWIVTSMC								417																			
	ILYHRWIVTSM		799																									
	IVTSMCQSQRW																											22
	LLVTCILYHRW																											286
	LVTCILYHRW																											30
	QRWLTSTTS								741																			
	QRWLTSTTSR								73																			
	QRWLTSTTSRS								1254																			
	REVLKKHWNSA																424											
	RTLLVTCIL																											123
	RTLLVTCILY																											100
	SAYLGRTL		1639																									
	SMCQSQRW																											688
	SRSSALMW																											856
	STTSRSSAL		559																									
	STTSRSSALMW																											36
	TCILYHRW																											1428
	TSMCQSQRW																											10
	TSRSSALMW																											16
	TTSRSSALM																											691
	TTSRSSALMW																											12
	TTSRSSALMWT																											641
	VLKKHWNSA		60																									
	VLKKHWNSAYL		433																									
	VTCILYHRW																											12
	VTSMCQSQRW																											11
NKKBIE	LTSSPRTETRW																										52	25
	RTETRWSTW																										32	24
	SPQQLEALTS																								5312	9943		
	SPQQLEALTSS																								6058	10284		
	SPRTETRWS	725																							5258			
	SPRTETRWST	126																						2317	1139	2555		
	SPRTETRWSTW	1117										10680															969	1651
	SSPRTETRW																										406	175
	TETRWSTW							5175										341	535									
	TSSPRTETRW																										58	17
	TSSPRTETRWS																										2677	1110
	PRTETRWSTW																										807	
	SPQQLEAL											9375																
	SPRTETRW	6076																										
RNF43	HPQRKRRGV	73	161																					9057	5538	8877		
	RGVPPSPPL	769		3739																17350	4037							

Prediction of mutated mut-Ags associated with HLA-C haplotypes

Predicted mut-Ags associated with HLA-C haplotypes are identified in the same 11 mutated proteins characterized by SNVs and in 5 mutated proteins characterized by InDels. Indeed, no mut-Ags are predicted in the $RPL22_{K15Rfs*5}$ and $NFKBIE_{Y254Sfs*13}$ proteins (Fig. 8A, B). Considering the total number of both types of mut-Ags, 91.9% are scored PB, 2.2% are scored WBs and 5.8% SB binders, on average. Moreover, the majority of WB + SB

Table 5 (continued)

PROTEIN	PEPTIDE	HLA-B																										
		07:02	08:01	13:01	15:01	15:02	15:21	18:01	27:05	35:01	35:05	35:19	35:43	38:01	39:01	40:01	40:02	44:02	44:03	46:01	48:03	51:01	51:10	54:01	55:02	56:01	57:01	58:01
	RGVPPSPPLAL	204																					22519					
	RRGVPPSPPL								328						3485						8219							
	VPPSPPLA																					10860	12775	6226	5386	9966		
	VPPSPPLAL	84	9164			7388	18225			1376	1164	1055	3954	20258	2735					28270	18160	3247	4971	12381	7911	8248		
	GVPPSPPLAL	794																										
	KRRGVPPSP								6824																			
	KRRGVPPSPPL								1136																			
	PPSPPLAL	3907																										
	RHPQRKRRGV	299																										
	RRGVPPSPPLA								2690																			
	VPPSPPLALG	2936																										
UBR5	GHLLMILL													7166	6251													
	LRVQNQGHL								882					7444	3287													
	LRVQNQGHLL								190					4661	2831													
	NQGHLLMIL			644										2942	529						1844							
	RVQNQGHLL	788		1606																	8716							
	RVQNQGHLLM			835	73																1858							
	VQNQGHLL			5518										14308	8490						5147							
	VQNQGHLLM			432	38	664	2909							2255	932	2257				14042	682							
	VQNQGHLLMIL			2480																	4065							
	KLRVQNQGHL								1454																			
	LRVQNQGHLLM								605																			
	VQNQGHLLMI			593																								

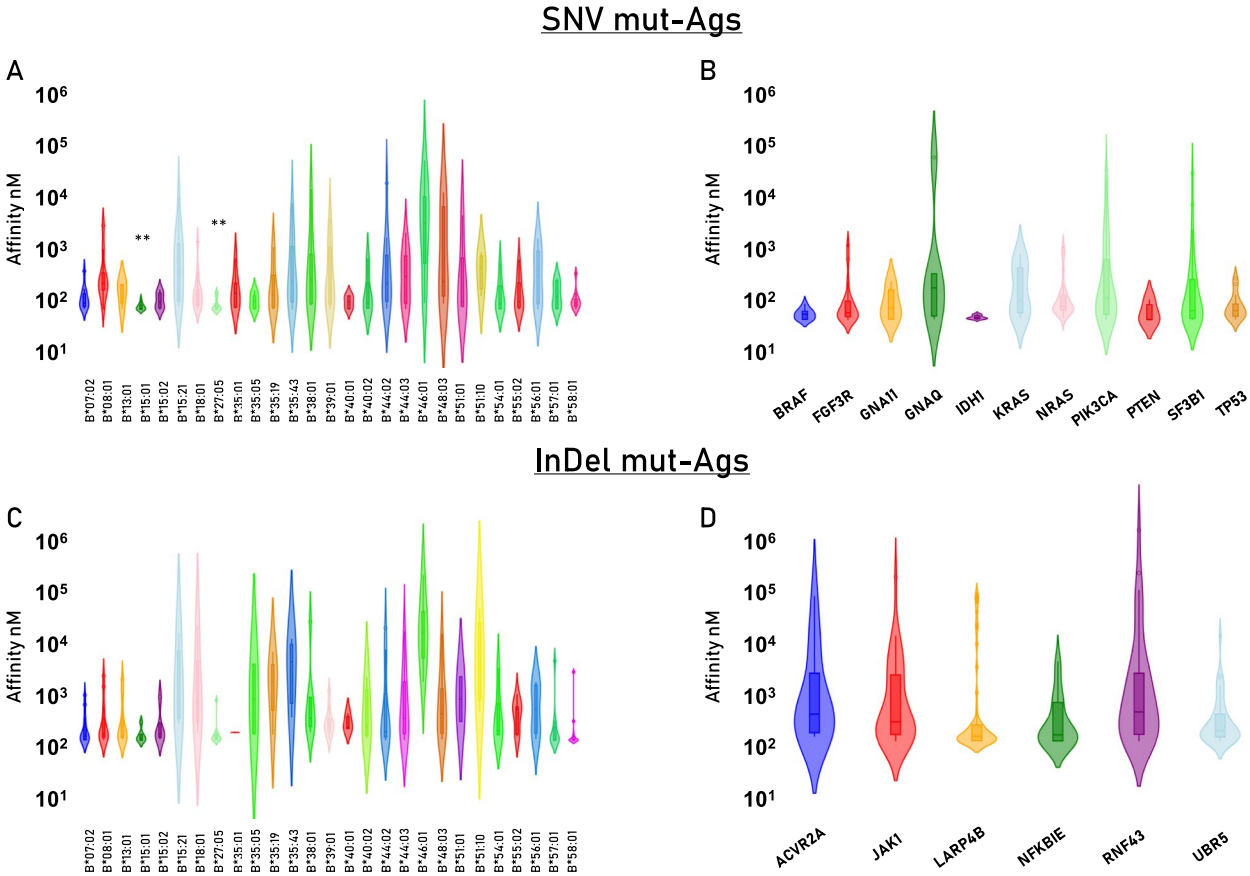

Fig. 7 Affinity of shared SNV and InDel mut-Ags to HLA-B haplotypes. The affinity to HLA-B haplotypes of all predicted mut-Ags (PB, WB and SB) are plotted. Values are shown according to each haplotype (**A** and **C**) or mutated protein from which the mut-Ags are derived (**B** and **D**). *nM* nanomolarity

Table 6 WB and SB targetable Neo-Ags binding to HLA-B haplotypes

	PROTEIN	PEPTIDE	HLA-B						DAI	
			07:02	13:01	15:01	15:02	27:05	39:01		
SNV - NeoAGs	$GNA11_{Q209L}$	FRMVDVGGL					99.1	92.2	**26.8**	**234.9**
	$KRAS_{G12R}$	GARGVGKSAL	165.8						**53.9**	
	$PIK3CA_{H1047R}$	ARHGGWTTKM					149.0		**67.0**	
		ARHGGWTTK					152.4		**320.3**	
	$PIK3CA_{R88Q}$	RQLCDLRLF		104.0	47.0				**97.9**	**69.4**
	$PTEN_{R130Q}$	GQTGVMICAY			22.4				**74.6**	
	$SF3B1_{R625C}$	YVCNTTARAF				66.7			**18.6**	

	PROTEIN	PEPTIDE	HLA-B								DAI	
			07:02	08:01	15:01	27:05	39:01	44:02	44:03	57:01	58:01	
InDel - NeoAgs	$JAK1_{K860Nfs}$	IVSEKNQQLKW									175.5	159.6	
		SEKNQQLKW						35.0	60.88			43.2	48.9
		VSEKNQQLKW								67.4	43.8	224.6	386.2
	$LARP4B_{T163Hfs}$	HRWIVTSM				152.4						n/a	
		VLKKHWNSAY			32.2							684.0	
		YHRWIVTSM					71.7					n/a	
		GRTLLVTCI				152.1						n/a	
		GRTLLVTCIL				69.9						n/a	
		IVTSMCQSQRW									21.6	n/a	
		LVTCILYHRW									29.8	n/a	
		QRWLTSTTSR				73.0						n/a	
		RTLLVTCIL									122.9	n/a	
		RTLLVTCILY									100.2	n/a	
		STTSRSSALMW									36.0	n/a	
		TSMCQSQRW									10.4	n/a	
		TSRSSALMW									15.5	n/a	
		TTSRSSALMW									11.7	n/a	
		VLKKHWNSA		59.5								228.4	
		VTCILYHRW									12.2	n/a	
		VTSMCQSQRW									10.5	n/a	
	$NFKBIE_{Y254Sfs}$	LTSSPRTETRW								51.7	25.3	137.7	159.1
		RTETRWSTW								32.0	24.2	n/a	n/a
		SPRTETRWST	125.9									n/a	
		TSSPRTETRW								57.8	17.4	621.2	1559.8
		SSPRTETRW									174.8	101.8	
	$RNF43_{G659Vfs}$	HPQRKRRGV	73.0	160.6								180.7	19.3
		VPPSPPLAL	84.2									99.1	
	$UBR5_{E2121Kfs}$	LRVQNQGHLL				189.9						n/a	
		RVQNQGHLLM			72.7							n/a	
		VQNQGHLLM			37.8							n/a	

n/a = not applicable because they are de novo protein sequences which do not have a match in the wt corresponding sequence

antigens (92%) are predicted to bind more than a single haplotype (Suppl. Table 3A and B).

Regarding the SNVs, the number and the score of the predicted mut-Ags greatly varied among the 11 proteins. However, WB and SB SNV mut-Ags are only 2 in SF3B1 (1 from the R625H and 1 from the R625C mutations) and 1 in $BRAF_{V600M}$. 2 of them are predicted to bind more than a single haplotype. All other SNVs did not generate WB and SB (Fig. 8A and Suppl. Table 3A).

Similar to SNVs, the number and the score of the predicted mut-Ags derived from the InDels greatly varied among the 5 mutated proteins. In particular, considering the WB and SB InDel-NeoAgs, the largest number of mut-Ags are predicted in the $LARP4B_{T163Hfs\_47}$ (4 WBs and 10 SBs). On the contrary, InDels identified

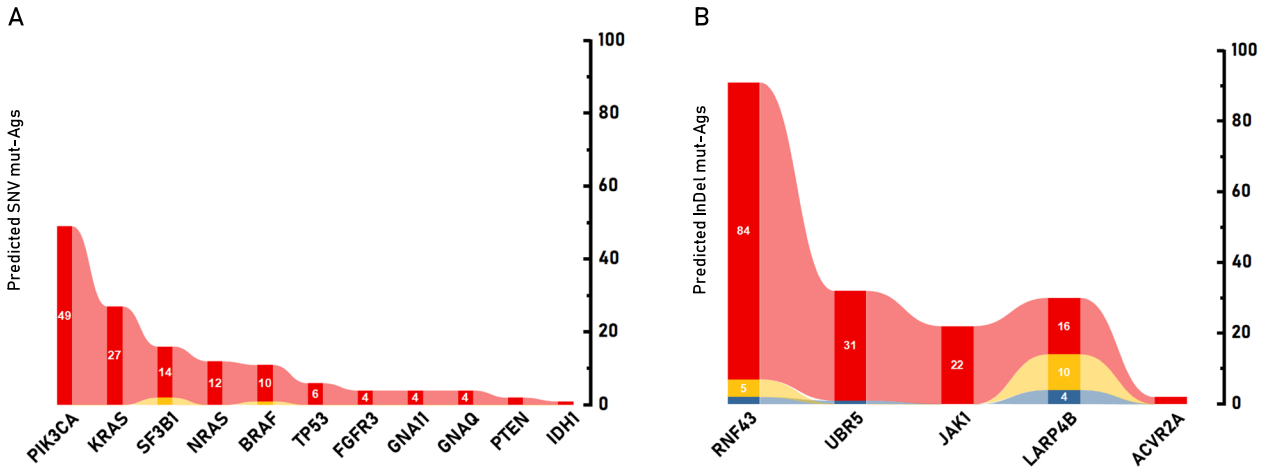

Fig. 8 Predicted shared mut-Ags for HLA-C. The diagrams show the number of putative mut-Ags derived from SNV (**A**) and InDels (**B**). The antigens are grouped by the affinity to the HLA molecules (*PB* poor binders, *WB* weak binders, *SB* strong binders)

in $JAK1_{K860Nfs*16}$ and $ACVR2A_{K437Rfs*5}$ did not generate any top scoring mut-Ags (Fig. 8B and Suppl. Table 3B).

HLA-C haplotypes presenting the predicted mut-Ags

The stratification of the predicted mut-Ags based on the HLA-C haplotypes, showed a significant variability for both SNVs and InDel mutations.

Considering the mutated SNV mut-Ags, the largest number are predicted to be associated with the C*06:02 (nr. 18), the other haplotypes are associated with 10 mut-Ags, on average, except the C*08:01 and C*15:02 which are associated only with 5 and 4 mut-Ags. Restricting the analysis, no WB and only 2 SB with the selected haplotypes are identified, specifically 1 for the HLA-C*03:02 and one for the HLA-C*05:01 (Fig. 9A and Suppl. Figure 6).

Considering the mut-Ags derived from InDels, the largest number are predicted to be associated with the

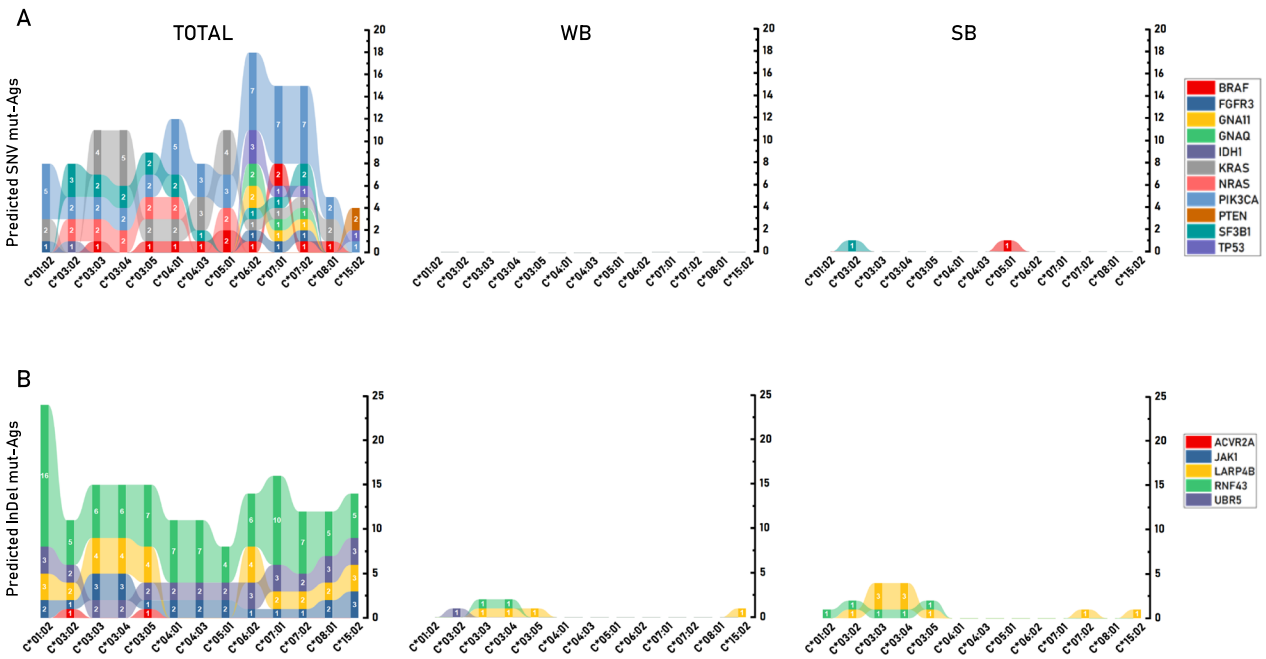

Fig. 9 Stratification of shared mut-Ags based on HLA-C haplotypes. The plots indicate the number of SNV mut-Ags (**A**) and InDel mut-Ags (**B**) for each of the selected HLA-C haplotypes, grouped by the level of binding (WB; SB)

C*01:02 (nr. 24), the other haplotypes are associated with 13 mut-Ags, on average. The evaluation of WB and SB shows that 7 WB are associated with HLA-C*03:02, 03:03, 03:04, 03:05, 15:02 haplotypes. Focusing on the 15 SB, these are associated with HLA-C*01:02 (nr.1), C*03:02 (nr.2), C*03:03 (nr.4), C*03:04 (nr. 4), C*03:05 (nr.2), C* 07:02 (nr.1) and C*15:02 (nr.1) (Fig. 9B and Suppl. Figure 7).

Affinity map of individual predicted mutated mut-Ags to HLA-C haplotypes

The affinity to HLA-C haplotypes of each predicted mut-Ags from SNVs and InDels is shown, including those with single and multiple affinities (Tables 7 and 8).

The evaluation of the SNV mut-Ags showed that the average affinity to the individual haplotypes is 5165.8 nM, ranging from 11,951.2 nM (C*04:01) to 1562.5 nM (C*08:01). In detail, the average affinity to C*03:02 is higher than to the other remaining haplotypes but does not reach the statistical significance, except for C*01:02, 04:01 and 08:01 which show an average affinity significantly lower. SB SNV mut-Ags are identified only in association with HLA- C*03:02 (Nr. 2) and C*05:01 (Nr. 1) haplotypes (Table 7 and Fig. 10A). Looking at the same results from the protein perspective, the average affinity of mut-Ags to all haplotypes is 3909,6 nM, ranging from 6447.7 nM (KRAS) to 1580,06 nM (IDH1) (Table 7 Fig. 10B). None of the proteins showed mut-Ags with an overall average affinity significantly different from all other proteins. However, SB are predicted only for BRAF (Nr. 1) and SF3B1 (Nr. 2) (Fig. 8A and Suppl Table 3A).

Considering the mut-Ags from InDels, the average affinity to the individual haplotypes is 5158.7 nM, ranging from 13,262,8 nM (C*04:01) to 1299.1 nM (C*03:02) (Table 8 and Fig. 10C). In detail, the average affinity to HLA-C*03:02 is higher than to the remaining haplotypes but does not reach the statistical significance, except for C*04:01, 04:03 and 08:01 which show an average affinity significantly lower (Fig. 8B and Suppl. Table 3B). SB InDel mut-Ags are predicted for 01:02, 03:02, 03:03, 03:04, 03:05, 07:02 and 15:02 haplotypes, specifically 1 each of the C*01:02, 07:02 and 15:02; 2 for the C*03:02, 03:05; 4 for the C*03:03, 03:04. No SB InDel mut-Ags are predicted for the remaining HLA-C haplotypes.

From the protein perspective, the average affinity of mut-Ags to all haplotypes is 4698.5 nM, ranging from 7243.1 nM (JAK1) to 779.4 nM (LARP4B). LARP4B protein showed a number of predicted mut-Ags with an overall average affinity significantly higher from all other proteins. In detail, SB InDel mut-Ags are predicted for this protein (nr. 5) and RNF43 (nr. 2). None are predicted for the ACVR2A, JAK1 and UBR5 proteins (Table 8 and Fig. 10D).

Targetable predicted mutated neoantigens linked to HLA-C haplotypes

In order to verify whether the predicted WB and SB mut-Ags derived from both SNVs and InDels may be effective targets for shared off-the-shelf cancer immunotherapy strategies, the differential agretopicity index (DAI) was calculated. According to such analysis, none of the SNV mut-Ags shows a DAI > 10. On the contrary, the 10 InDel mut-Ags identified in $LARP4B_{T163Hfs}$ (6 NeoAgs), $RNF43_{G659Vfs}$ (3 NeoAgs) and $UBR5_{E2121Kfs}$ (1 NeoAg) are potentially targetable because they show either a DAI > 10 ($RNF43_{G659Vfs}$ RGVPPSPPL) or they are de novo protein sequences which do not have a match in the wt corresponding sequence (Table 9). Of these, 3 are WB and 7 SB. The latter are predicted in $LARP4B_{T163Hfs}$ (5 NeoAgs) and in $RNF43_{G659Vfs}$ (2 NeoAgs). Moreover, while $LARP4B_{T163Hfs}$ YHRWIVTSM and SAYLGRTLLV and $RNF43_{G659Vfs}$ ITPPVWHIL are predicted to bind a single HLA-C haplotype, all others are promiscuous binding to more than one haplotype (Table 9).

Total targetable predicted mutated shared neoantigens

A recapitulation of the targetable mutated shared neoantigens derived from the SNV and InDel mutations most frequently identified in cancer and associated to the haplotypes of the three major HLA alleles is performed.

Considering the SNVs, $PIK3CA_{H1047R}$ may generate 5 NeoAgs associated to three different haplotypes, namely HLA-A*31:01 and 33:03 as well as HLA-B*27:05. All other SNVs may generate 1 NeoAg associated to a single haplotype, except for $GNA11_{Q209L}$ FRMVDVGGL, $PIK3CA_{E542K}$ AISTRDPLSK and $SF3B1_{R625C}$ YVCNTTARAF which are predicted to bind two haplotypes. From the HLA perspective, the HLA-B*27:05 is associated with the highest number of SNV-NeoAgs (2 $PIK3CA_{H1047R}$ and 1 $GNA11_{Q209L}$). All others are associated to 1 or maximum 2 SNV-NeoAgs, and in the latter case, these are derived from different mutations (Fig. 11; Table 10).

Considering the InDels, the number of targetable NeoAgs is much larger. Indeed $LARP4B_{T163Hfs}$ may generate 30 NeoAgs, $RNF43_{G659Vfs}$ 9, $NFKBIE_{Y254Sfs}$ 7, $JAK1_{K860Nfs}$ 5 and $UBR5_{E2121Kfs}$ 4. In particular, the 30 InDel-NeoAgs derived from $LARP4B_{T163Hfs}$ are associated with 14 different haplotypes, of which, 10 with the HLA-B*58:01, 4 with HLA-A*31:01 and 3 with each HLA-B*27:05, C*03:03, C*03:04, respectively.

Remarkably, the HLA-B*58:01 is associated with 16 InDel-NeoAgs (10 $LARP4B_{T163Hfs}$, 4 $NFKBIE_{Y254Sfs}$ and 2 $JAK1_{K860Nfs}$). Second in the ranking is the HLA-A*31:01 associated with 8 InDel-NeoAgs (5 $LARP4B_{T163Hfs}$, 2 $RNF43_{G659Vfs}$ and 1 $NFKBIE_{Y254Sfs}$). All others are associated with 1 or maximum 6

Table 7 Affinity map of SNV mut-Ags predicted to bind HLA-C haplotypes

PROTEIN	PEPTIDE	HLA-C											
		01:02	03:02	03:03	03:04	03:05	04:01	04:03	05:01	06:02	07:01	07:02	08:01	15:02
BRAF	EKSRWSGSHQF											11276		
	KIGDFGLATM								1812					
	IGDFGLATM			561		1705	4937	449	97.0				529	
	MKSRWSGSHQF									6044	3119	2261		
FGFR3	RCPHRPIL	5374												
	ERCPHRPIL									2572	2085	3301		
IDH1	IIIGCHAY		1580											
KRAS	AGDVGKSAL	8694		6635	6635	10961	11958	3223	677				6021	
	ARGVGKSAL									8183	4940	5910		
	AVGVGKSAL	8838		5255	5255									
	GADGVGKSAL			3787	3787			4844	1127				6021	
	GAGDVGKSAL													
	GAGDVGKSAL							12676	4632				16056	
	GAVGVGKSAL			982	982	2773							6021	
	GADGVGKSA												19403	
	GARGVGKSAL					4245								
NRAS	AGKEEYSAM		1374	2912	2912	3478								
	AGREEYSAM		739	1582	1582	1744								
	ILDTAGKEEY						19997		6735					
	ILDTAGREEY						17751		5964					
PIK3CA	ARHGGWTTKM									1986	1814	3311		
	EFFDETRQL	16579					20157			2021	3556	1995		
	FFDETRQL	9326	4761	4613	4613	6359	239	1269	2311	7999	5150	1570	7415	
	FFDETRQLC	11323					1651	3624	5063			5753	11819	
	FFDETRQLCDL	13518					1003	2332	3478			4508		
	RHGGWTTKM									7119	3361	2160		
	STRDPLSKI	10171	3570	4412	4412	3556				2968	5446			453
	TRDPLSKI						25045			10401				
	TRQLCDLRL									3039	1906	2596		
	ARHGGWTTK										8536			
	ITKQEKDFL													5795
PTEN	KAGKGGTGV													2426
	KAGKGQTGV													2977
SF3B1	NMDEYVHNT							11694		8332				
	VCNTTARAF		372	4213	4213	6481								
	VHNTTARAF						16768			4629	1689	526		
	YVHNTTARAF		25	511	511	1398						989		
	YVCNTTARAF		75											
TP53	MNQRPILTI													590
	NQRPILTII									2832				
	QRPILTII									5856				
	QRPILTIITL									2944	3381.	2639		

InDels-NeoAgs, and in the latter case, these are derived from different mutations (Fig. 12; Table 11).

In order to verify whether the targetable shared Neo-Ags described in the present study are novel, a search in different databases was conducted Considering the 12 targetable SNV-NeoAgs, 10 have been already reported in literature or covered by a patent [21–24]. The remaining 2 NeoAgs are novel, namely PIK3CAE545K QAMESEITK (HLA-A*11:01) and SF3BIER625C YVC-NTTARAF (HLA-B*15:02) (Table 10).

Table 8 Affinity map of InDel mut-Ags predicted to bind HLA-C haplotypes

PROTEIN	PEPTIDE	HLA-C												
		01:02	03:02	03:03	03:04	03:05	04:01	04:03	05:01	06:02	07:01	07:02	08:01	15:02
ACVR2A	VVHKKRGLF		1785											
	VVVHKKRGL					8782								
JAK1	IVSEKNQQL	2551	942	318	318	1118	12560	3973	2649	4961	3444	6831	1029	5279
	VSEKNQQL	15870		14314	14314		21215	8307	4334				19224	10399
	QQLKWTPHI													5394
LARP4B	HRWIVTSM									1741	1847	1603		
	NSAYLGRTL			78	78	298								
	SAYLGRTL	3016	339	127	127	113							7247	
	SAYLGRTLL	626	35	12	12	26				826			1361	164
	STTSRSSAL	502		59	59	266								414
	YHRWIVTSM									509	380	98		
	AYLGRTLLV									1351				
	SAYLGRTLLV													70
RNF43	FFPITPPVW	5970	3252				10046	15664		7318	5388	1093		
	FPITPPVWHIL	2613		4184	4184	5325							14319	
	GVPPSPPLAL	752		10043	10043	13278		17104						
	ITPPVWHI	3460						13619						4223
	ITPPVWHIL	77	1289	765	765	1124	10025	3784	4813	3405	2244	1566	6564	1270
	LALGPRMQL	3720	41	14	14	36	17571	8960	9100	5503	3894	6005	2675	290
	PRMQLCTQL									6175	5009			
	RFFPITPPV	5602					9590			2057	2612	1011		
	RGVPPSPPL	1346	345	114	114	214	13342	9295	7709		10971		3796	1633
	RGVPPSPPLAL	2874				6060	21538				3672	4551		
	VPPSPPLAL	435	3243	1397	1397	2899	9799	8212	10376	20203	15384	8838	8073	13592
	FFPITPPV	2957												
	GVPPSPPL	1014												
	GVPPSPPLA	9137												
	KRRGVPPSPPL											7892		
	PITPPVWHIL	5017												
	PPSPPLAL	18597												
	RFFPITPPVW										10676			
	TPPVWHIL	5863												
	WHILGPQRH											8153		
	LRVQNQGHLL									2014				
UBR5	LRVQNQGHL									1465	2067	1966		
	RVQNQGHLL	6087	2840	2278	2278	3582	9185	5214	3907		4609		11840	512
	VQNQGHLLM	8731	178	530	530	1775	11020	6220	3104	4313	2487	1797	4254	1299
	NSPCCQKKL	2228												
	QNQGHLLMI													1450
	RVQNQGHLLM													1311
	VQNQGHLL												24009	

Considering the 53 targetable InDel-NeoAgs, 13 antigens, 4 derived from $JAK1_{K860Nfs}$ and 9 $RNF43_{G659Vfs,}$ are already covered by a patent [24, 25]. The remaining 40 NeoAgs are novel, namely all those derived from $LARP4B_{T163Hfs}$, $NFKBIE_{Y254Sfs}$, $UBR5_{E2121Kfs}$ and 1 derived from the $JAK1_{K860Nfs}$ mutation, covering a broad range of haplotypes in the HLA-A, B and C alleles (Table 11).

Potential clinical application of mutated shared NeoAgs

The subsequent step was to verify whether the targetable SNV and InDel-NeoAgs may have a potential clinical application in cancers. The mutated proteins from which such NeoAgs derive, are identified in a number of tumor types at a frequency > 5%. This ranges from 7% in Bladder&Prostate (BLCA) cancers to 50% in Uveal melanoma (UVM). Notably, tumor types with a 5-year overall survival < 10%, namely pancreatic (PAAD) and stomach (STAD) cancers, are characterized by mutated proteins generating targetable shared NeoAgs, which can be useful in 30 and 24% of patients, respectively (Fig. 13).

Considering that the targetable shared SNV and InDel-NeoAgs are associated with haplotypes showing a uneven prevalence in different populations, vaccines based

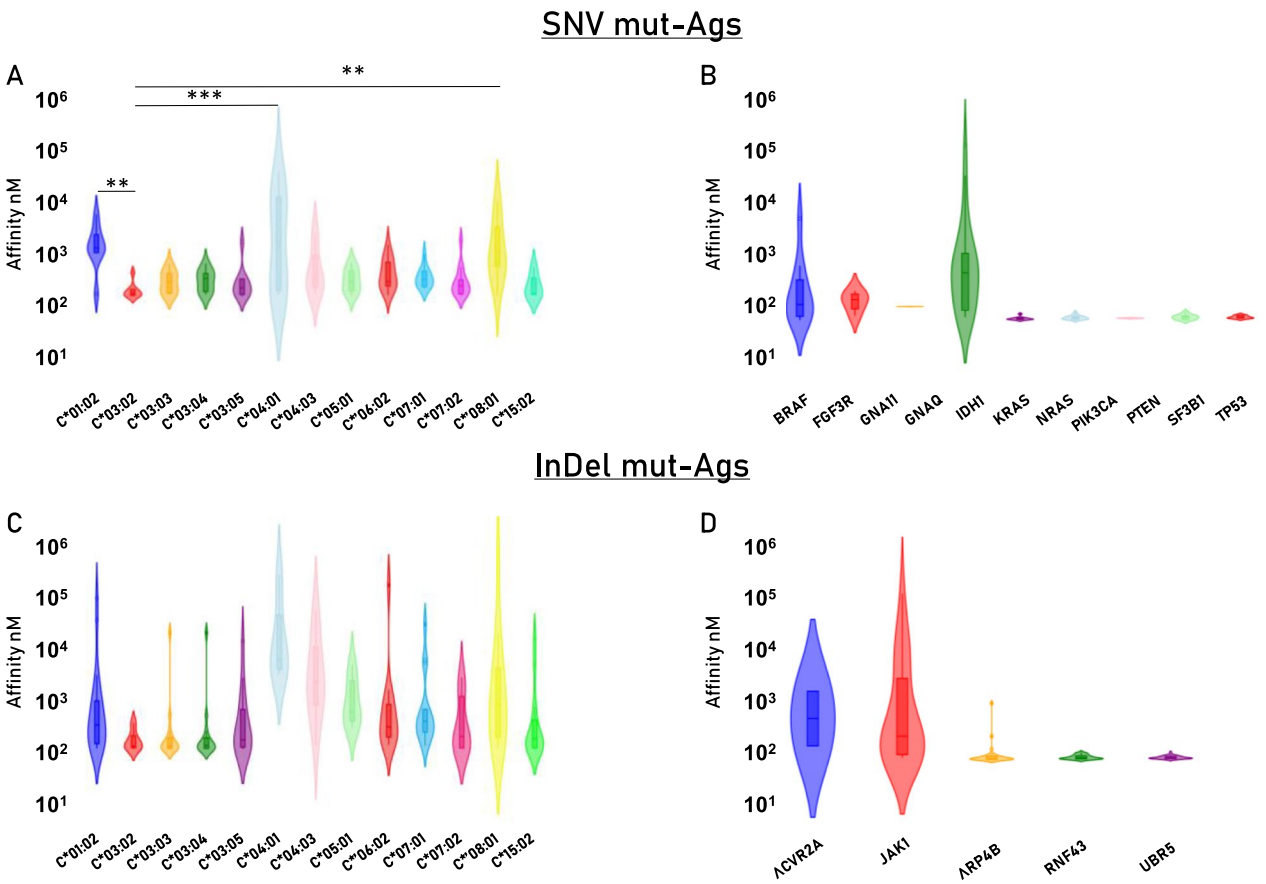

Fig. 10 Affinity of shared SNV and InDel mut-Ags to HLA-C haplotypes. The affinity to HLA-C haplotypes of all predicted mut-Ags (PB, Wb and SB) are plotted. Values are shown according to each haplotype (**A** and **C**) or mutated protein from which the mut-Ags are derived (**B** and **D**). *nM* nanomolarity

Table 9 WB and SB targetable Neo-Ags binding to HLA-C haplotypes

	PROTEIN	PEPTIDE	HLA-C							DAI
			01:02	03:02	03:03	03:04	03:05	07:02	15:02	
InDel-NeoAgs	$LARP4B_{Tl63Hfs}$	NSAYLGRTL			78	78				n/a
		SAYLGRTL			127	127	113			n/a
		SAYLGRTLL		35	12	12	26		164	n/a
		STTSRSSAL			59	59				n/a
		YHRWIVTSM						98		n/a
		SAYLGRTLLV							70	n/a
	$RNF43_{G659Vfs}$	ITPPVWHIL	77							n/a
		LALGPRMQL		41	14	14	36			n/a
		RGVPPSPPL			114	114				396
	$UBR5_{E2121Kfs}$	VQNQGHLLM		178						n/a

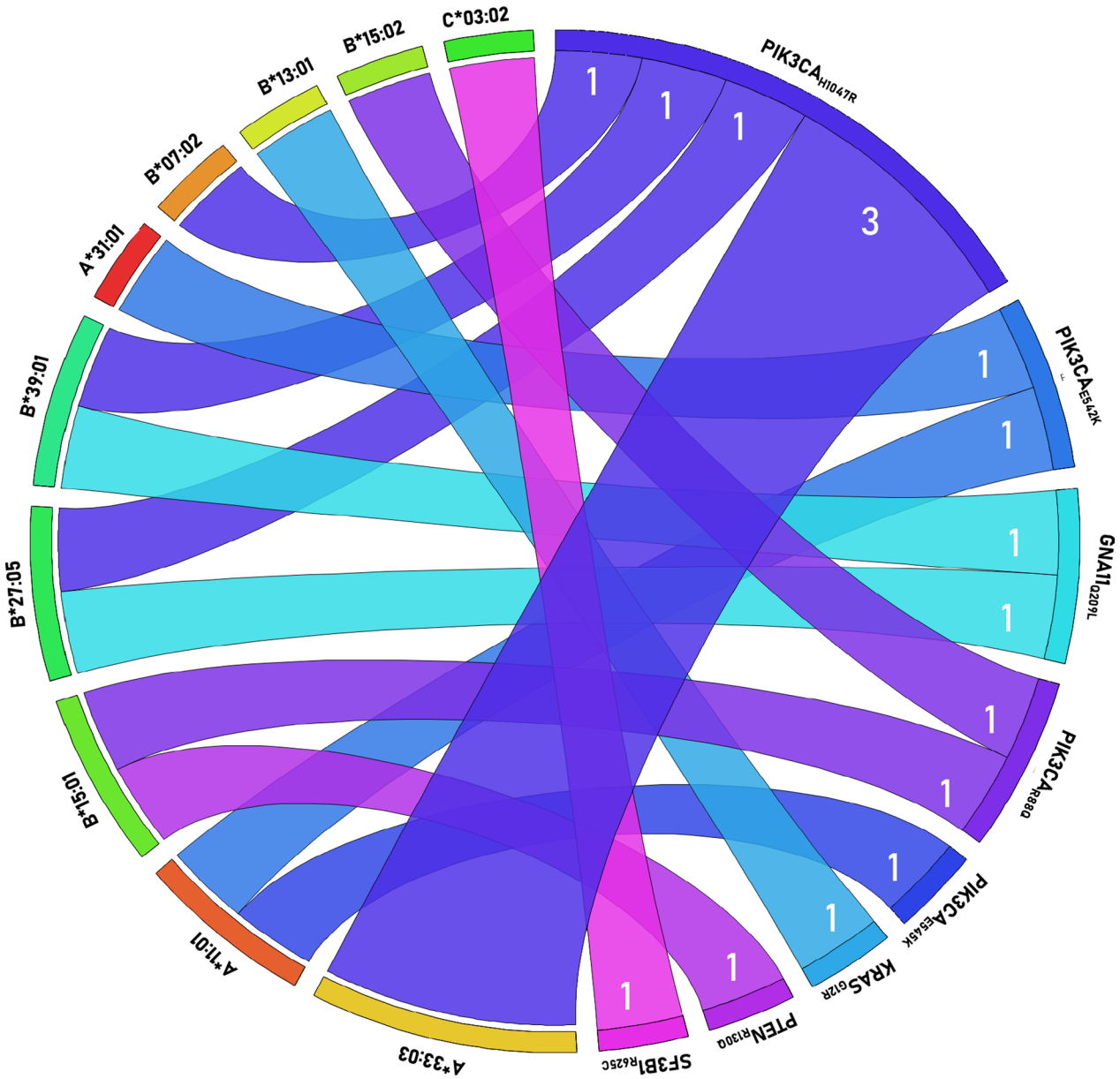

Fig. 11 Targetable shared SNV-NeoAgs. The shared NeoAgs with DAI > 10 (targetable SNV-NeoAgs) derived from each protein characterized by SNV mutation are plotted according to the associated HLA-A, B or C haplotypes. Indicated are the numbers of NeoAgs from each of the mutations

on such NeoAgs may have a global or a regional application. As example, an off-the-shelf cancer vaccine for colorectal (COAD), stomach (STAD) and uterus endometrial (UCEC) cancers may have a global coverage. Indeed, the SNV and InDel-NeoAgs derived from their tumor-specific mutated proteins are associated with haplotypes, including HLA-A*02:01, 03:01 and 11:01, which collectively cover > 10% of populations all over the globe. On the contrary, off-the-shelf cancer vaccines for pancreatic (PAAD) cancer or uveal melanoma (UVM) may be developed only for a strict regional coverage. Indeed, the SNV and InDel-NeoAgs derived from their tumor-specific mutated proteins are associated with HLA-A*03:01, which covers > 10% of populations only in Europe (LUAD and PAAD) or with HLA-A*33:03 and B*15:02, which covers > 10% of populations only in Oceania and South-East Asia (UVM) (Fig. 14).

Table 10 Novelty of targetable SNV-NeoAgs

Mutation	Sequence	Haplotype	Novel	Reference
$GNA11_{Q209L}$	FRMVDVGGL	B*27:05/39:01	N	(24)
$KRAS_{G12R}$	GARGVGKSAL	B*07:02	N	(21), (22)
$PIK3CA_{E542K}$	AISTRDPLSK	A*03:01/11:01	N	(22), (23)
$PIK3CA_{E545K}$	QAMESEITK	A*11:01	Y	
$PIK3CA_{H1047R}$	YFMKQMNDAR	A*33:03	N	(22)
	EYFMKQMNDAR	A*33:03	N	(22)
	FMKQMNDAR	A*31:01	N	(22)
	ARHGGWTTKM	B*27:05	N	(24)
	ARHGGWTTK	B*27:05	N	(24)
$PIK3CA_{R88Q}$	RQLCDLRLF	B*13:01/15:01	N	(22)
$PTEN_{R130Q}$	GQTGVMICAY	B*15:01	N	(24)
$SF3B1_{R625C}$	YVCNTTARAF	B*15:02	Y	

Discussion

The SNV and InDel mutations reported in the TCGA database as those identified in at least 5% of cancer patients are selected for predicting shared mutated neoantigens (SNV and InDel-NeoAgs). 50 haplotypes in the three MHC class I loci (10 HLA-A, 27 HLA-B and 13 HLA-C) were selected to cover > 80% of the global population.

Out of the 100 + mutated proteins falling in the selection parameter, only 18 proteins were identified to generate mutated antigens (mut-Ags) in association with one or more of the selected haplotypes. In particular, 11 proteins are characterized by single nucleotide variations (SNV), leading to a single aminoacid point mutation and 7 proteins are characterized by insertion/deletions (InDel), leading to a frameshift with an alternative de novo codon translation. Overall, 568 SNV and 502 InDel mut-Ags are predicted from all the 18 mutated proteins in association with the 50 haplotypes; however, only 72 (12.7%) SNV and 79 (15.7%) InDel show high affinity to the selected haplotypes (WB and SB). Furthermore, taking into consideration only the predicted mut-Ags with the highest affinity to the haplotypes (< 100 nM = SBs), only 42 SNV (7.4%) and 51 InDel (10.1%) are identified. Interestingly, SNVs in NRAS, IDH1 and PTEN as well as InDels in NFKBIE, RPL22, UBR5 and ACVR2 proteins do not generate SBs in any of the three HLA alleles. The biological explanation for this observation requires further experimental assessments. In the quest of targetable shared neoantigens (NeoAgs) from the WB and SBs, only 12/72 (16.7%) mut-Ags derived from SNV mutations and 55/79 (69.6%) mut-Ags derived from InDel mutations can be defined as neoantigens. Indeed, they either show an affinity to the same MHC > tenfold compared to the corresponding nonmutated counterpart (differential agretopicity index (DAI) > 10) or derive from de novo protein sequences which do not have a match in the wt corresponding sequence. In particular, the latter derive from "abnormal" mRNAs generated by the frameshift, contain premature termination codons (PTCs) which may be recognized and degraded by nonsense-mediated mRNA decay (NMD) or undergo a translational repression [26–29]. Nevertheless, InDel-NeoAgs have been identified by MS/MS on a set of tumor cells and their immunogenicity has been proven by ex vivo stimulation of PBMCs from both healthy donors (HD) as well as tumor patients [30]. Therefore, both SNV-NeoAgs and InDel-NeoAgs are presented by tumor cells and recognized by the immune system as "non-self" antigens representing potent immunogenic targets.

The $PIK3CA_{H1047R}$ SNV mutation encodes the largest number of SNV-NeoAgs (nr. 5) providing a set of potential shared mutated neoantigens for developing off-the-shelf cancer vaccines targeting 12% of breast cancer (BRCA), 7% of uterine carcinosarcoma (UCS) as well as 5% of Uterine Corpus Endometrial Carcinoma (UCEC) patients.

Such SNV-NeoAgs are associated with three HLA haplotypes, namely HLA-A*31:01, 33:03, present in > 10% of the population in South America, North-East and South-East Asia, and HLA-B*27:05, present in < 5% only in Europe. All the other 7 SNV mutations encode a single NeoAg each.

The $LARP4B_{T163Hfs}$ InDel mutation encodes for the largest number of InDel-NeoAgs (nr. 29) providing a set of potential shared mutated neoantigens for developing off-the-shelf cancer vaccines targeting 5.3% of stomach cancer (STAD) patients.

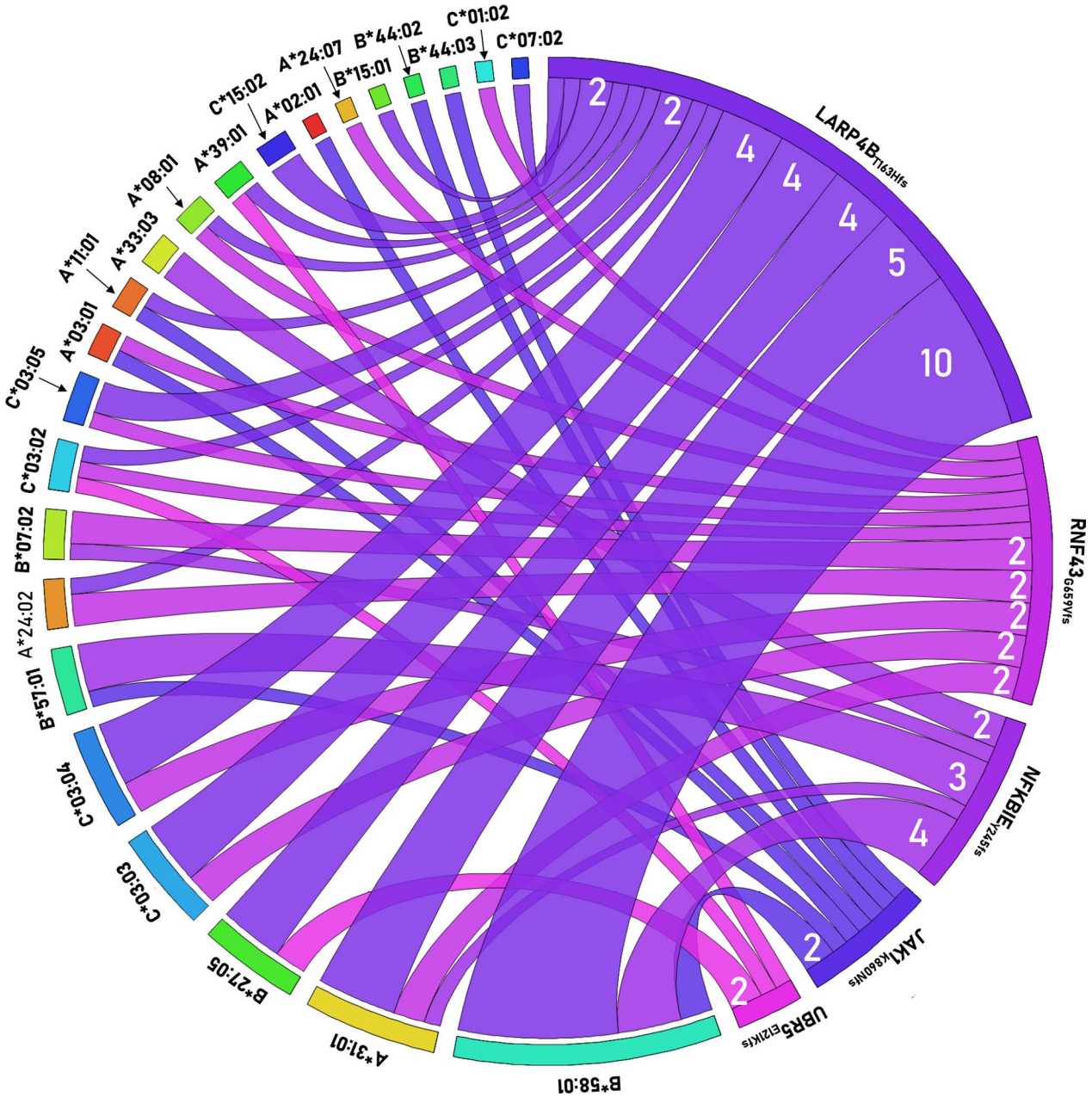

Fig. 12 Targetable shared InDel-NeoAgs. The shared NeoAgs with DAI > 10 or de novo (targetable InDel-NeoAgs) derived from each protein characterized by InDel mutation are plotted according to the associated HLA-A, B or C haplotypes. Indicated are the numbers of NeoAgs, starting from 2, from each of the mutations

Such SNV-NeoAgs are associated with several HLA haplotypes, covering most of the Continents. The other 4 InDel mutations encode 3 to 9 NeoAgs each.

The HLA haplotypes more frequently associated with both SNV and InDel-NeoAgs are the A*31:01 (South America, North-East and South-East Asia), B*58:01 (South-East Asia) and C*03:03/03:04 (North-East and South-East Asia), considering the distribution in > 10% of the population.

The length of all predicted NeoAgs spans from 8 to 11 aa, with a predominance of 9mer (36/67) which is the most frequently identified in the peptides naturally presented by MHC class I molecules [30]. Of these, seven 9mer are among the SNV-NeoAgs and 29 are among the InDel-NeoAgs and they should be considered as the ones

Table 11 Novelty of targetable InDel-NeoAgs

Mutation	Sequence	Haplotype	Novel	Reference
JAK1$_{K860Nfs}$	QLKWTPHILK	A*03:01/11:01	N	[25]
	QQLKWTPHI	A*02:01	N	
	IVSEKNQQLKW	B*08:01	Y	
	SEKNQQLKW	B*44:02/44:03	N	[24]
	VSEKNQQLKW	B*57:01/58:01	N	[24]
LARP4B$_{T163Hfs}$	HWNSAYLGR	A*31:01	Y	
	AYLGRTLLV	A*24:02	Y	
	LVTCILYHR	A*31:01	Y	
	RWLTSTTSR	A*31:01	Y	
	RTLLVTCILY	A*11:01/B*58:01	Y	
	SQRWLTSTTSR	A*31:01	Y	
	VTSMCQSQR	A*31:01	Y	
	HRWIVTSM	B*27:05	Y	
	VLKKHWNSA	B*08:01	Y	
	VLKKHWNSAY	B*15:01	Y	
	YHRWIVTSM	B*39:01	Y	
	GRTLLVTCI	B*27:05	Y	
	GRTLLVTCIL	B*27:05	Y	
	IVTSMCQSQRW	B*58:01	Y	
	LVTCILYHRW	B*58:01	Y	
	QRWLTSTTSR	B*27:05	Y	
	RTLLVTCIL	B*58:01	Y	
	STTSRSSALMW	B*58:01	Y	
	TSMCQSQRW	B*58:01	Y	
	TSRSSALMW	B*58:01	Y	
	TTSRSSALMW	B*58:01	Y	
	VTCILYHRW	B*58:01	Y	
	VTSMCQSQRW	B*58:01	Y	
	NSAYLGRTL	C*03:03/03:04	Y	
	SAYLGRTL	C*03:03/03:04/03:05	Y	
	SAYLGRTLL	C*03:03/03:04/03:05/15:02	Y	
	STTSRSSAL	C*03:03/03:04	Y	
	SAYLGRTLLV	C*15:02	Y	
	YHRWIVTSM	B*39:01/C*07:02	Y	
NFKBIE$_{Y254Sfs}$	LTSSPRTETR	A*31:01	Y	
	TSSPRTETR	A*33:03	Y	
	LTSSPRTETRW	B*57:01/58:01	Y	
	RTETRWSTW	B*57:01/58:01	Y	
	SPRTETRWST	B*07:02	Y	
	TSSPRTETRW	B*57:01/58:01	Y	
	SSPRTETRW	B*58:01	Y	
RNF 43 $_{G659Vfs}$	MQLCTQLAR	A*33:03	N	[24]
	FFPITPPVW	A*24:02	N	[24]
	RFFPITPPVW	A*24:02/27:02	N	[24]
	RMQLCTQLAR	A*03:01/31:01	N	[24]
	HPQRKRRGV	B*07:02/08:01	N	[24]
	VPPSPPLAL	B*07:02	N	[24]
	ITPPVWHIL	C*01:01	N	[24]
	LALGPRMQL	C*03:02/03:03/03:04/03:05	N	[24]
	RGVPPSPPL	C*03:03/03:04	N	[24]
UBR5$_{E2121Kfs}$	LRVQNQGHLL	B*27:05	Y	
	RVQNQGHLLM	B*15:01	Y	
	VQNQGHLLM	B*15:01/C*03:02	Y	

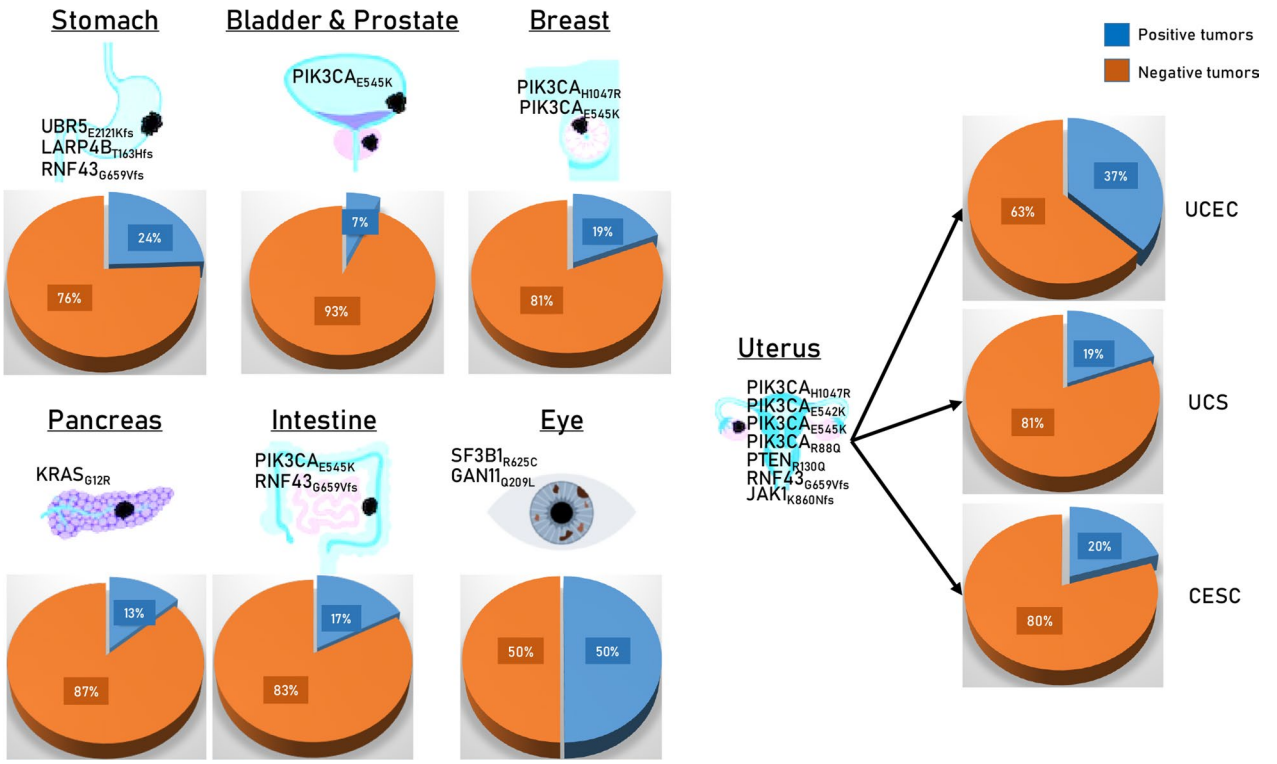

Fig. 13 Frequency of targetable shared NeoAgs in cancers. The distribution and percentage in cancers of the mutated proteins generating targetable shared SNVs and InDel-NeoAgs is shown. The percentage (blue) indicates the sum of the percentages for each mutated protein

with the highest probability to be expressed by cancer cells. Only the shared NeoAg derived from $PIK3CA_{E542K}$ (AISTRDPLSK) have been previously experimentally validated [22–24]. For the remaining predicted NeoAgs, although the highly stringent affinity values applied in the present study strongly suggest the natural presentation by cancer cells, a definitive experimental validation is required.

Overall, considering the mutations encoding the SNV and InDel-NeoAgs as well as the associated HLA alleles, our findings provide the most comprehensive set of immunologically relevant shared mutated neoantigens for development of cancer vaccines. In particular, they may have a global reach or be more appropriate for regional applications, depending on the frequency of the relevant haplotypes in different populations.

Specifically, colorectal cancer (COAD), stomach cancer (STAD), and uterine endometrial cancer (UCEC), are characterized by mutations encoding NeoAgs associated with a broad range of different haplotypes. Consequently, a universal "off-the-shelf" vaccine for these tumors can be developed. On the contrary, pancreatic cancer (PAAD), and uveal melanoma (UVM) are characterized by mutations encoding NeoAgs associated with haplotypes prevalent in certain ethnic or regional groups only. Therefore, a "regional" vaccine approach may be considered.

In the present study, the analysis has been focused only on MHC-class I neoantigens, which are the final effective target of $CD8^+$ T cells cytotoxic effect. Nevertheless, both the SNV and InDel mutations might likely generate MHC-class II neoantigens which could elicit mutation-specific $CD4^+$ T helper cells to potentiate the $CD8^+$ CTLs. However, such analysis requires a subsequent follow up with integrated approaches, considering that prediction tools for MHC-class II epitopes are not as robust as those for the MHC-class I epitopes.

The identification and validation of shared mutated neoantigens, represent a key finding for the development of "off-the-shelf" cancer vaccines, which could completely change cancer treatment worldwide. The next step will be to develop a validation platform to experimentally prove that the predicted shared mutated neoantigens are identified in the ligandome of tumor cells and are truly immunogenic. Indeed, SNV-NeoAgs could be poorly immunogenic if the corresponding wt is tolerogenic and InDel-NeoAgs could be poorly represented in the cancer ligandome.

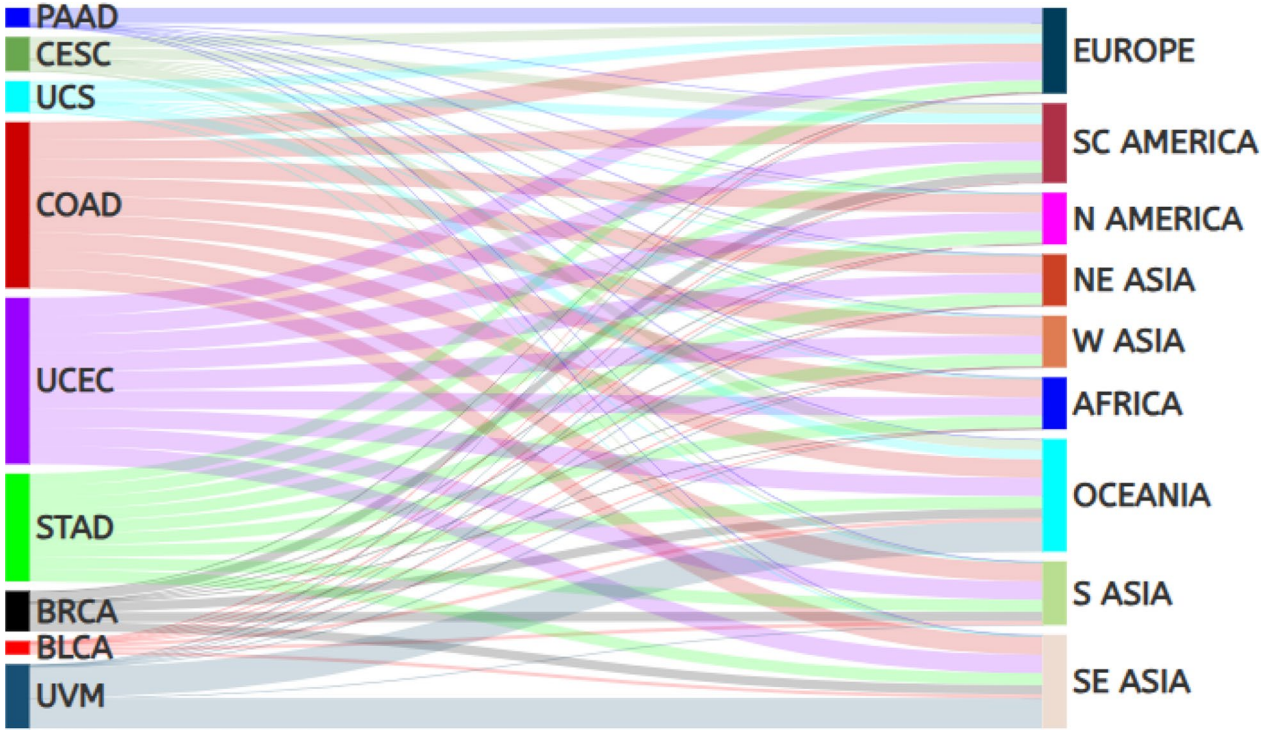

Fig. 14 Coverage of off-the-shelf cancer vaccines based on targetable shared NeoAgs. The geographic applicability of off-the-shelf cancer vaccines based on targetable shared NeoAgs is shown according the HLA haplotypes frequency in distinct world populations. The size of the connecting ribbons correlates with the percentage of the specific SNV or InDel mutation in each tumor type (left). The height of the bar indicates the total percentage of cancer types targetable in each Continent, based on the HLA haplotype prevalence (right)

In conclusion if confirmed, such vaccines would provide a fast, cost-effective solution to the challenge of immunotherapy, enabling treatment to be quickly employed across various patient populations. This would help to address the global cancer burden by providing new options for patients who may not have access to personalized therapies or for cancers with low tumor mutational burden (TMB) and a low number of unique neoantigens.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12967-025-06478-3.

Supplementary material 1: Fig. 1 Strategy for selecting the protein sequence for antigen prediction. Fig. 2 Shared mut-Ags predicted in HLA-A haplotypes from each protein carrying a SNV mutation. The plots indicate the number of SNV mut-Ags for each of the mutated protein in association with indicated HLA-A haplotypes. Mut-Ags are grouped by the level of binding (WB; SB). Fig. 3 Shared mut-Ags predicted in HLA-A haplotypes from each protein carrying an InDel mutation. The plots indicate the number of InDel mut-Ags for each of the mutated protein in association with indicated HLA-A haplotypes. Mut-Ags are grouped by the level of binding (WB; SB). Fig. 4 Shared mut-Ags predicted in HLA-B haplotypes from each protein carrying a SNV mutation. The plots indicate the number of SNV mut-Ags for each of the mutated protein in association with indicated HLA-B haplotypes. Mut-Ags are grouped by the level of binding (WB; SB). Fig. 5 Shared mut-Ags predicted in HLA-B haplotypes from each protein carrying an InDel mutation. The plots indicate the number of InDel mut-Ags for each of the mutated protein in association with indicated HLA-B haplotypes. Mut-Ags are grouped by the level of binding (WB; SB). Fig. 6 Shared mut-Ags predicted in HLA-C haplotypes from each protein carrying a SNV mutation. The plots indicate the number of SNV mut-Ags for each of the mutated protein in association with indicated HLA-C haplotypes. Mut-Ags are grouped by the level of binding (WB; SB). Fig. 7 Shared mut-Ags predicted in HLA-C haplotypes from each protein carrying an InDel mutation. The plots indicate the number of InDel mut-Ags for each of the mutated protein in association with indicated HLA-C haplotypes. Mut-Ags are grouped by the level of binding (WB; SB).

Supplementary material 2: Table 1A. Missense mutations predicted in HLA-A haplotypes. Table 1B. Frameshift mutations predicted in HLA-A haplotypes. Table 2A. Missense mutations predicted in HLA-B haplotypes. Table 2B. Frameshift mutations predicted in HLA-B haplotypes. Table 3A. Missense mutations predicted in HLA-C haplotypes. Table 3B. Frameshift mutations predicted in HLA-C haplotypes.

Acknowledgements
Not applicable.

Author contributions
AM and BC performed 80% of all the antigen prediction analyses; CR performed the remaining 20% of the antigen prediction analyses. MT and LB designed the structure of the review article, supervised the analysis. AM, BC and LB drafted the manuscript. All the Authors revised the manuscript.

Funding
The study was funded by the Italian Ministry of Health through Institutional "Ricerca Corrente" (Project L2/3 to LB; Project L2/13 to MT); the PNRR Ministero Salute PNRR-POC-2022-12375769 "Molecular mimicry to improve liver cancer immunotherapy" (2023–2025) CUP Master H63C22000420006 (to LB).

Availability of data and materials
Data and material will be deposited and publicly available.

Declarations

Ethics approval and consent to participate
N/A.

Consent for publication
The corresponding author has received consent for publication.

Competing interests
The authors declare no potential competing interests.

Received: 28 February 2025 Accepted: 10 April 2025
Published online: 19 May 2025

References

1. Martincorena I, Raine KM, Gerstung M, et al. Universal patterns of selection in cancer and somatic tissues. Cell. 2017;171:1029-1041.e21.
2. Greenman C, Stephens P, Smith R, et al. Patterns of somatic mutation in human cancer genomes. Nature. 2007;446:153–8.
3. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. Nature. 2009;458:719–24.
4. Martincorena I, Campbell PJ. Somatic mutation in cancer and normal cells. Science. 2015;349:1483–9.
5. Tomasetti C, Marchionni L, Nowak MA, Parmigiani G, Vogelstein B. Only three driver gene mutations are required for the development of lung and colorectal cancers. Proc Natl Acad Sci USA. 2015;112:118–23.
6. Perrin-Vidoz L, Sinilnikova OM, Stoppa-Lyonnet D, Lenoir GM, Mazoyer S. The nonsense-mediated mRNA decay pathway triggers degradation of most BRCA1 mRNAs bearing premature termination codons. Hum Mol Genet. 2002;11:2805–14.
7. Wagner E, Lykke-Andersen J. mRNA surveillance: the perfect persist. J Cell Sci. 2002;115:3033–8.
8. You KT, Li LS, Kim NG, et al. Selective translational repression of truncated proteins from frameshift mutation-derived mRNAs in tumors. PLoS Biol. 2007;5:e109.
9. Cresswell P, Ackerman AL, Giodini A, Peaper DR, Wearsch PA. Mechanisms of MHC class I-restricted antigen processing and cross-presentation. Immunol Rev. 2005;207:145–57. https://doi.org/10.1111/j.0105-2896.2005.00316.x.
10. Buonaguro L, Tagliamonte M. Peptide-based vaccine for cancer therapies. Front Immunol. 2023 Aug 16;14:1210044. https://doi.org/10.3389/fimmu.2023.1210044. Erratum in: Front Immunol. 2023 Oct 26;14:1324894. https://doi.org/10.3389/fimmu.2023.1324894. PMID: 37654484; PMCID: PMC10467431.
11. Buonaguro L, Tagliamonte M. Selecting target antigens for cancer vaccine development. Vaccines (Basel). 2020;8(4):615. https://doi.org/10.3390/vaccines8040615.
12. Castle JC, Uduman M, Pabla S, Stein RB, Buell JS. Mutation-Derived Neoantigens for Cancer Immunotherapy. Front Immunol. 2019;10:1856.
13. Bailey MH, Tokheim C, Porta-Pardo E, et al. Comprehensive characterization of cancer driver genes and mutations [published correction appears in Cell. 2018 Aug 9;174(4):1034-1035. 10.1016/j.cell.2018.07.034]. Cell. 2018;173(2):371-385.e18. https://doi.org/10.1016/j.cell.2018.02.060.
14. Charoentong P, Finotello F, Angelova M, et al. Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. Cell Rep. 2017;18(1):248–62. https://doi.org/10.1016/j.celrep.2016.12.019.
15. Zhao W, Wu J, Chen S, Zhou Z. Shared neoantigens: ideal targets for off-the-shelf cancer immunotherapy. Pharmacogenomics. 2020;21(9):637–45. https://doi.org/10.2217/pgs-2019-0184.
16. Duan F, Duitama J, Al Seesi S, et al. Genomic and bioinformatic profiling of mutational neoepitopes reveals new rules to predict anticancer immunogenicity. J Exp Med. 2014;211(11):2231–48. https://doi.org/10.1084/jem.20141308.
17. Pearlman AH, Hwang MS, Konig MF, et al. Author correction: targeting public neoantigens for cancer immunotherapy. Nat Cancer. 2021;2(8):865–7. https://doi.org/10.1038/s43018-021-00246-0.
18. Ragone C, Cavalluzzo B, Mauriello A, Tagliamonte M, Buonaguro L. Lack of shared neoantigens in prevalent mutations in cancer. J Transl Med. 2024;22(1):344. https://doi.org/10.1186/s12967-024-05110-0.
19. Rech AJ, Balli D, Mantero A, et al. Tumor immunity and survival as a function of alternative neopeptides in human cancer. Cancer Immunol Res. 2018;6(3):276–87. https://doi.org/10.1158/2326-6066.CIR-17-0559.
20. Sercarz EE, Lehmann PV, Ametani A, Benichou G, Miller A, Moudgil K. Dominance and crypticity of T cell antigenic determinants. Annu Rev Immunol. 1993;11:729–66. https://doi.org/10.1146/annurev.iy.11.040193.003501.
21. Choi J, Goulding SP, Conn BP, et al. Systematic discovery and validation of T cell targets directed against oncogenic KRAS mutations. Cell Rep Methods. 2021;1(5):100084. https://doi.org/10.1016/j.crmeth.2021.100084.
22. Iizuka A, Akiyama Y, Sakura N, et al. Generation of novel complete HLA class I monoallelic cell lines used in an MHC stabilization assay for neoantigen evaluation. Oncol Lett. 2023;26(2):324. https://doi.org/10.3892/ol.2023.13910.
23. Shen M, Chen S, Han X, et al. Identification of an HLA-A*11:01-restricted neoepitope of mutant PIK3CA and its specific T cell receptors for cancer immunotherapy targeting hotspot driver mutations. Cancer Immunol Immunother. 2024;73(8):150. https://doi.org/10.1007/s00262-024-03729-y.
24. Chen TF et al. Engineered erythroid cells including loadable antigen-presenting polypeptides and methods of use (Patent Application: USP 20200291355-United States). 2020.
25. Juneja V. Neoantigens and their uses (Patent Application: BR112020025764A2-Brazil). 2019.
26. Perrin-Vidoz L, Sinilnikova OM, Stoppa-Lyonnet D, Lenoir GM, Mazoyer S. The nonsense-mediated mRNA decay pathway triggers degradation of most BRCA1 mRNAs bearing premature termination codons. Hum Mol Genet. 2002;11:2805–14. https://doi.org/10.1093/hmg/11.23.2805.
27. Wagner E, Lykke-Andersen J. mRNA surveillance: the perfect persist. J Cell Sci. 2002;115:3033–8. https://doi.org/10.1242/jcs.115.15.3033.
28. You KT, Li LS, Kim NG, et al. Selective translational repression of truncated proteins from frameshift mutation-derived mRNAs in tumors. PLoS Biol. 2007;5:e109. https://doi.org/10.1371/journal.pbio.0050109.
29. Roudko V, Bozkus CC, Orfanelli T, et al. Shared immunogenic poly-epitope frameshift mutations in microsatellite unstable tumors. Cell. 2020;183(6):1634-1649.e17. https://doi.org/10.1016/j.cell.2020.11.004.
30. Trolle T, McMurtrey CP, Sidney J, et al. The length distribution of class i-restricted T Cell epitopes is determined by both peptide supply and MHC allele-specific binding preference. J Immunol. 2016;196(4):1480–7. https://doi.org/10.4049/jimmunol.1501721.

Publisher's Note