Journal of Translational Medicine

RESEARCH

Open Access

# Lack of shared neoantigens in prevalent mutations in cancer

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

## Abstract

Tumors are mostly characterized by genetic instability, as result of mutations in surveillance mechanisms, such as DNA damage checkpoint, DNA repair machinery and mitotic checkpoint. Defect in one or more of these mechanisms causes additive accumulation of mutations. Some of these mutations are drivers of transformation and are positively selected during the evolution of the cancer, giving a growth advantage on the cancer cells. If such mutations would result in mutated neoantigens, these could be actionable targets for cancer vaccines and/or adoptive cell therapies. However, the results of the present analysis show, for the first time, that the most prevalent mutations identified in human cancers do not express mutated neoantigens. The hypothesis is that this is the result of the selection operated by the immune system in the very early stages of tumor development. At that stage, the tumor cells characterized by mutations giving rise to highly antigenic non-self-mutated neoantigens would be efficiently targeted and eliminated. Consequently, the outgrowing tumor cells cannot be controlled by the immune system, with an ultimate growth advantage to form large tumors embedded in an immunosuppressive tumor microenvironment (TME). The outcome of such a negative selection operated by the immune system is that the development of off-the-shelf vaccines, based on shared mutated neoantigens, does not seem to be at hand. This finding represents the first demonstration of the key role of the immune system on shaping the tumor antigen presentation and the implication in the development of antitumor immunological strategies.

**Keywords** Mutations, Neoantigens, Tumor-associated antigens, Tumor-specific antigens, Cancer vaccines, Molecular mimicry, T cell immunity

## Introduction

Somatic mutations occur in the genomes of all normal and neoplastic dividing cells. They are the result of errors occurring during DNA replication as well as exposure to exogenous or endogenous mutagens. However, if most of these mutations are repaired by cellular mechanisms, a minority remains fixed in the cell genome. Most of such fixed mutations are biologically neutral and already present in the progenitor cell, before the transformation into the final clonal cancer cell ("passenger" mutations). The remaining ones are "driver" mutations that confer growth advantage on the cell, increasing survival or proliferation, and are selected. The accumulation of the driver mutations over the lifetime of an individual will induce cell transformation and cancer development [1–3]. The number of mutations required to drive a cancer significantly varies across tumor types [4]. Studies have shown that carcinogenesis may be driven by a small number of driver mutations. In particular, one driver mutation per patient is sufficient in sarcomas, thyroid, and testicular cancers; and about four driver mutations per patient are needed in bladder, endometrial, and colorectal cancers [1, 2, 5]. The different mutations in cancer cells show different

*Correspondence:
Maria Tagliamonte
m.tagliamonte@istitutotumori.na.it
Luigi Buonaguro
l.buonaguro@istitutotumori.na.it
[1] Lab of Innovative Immunological Models Unit, Istituto Nazionale Tumori, IRCCS - "Fondazione Pascale", Via Mariano Semmola, 52, 80131 Naples, Italy

rates. In particular, most cancers carry 1000 to 20,000 somatic point mutations and a few to hundreds of insertions, deletions, and rearrangements [1].

Such mutations in the genomic sequences of cancer cells may generate modified protein sequences, which may give rise to new epitopes unique to cancer cells. These mutated epitopes ("neoantigens") are tumor-specific non-self-antigens efficiently recognized by the immune system. Therefore, therapeutic vaccines based on such neoantigens would elicit a T cell immune response that can exclusively target the tumor while sparing healthy tissue [6]. The presence and biological relevance of the T cell immunity against neoantigens in cancer patients is demonstrated by the higher clinical efficacy of Immune checkpoint inhibitors (ICI) in tumors with high tumor mutational burden (TMB) [7–9] and with neoantigen-specific CD8 + T cells [10].

However, mutations and neoantigens are strictly individual (private) and their identification requires a combination of high throughput omics bioinformatics pipeline for each cancer patient, whose reliability has not been fully proven yet. Indeed, a comprehensive meta-analysis of the literature showed that only < 2.7% of prioritized predicted neoantigens are recognized by patient-derived T cells [11]. This has been further confirmed by the tumor neoantigens selection alliance (TESLA) global consortium [12]. Neoantigens were predicted with different pipelines by each participating member from the same tumor sequencing data but only approximately 6% of such predicted neoantigens were recognized by the T cells.

In addition to the complexity and reliability of the approaches, which appear highly difficult to be applied on a large scale, this strictly personalized strategy may fail due to the high mutational rate of tumors, which drives a constant generation of new target mutated neoantigens in the same patient. This would require subsequent rounds of neoantigens identification and vaccine production. More than 100 active or completed clinical trials are listed in clinicaltrials.gov when searching for the terms 'vaccine' and 'neoantigens', but a clear clinical benefit has not been demonstrated [13]. Only recently, an early phase trial in pancreatic cancer has generated a clinical benefit in terms of prolonged recurrence free survival (RFS) [14].

In this framework, it would be of the highest priority to identify mutated neoantigens, derived from the most frequent mutations and shared among cancer patients, to develop off-the-shelf cancer vaccines.

The results of the present study show that, indeed, such shared mutated neoantigens are not predicted for the most frequent cancer mutations (substitutions and insertion/deletion) in association to the most frequent HLA alleles. This would strongly suggest that only cancer cells lacking immunogenic tumor-specific non-self neoantigens, "poorly-visible" to the immune system, have a growth advantage and proliferate to generate clinically visible tumors. Therefore, off-the-shelf cancer vaccines based on shared mutated neoantigens have low chance to be a feasible strategy.

## Materials and methods

### Selection of cancer mutations from TCGA

The first 100 mutations reported at the TCGA database were selected for the study. Collectively, they represent 55.8% of all mutations identified in human cancers.

### Prediction of mutated neoantigens

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 NetMHCpan 4.1 algorithm (https://services.healthtech.dtu.dk/service.php?NetMHCpan-4.1) to predict the best nonamers with affinity values 0—400 nM to the 12 most frequent HLA-A and B alleles. Only those with an affinity value < 100 nM (strong binders – SB) were then selected for subsequent analyses.

### Homology search for neoantigens in literature

The mutated neoantigens, identified as SB according to the NetMHCpan 4.1 prediction tool, were submitted to the Immune Epitope Database & Tools (www.iedb.org) to verify whether the predicted epitopes have been already described and validated in literature. The analysis was performed setting the parameters to search for epitopes with exact match in any host.

### Statistical analysis

The statistical significance of the observed predicted neoantigens derived from either missense or InDel mutations, was calculated based on the observed predicted neoantigens in all samples at TCGA. The normal distribution was calculated as $Z = (X - \mu)/\sigma$, where X is the experimental result; μ is the mean value; σ is the standard deviation. P value was calculated as left-tailed. The confidence interval was calculated as $\mu \pm Z\frac{\sigma}{\sqrt{n}}$, where μ is the mean value; Z is the Z-score; σ is the standard deviation; n is the sample population.

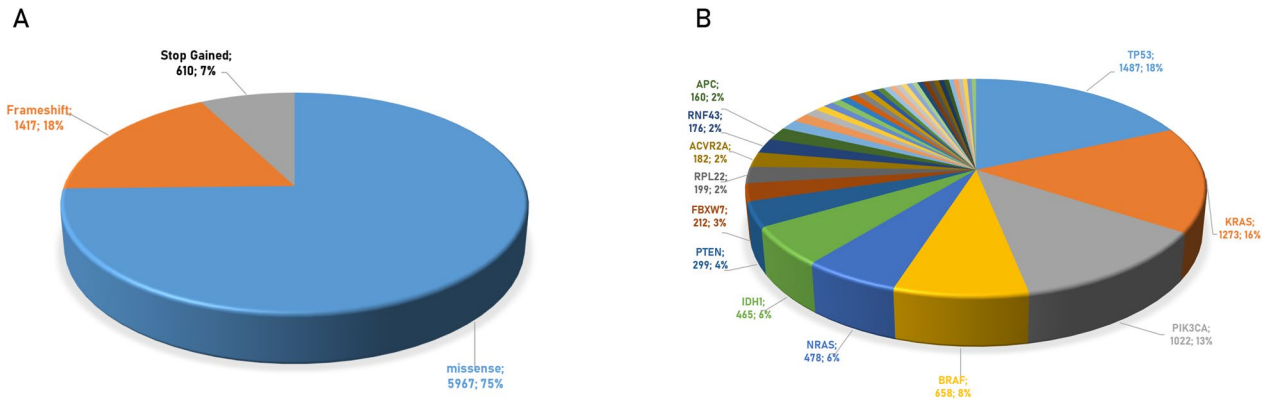


**Fig. 1** Top 100 mutations identified in cancers at TCGA database. **A** Percentage of type of mutations; **B** percentage of tumors presenting mutations of the indicated proteins

## Results

### Most frequent mutations in cancers

The total number of somatic mutations reported in the TCGA database is 190,632. They have been identified in 14,254 cancer cases. The most frequent 100 mutations occur in 8074 cases, which represent 56.65% of all cases, and the top frequent mutation is the $BRAF_{V640E/V600E}$, found in 619/14,254 cases (4.34%) (Additional file 1: Table S1).

Among these 100 hot-spot mutations, 62% are missense mutations identified in 5967 cases (73.9%) and 23% are frameshift mutations identified in 1417 cases (17.55%). In addition, 13% are stop-gained mutations identified in 610 cases (7.56%) (Fig. 1A). The TP53 protein is characterized by the highest number of different mutations (nr. 20), which cumulatively are identified in the highest number of cases (1487 of the 8074 cases, 18%) (Fig. 1B; Additional file 1: Table S2). Among the 100 hot-spot mutations are included all the hot-spot mutations

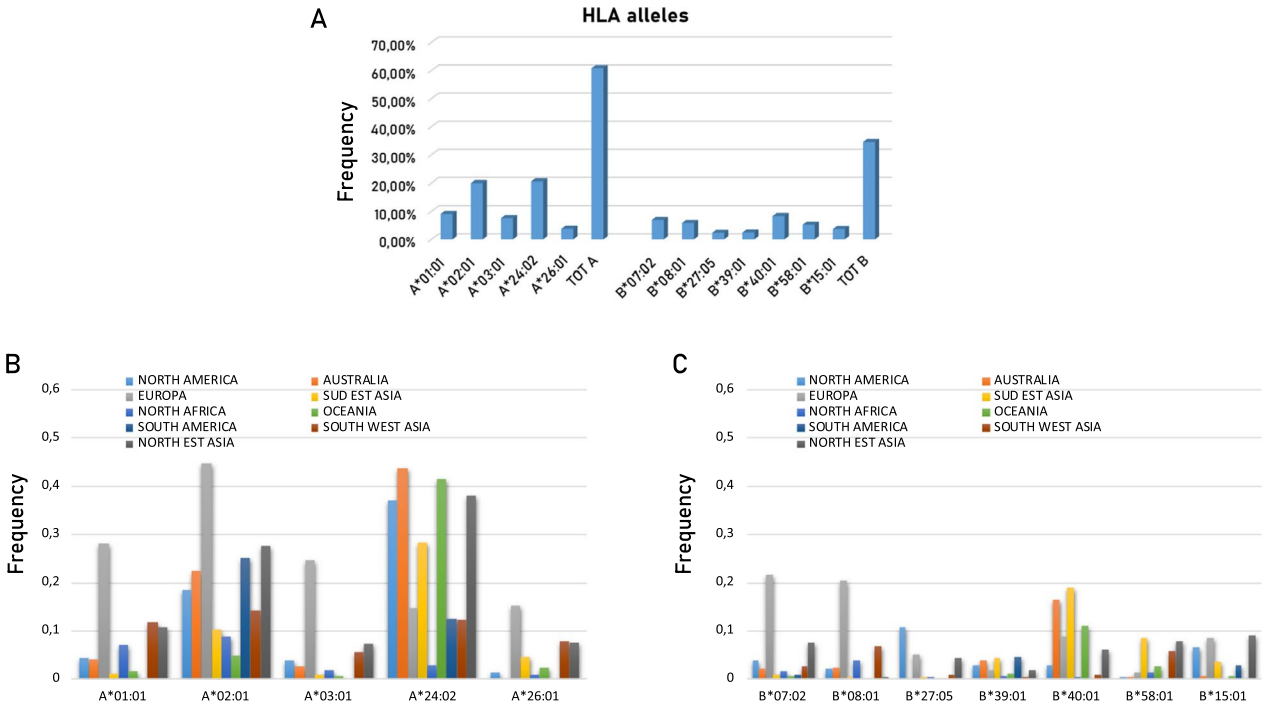


**Fig. 2** Frequency of HLA-A and B alleles considered in the study. **A** Frequency of individual alleles at global level; **B** and **C** Frequency of individual HLA-**A** and **B** alleles in each world population

**Table 1** Example of overlapping peptides from wt and missense mutated protein sequences for neo-epitope prediction. Mutated aminoacid residue is indicated in bold. In each overlapping peptide, the residue involved in the missense mutation is indicated in red

| | |
|---|---|
| $BRAF_{wt}$ | 592 – IGDFGLATVKSRWSGSH – 608 |
| $BRAF_{V600E}$ | 592 – IGDFGLATEKSRWSGSH – 608 |

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| BRAF WT | **V** | K | S | R | W | S | G | S | H |
| BRAF WT | T | **V** | K | S | R | W | S | G | S |
| BRAF WT | A | T | **V** | K | S | R | W | S | G |
| BRAF WT | L | A | T | **V** | K | S | R | W | S |
| BRAF WT | G | L | A | T | **V** | K | S | R | W |
| BRAF WT | F | G | L | A | T | **V** | K | S | R |
| BRAF WT | D | F | G | L | A | T | **V** | K | S |
| BRAF WT | G | D | F | G | L | A | T | **V** | K |
| BRAF WT | I | G | D | F | G | L | A | T | **V** |

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| $BRAF_{V600E}$ | **E** | K | S | R | W | S | G | S | H |
| $BRAF_{V600E}$ | T | **E** | K | S | R | W | S | G | S |
| $BRAF_{V600E}$ | A | T | **E** | K | S | R | W | S | G |
| $BRAF_{V600E}$ | L | A | T | **E** | K | S | R | W | S |
| $BRAF_{V600E}$ | G | L | A | T | **E** | K | S | R | W |
| $BRAF_{V600E}$ | F | G | L | A | T | **E** | K | S | R |
| $BRAF_{V600E}$ | D | F | G | L | A | T | **E** | K | S |
| $BRAF_{V600E}$ | G | D | F | G | L | A | T | **E** | K |
| $BRAF_{V600E}$ | I | G | D | F | G | L | A | T | **E** |

identified in each of the 51 primary cancer sites present in TCGA. The frequency of such hot-spot mutations in the different cancer sites is quite variable and broad, going from 3.10% of $TP53_{R175H}$ identified in the retro peritoneum ca to 61.52% of $BRAF_{V640E/V600E}$ in the thyroid ca. In particular, considering cancers with a high unmet clinical need (namely, < 20% 5 year overall survival — OS), the $IDH1_{R132H}$ is found in 37.65% of brain ca; the $KRAS_{G12D}$ is found in 32.87% of pancreas ca; the $ACVR2A_{K437Rfs*5}$ is found in 14.13% of stomach ca (Additional file 1: Table S2).

### Selection of HLA alleles for epitope prediction from top 100 mutations.

In the quest for such shared TSAs, the peptide sequences including each of the top 62 missense mutations or derived from each of the 23 InDel mutations were analyzed for prediction of epitope binding to MHC class I molecules. Such analysis was performed including the 12 most frequent HLA-A and B alleles that, collectively, cover 60% (HLA-A alleles) and 35% (HLA-B alleles) of the world population (Fig. 2A). In particular, HLA-A*02:01 is present in 44% of the European population and in more than 10% in all other populations, with exception of Southeast Asian, North African and Oceania. The HLA-A*24:02 is present more than 10% in all populations, with exception of North African. Among the HLA-B alleles, the B*07:02 and 08:01 alleles show in Europeans a high prevalence of 21.8% and 20.6%, respectively. Furthermore, the B*40:01 allele shows a high prevalence in Australians (16.4%) and Southeast Asians (19.1%). All other

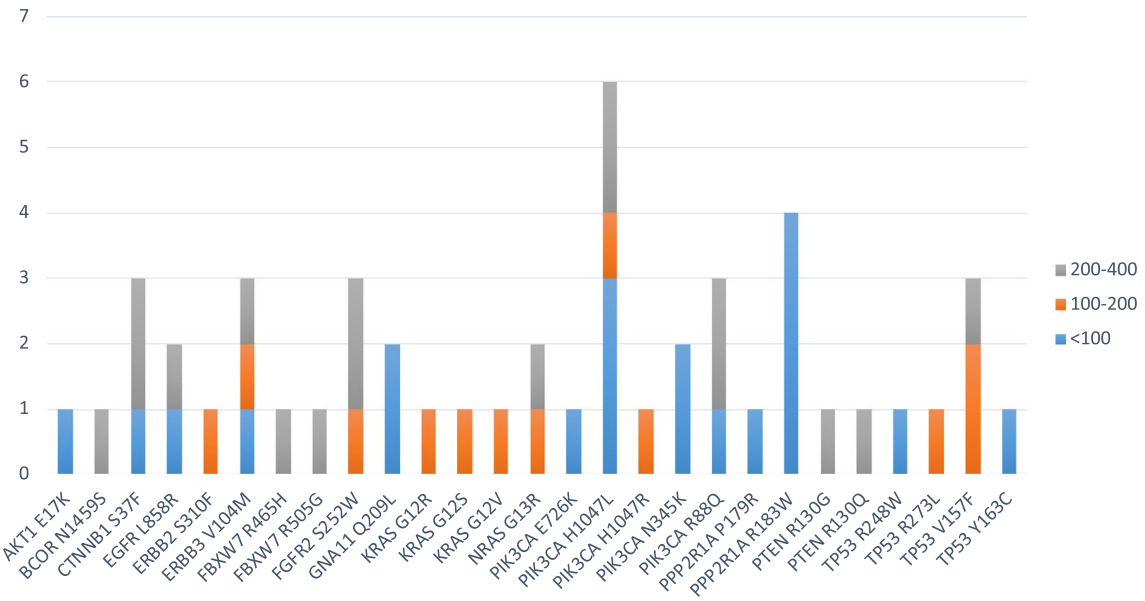


**Fig. 3** Number of predicted neoantigens from missense mutations. The number of predicted neoantigens for each missense mutations are reported. The predicted affinity of such neoantigens, expressed in nM, is indicated with color-code

**Table 2** Predicted neo-epitopes derived from missense mutations with an affinity value to the HLA alleles < 100 nM (green highlighted).

| PROTEIN | PEPTIDE | A*01:01 | A*02:01 | A*03:01 | A*24:02 | A*26:02 | B*07:02 | B*08:02 | B*27:05 | B*39:01 | B*40:01 | B*58:01 | B*15:01 | PASS | frequency |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $AKT1_{17wt}$ | WLHKRGEYI | 23091,35 | 1023,70 | 26483,53 | 14709,45 | 22904,74 | 10068,68 | 75,55 | 13611,61 | 11392,17 | 24529,23 | 22719,12 | 4002,82 | | |
| $AKT1_{E17K}$ | WLHKRG**K**YI | 30311,99 | 8799,75 | 27066,69 | 24635,88 | 26877,00 | 7869,28 | 39,58 | 14807,65 | 20770,70 | 30641,08 | 31130,31 | 6700,30 | N | |
| $CTNNB1_{37wt}$ | SYLDSGIHS | 42222,49 | 31730,85 | 34347,66 | 17184,22 | 41225,75 | 38486,34 | 39521,47 | 34683,75 | 34874,91 | 40663,58 | 38734,07 | 34477,59 | | |
| $CTNNB1_{S37F}$ | SYLDSGIH**F** | 34152,35 | 23546,25 | 30205,58 | 44,06 | 24053,26 | 26536,89 | 27547,07 | 19749,20 | 19610,36 | 27054,38 | 16511,66 | 6835,77 | Y | 0.27% |
| $EGFR_{858wt}$ | KITDFGLAK | 27224,70 | 19690,52 | 31,68 | 37442,23 | 34662,73 | 25403,55 | 37308,39 | 14502,27 | 39272,55 | 37522,54 | 24009,32 | 18459,04 | | |
| $EGFR_{L858R}$ | KITDFG**R**AK | 37166,17 | 28344,68 | 72,17 | 41075,70 | 36202,89 | 23573,78 | 36624,86 | 19028,25 | 41784,39 | 39711,38 | 30819,29 | 19714,81 | N | |
| $ERBB3_{104wt}$ | RVVRGTQVY | 10326,97 | 30672,60 | 1124,64 | 27620,50 | 3764,07 | 10454,12 | 27289,88 | 13340,85 | 28536,38 | 26377,45 | 460,74 | 24,00 | | |
| $ERBB3_{V104M}$ | R**M**VRGTQVY | 8985,44 | 17179,77 | 305,54 | 13934,83 | 11096,56 | 14600,20 | 19161,52 | 4636,22 | 17751,17 | 16665,84 | 707,33 | 6,05 | N | |
| $GNA11_{209wt}$ | FRMVDVGGQ | 43235,86 | 39347,39 | 41085,93 | 44289,34 | 37012,06 | 38910,05 | 30614,24 | 2659,46 | 21669,44 | 40332,75 | 37297,49 | 33616,37 | | |
| $GNA11_{Q209L}$ | FRMVDVGG**L** | 33485,68 | 14356,45 | 37259,17 | 27555,14 | 22164,08 | 18486,82 | 5197,23 | 99,13 | 92,24 | 16113,33 | 28513,24 | 16842,76 | Y | 0.26% |
| $PIK3CA_{345wt}$ | ILCATYVNV | 25354,95 | 36,09 | 12666,91 | 17502,66 | 33496,18 | 35158,30 | 19488,95 | 29229,17 | 28996,40 | 32642,52 | 17391,47 | 7209,36 | | |
| $PIK3CA_{N345K}$ | ILCATYV**K**V | 25665,45 | 80,57 | 15146,48 | 17958,45 | 34968,62 | 36385,11 | 20524,52 | 29801,43 | 30593,71 | 33246,31 | 19173,12 | 9870,21 | N | |
| $PIK3CA_{345wt}$ | KILCATYVN | 36722,88 | 20845,23 | 16792,73 | 29390,9 | 41961,54 | 32509,64 | 31764,51 | 29288,68 | 38567,61 | 39202,08 | 5395,25 | 18471,02 | | |
| $PIK3CA_{N345K}$ | KILCATYV**K** | 32807,87 | 18294,80 | 39,61 | 24155,23 | 38363,28 | 28598,50 | 28996,40 | 11823,16 | 37563,96 | 37830,32 | 14049,58 | 20325,62 | Y | 0.29% |
| $PIK3CA_{726wt}$ | ETQKVQMKF | 9158,29 | 38289,06 | 30151,71 | 8364,17 | 374,18 | 27899,23 | 20483,48 | 28553,69 | 32705,1 | 34535,07 | 1757,64 | 10158,52 | | |
| $PIK3CA_{E726K}$ | **K**TQKVQMKF | 12879,61 | 25829,26 | 9939,87 | 953,74 | 18473,02 | 15638,41 | 25134,80 | 13142,42 | 36210,34 | 28957,83 | 15,89 | 1378,42 | Y | 0.21% |
| $PIK3CA_{1047wt}$ | FMKQMNDAH | 25480,9 | 23347,11 | 18678,43 | 39240,68 | 22569,18 | 19605,06 | 6249,72 | 19414,23 | 20286,29 | 30118,12 | 31352,38 | 331,41 | | |
| $PIK3CA_{H1047L}$ | FMKQMNDA**L** | 22962,79 | 89,17 | 31815,76 | 13261,70 | 22130,54 | 2713,52 | 40,19 | 12360,79 | 285,55 | 10406,38 | 20578,56 | 102,31 | Y | 0.31% |
| $PIK3CA_{1047wt}$ | AHHGGWTTK | 39036,15 | 34641,36 | 9840,68 | 29029,98 | 37188,28 | 30055,62 | 36595,54 | 10216,72 | 16893,14 | 33532,45 | 33219,72 | 27708,8 | | |
| $PIK3CA_{H1047L}$ | A**L**HGGWTTK | 32357,68 | 15050,91 | 35,26 | 32319,89 | 32272,73 | 25572,04 | 30715,77 | 12768,32 | 37186,67 | 34940,63 | 29501,46 | 12316,61 | Y | 0.31% |
| $PIK3CA_{88wt}$ | RRLCDLRLF | 34839,46 | 28033,88 | 29701,64 | 5600,69 | 29964,7 | 24519,41 | 17337,37 | 19,07 | 15073,41 | 21176,21 | 11340,28 | 12012,84 | | |
| $PIK3CA_{R88Q}$ | R**Q**LCDLRLF | 29147,70 | 6147,05 | 17440,66 | 495,74 | 19641,57 | 23491,80 | 17878,76 | 398,81 | 13721,18 | 3180,60 | 1660,39 | 47,05 | Y | 0.58% |
| $PPP2R1A_{179wt}$ | TPMVRRAAA | 33891,74 | 31684,18 | 28009,02 | 38971,58 | 30553,02 | 10,52 | 19,12 | 21391,99 | 4409,60 | 34481,69 | 38549,26 | 27721,69 | | |
| $PPP2R1A_{P179R}$ | T**R**MVRRAAA | 37500,21 | 35012,54 | 31721,93 | 36070,36 | 35845,09 | 6000,38 | 94,11 | 662,45 | 2045,57 | 33464,31 | 39796,11 | 26219,51 | N | |
| $PPP2R1A_{183wt}$ | MVRRAAASK | 31275,82 | 28904,61 | 44,02 | 34680,37 | 17042,02 | 4317,58 | 10245,17 | 11747,55 | 37220,09 | 35742,07 | 20709,44 | 6832,73 | | |
| $PPP2R1A_{R183W}$ | MVR**W**AAASK | 28957,21 | 20984,40 | 30,24 | 30629,15 | 11215,45 | 4154,21 | 12673,08 | 11889,49 | 33405,70 | 32240,96 | 17508,16 | 4930,68 | N | |
| $PPP2R1A_{183wt}$ | TPMVRRAAA | 33891,74 | 31684,18 | 28009,02 | 38971,58 | 30553,02 | 10,52 | 19,12 | 21391,99 | 4409,60 | 34481,69 | 38549,26 | 27721,69 | | |
| $PPP2R1A_{R183W}$ | TPMVR**W**AAA | 27902,55 | 19244,83 | 23737,32 | 32026,09 | 20292,22 | 16,76 | 24,34 | 16596,56 | 2449,86 | 27977,22 | 32201,57 | 22924,81 | N | |
| $TP53_{248wt}$ | SSCMGGMNR | 21319,21 | 34028,79 | 3897,27 | 37889,3 | 31477,46 | 36365,03 | 37245,07 | 25265,66 | 40209 | 41168,7 | 23587,8 | 29458,4 | | |
| $TP53_{R248W}$ | SSCMGGMN**W** | 13361,21 | 35375,80 | 23490,53 | 11386,50 | 24675,63 | 30402,97 | 31584,93 | 27410,62 | 34487,67 | 33845,19 | 14,01 | 6103,58 | Y | 0.79% |
| $TP53_{163wt}$ | RVRAMAIYK | 28640 | 23964,17 | 7,61 | 25419,76 | 33270,45 | 12084,8 | 21645,3 | 3539,52 | 41726,12 | 36241,29 | 16983,49 | 12469,19 | | |
| $TP53_{Y163C}$ | RVRAMAI**C**K | 31135,69 | 29683,96 | 21,08 | 29777,57 | 36972,84 | 13643,31 | 21100,97 | 6913,79 | 41925,67 | 37229,36 | 20495,67 | 14990,95 | N | |

The neo-epitopes pass the validation only when the corresponding wt epitope is a poor binder. The frequency of the validated neo-epitopes in the TCGA database is indicated. Identity to peptides in iedb is indicated

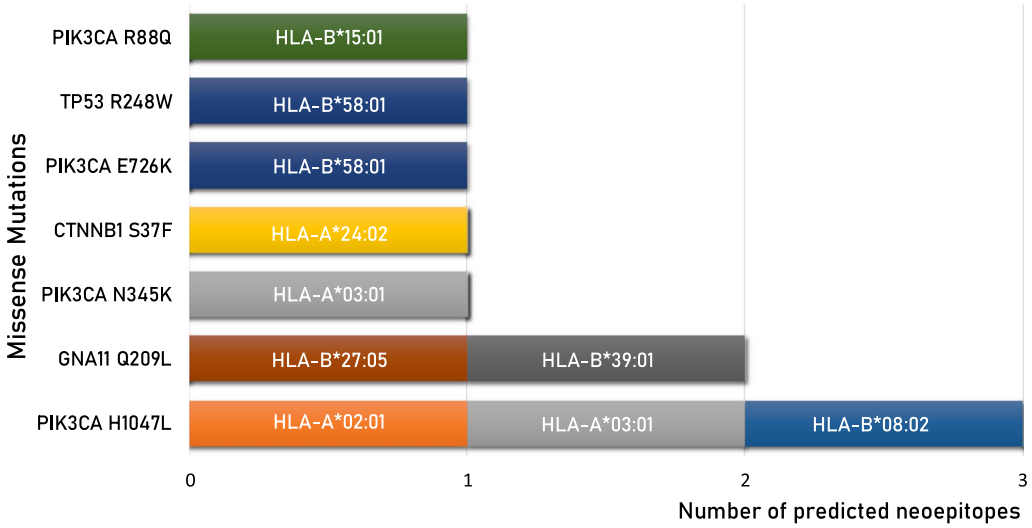


**Fig. 4** High-affinity predicted neoantigens from missense mutations and HLA restriction. The number of predicted neoantigens for each missense mutations are reported with indication of the HLA restriction

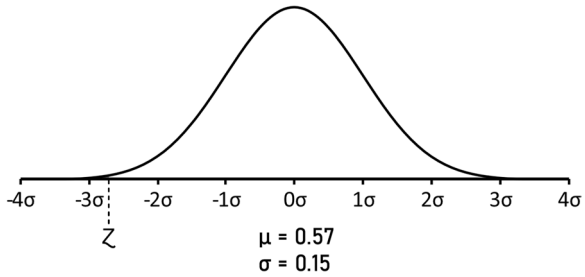


**Fig. 5** Z-score of the observed predicted neoantigens from the hot-spot missense mutations. The normal distribution of the percentage of predicted neoantigens from the 8547 samples present at TCGA. The Z-score of the observed predicted neoantigens from the hot-spot missense mutations is indicated. The result shows a statistically significant lower percentage than what expected (p-value = 0.006; 99.37% confidence level)

HLA-A and B alleles show low prevalence (< 10%) across populations (Fig. 2B,C).

**Neoantigen prediction from the missense mutations.**

In order to predict neoantigens from proteins with a single amino acid missense mutation, the amino acid sequence was downloaded from UniProt for each of the 62 proteins. A 17mer peptide was selected, centered around the mutated residue (from − 8 to + 8), and overlapping peptides were designed with the mutated residue at each of the 9 positions (Table 1). The wt and mutated peptides were subjected to the prediction analysis, to assess the affinity to the 12 HLA-A and B alleles. The results on the 945 peptides analyzed showed that only 49 mutated peptides (neoantigens) (5.18%) have an affinity < 400 nM (Fig. 3; Additional file 1: Table S3).

However, only 20 (2.11%) have an affinity value to the HLA alleles < 100 nM and only 10 (1.06%) can be considered optimal neoantigens. Indeed, only for these, the corresponding wt-epitope shows very low affinity values to the HLA alleles (102–41,900 nM) and are not antigenic (Table 2). Six of such neoantigens are strong binders to a single HLA allele; one epitope ($GNA11_{Q209L}$ FRMVDVGG**L**) is a strong binder to two HLA alleles (B*27:05 and B*39:01); two epitopes derived from the same $PIK3CA_{H1047L}$ mutation and are strong binders to three HLA alleles (FMKQMNDA**L,** A*02:01 and B*08:02; A**L**HGGWTTK, A*03:01) (Fig. 4).

In order to verify the statistical significance of the observed low number of predicted neoantigens, we have considered all the mutations generating predicted epitopes in the 8547 samples present at TCGA (https://gdc.cancer.gov/about-data/publications/panimmune). Overall, 56.86% of all 1,327,063 missense mutations generate predicted epitopes. On the contrary, the 62 hot spot missense mutations analyzed in the present study generate only 10 mutations (16.13%). Therefore, the number

**Table 3** Predicted neo-epitopes with an affinity value to the HLA alleles < 100 nM, derived from missense mutations, are listed with selected information

| MISSENSE | | | TOT FREQ | TOP FREQ | TUMOR |
|---|---|---|---|---|---|
| **PROTEIN** | **PEPTIDE** | **A*02:01** | | | |
| $PIK3CA_{H1047L}$ | FMKQMNDA**L** | 89.17 | 0.31% | N/A | N/A |
| **PROTEIN** | **PEPTIDE** | **A*03:01** | | | |
| $PIK3CA_{N345K}$ | KILCATYV**K** | 39.61 | 0.29% | N/A | N/A |
| $PIK3CA_{H1047L}$ | A**L**HGGWTTK | 35.26 | 0.31% | N/A | N/A |
| **PROTEIN** | **PEPTIDE** | **A*24:02** | | | |
| $CTNNB1_{S37F}$ | SYLDSGIH**F** | 44.06 | 0.27% | N/A | N/A |
| **PROTEIN** | **PEPTIDE** | **B*08:02** | | | |
| $PIK3CA_{H1047L}$ | FMKQMNDA**L** | 40.19 | 0.31% | N/A | N/A |
| **PROTEIN** | **PEPTIDE** | **B*27:05** | | | |
| $GNA11_{Q209L}$ | FRMVDVGG**L** | 99.13 | 0.42% | 42.50% | Eye & adnexa |
| **PROTEIN** | **PEPTIDE** | **B*39:01** | | | |
| $GNA11_{Q209L}$ | FRMVDVGG**L** | 92.24 | 0.42% | 42.50% | Eye & adnexa |
| **PROTEIN** | **PEPTIDE** | **B*58:01** | | | |
| $PIK3CA_{E726K}$ | **K**TQKVQMKF | 15.89 | 0.21% | N/A | N/A |
| $TP53_{R248W}$ | SSCMGGMN**W** | 14.01 | 1.05% | N/A | N/A |
| **PROTEIN** | **PEPTIDE** | **B*15:01** | | | |
| $PIK3CA_{R88Q}$ | R**Q**LCDLRLF | 47.05 | 0.57% | N/A | N/A |

The peptide sequences include the mutated aminoacid residue (bold & underlined). Values in the column of the haplotypes indicate the predicted affinity (nM). TOT FREQ: frequency in the TCGA database; TOP FREQ: top frequency in specific tumor; TUMOR: tumor type in which the top frequency is reported

**Table 4** Example of overlapping peptides from wt and frameshift mutated protein sequences for neo-epitope prediction.

| | |
|---|---|
| $JAK1_{wt}$ | NPDIVSEK**K**PATEVDPT |
| $JAK1_{K860Nfs\_1}$ | NPDIVSEK**N***QQLKWTPHILKSAS* |

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| WT | **K** | P | A | T | E | V | D | P | T |
| WT | K | **K** | P | A | T | E | V | D | P |
| WT | E | K | **K** | P | A | T | E | V | D |
| WT | S | E | K | **K** | P | A | T | E | V |
| WT | V | S | E | K | **K** | P | A | T | E |
| WT | I | V | S | E | K | **K** | P | A | T |
| WT | D | I | V | S | E | K | **K** | P | A |
| WT | P | D | I | V | S | E | K | **K** | P |
| WT | N | P | D | I | V | S | E | K | **K** |

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| K860Nfs_1 | T | P | H | I | L | K | S | A | S |
| K860Nfs_1 | W | T | P | H | I | L | K | S | A |
| K860Nfs_1 | K | W | T | P | H | I | L | K | S |
| K860Nfs_1 | L | K | W | T | P | H | I | L | K |
| K860Nfs_1 | Q | L | K | W | T | P | H | I | L |
| K860Nfs_1 | Q | Q | L | K | W | T | P | H | I |
| K860Nfs_1 | **N** | Q | Q | L | K | W | T | P | H |
| K860Nfs_1 | K | **N** | Q | Q | L | K | W | T | P |
| K860Nfs_1 | E | K | **N** | Q | Q | L | K | W | T |
| K860Nfs_1 | S | E | K | **N** | Q | Q | L | K | W |
| K860Nfs_1 | V | S | E | K | **N** | Q | Q | L | K |
| K860Nfs_1 | I | V | S | E | K | **N** | Q | Q | L |
| K860Nfs_1 | D | I | V | S | E | K | **N** | Q | Q |
| K860Nfs_1 | P | D | I | V | S | E | K | **N** | Q |
| K860Nfs_1 | N | P | D | I | V | S | E | K | **N** |

Mutated aminoacid residue is indicated in bold and the downstream sequence from the alternative reading frame is indicated in italics. In each overlapping peptide in the wt and mutated sequence, the mutated residue is indicated in red

of observed mutations is significantly lower than what expected, with a p-value = 0.006 and a 99.37% confidence level (Fig. 5).

No neoantigens were predicted for HLA-A*01:01, A*26:02, B*07:02 and B*40:01; one neoantigen was predicted for HLA-A*02.01, A*24:02, B*08:02, B*27:05, B*39:01 and B*15:01. Only for HLA-A*03:01 and B*58:01 were predicted more than a single neoantigen, three and two respectively.

The Immune Epitope Database & Tools (www.iedb.org) was interrogated in order to verify whether the predicted epitopes have been already described and validated in literature. The search returned only three peptides, the $CTNNB1_{S37F}$ SYLDSGIH**F** peptide (PMID: 8642260; PMID:35122353), the $PIK3CA_{H1047L}$ A**L**HGGWTTK peptide (PMID: 35484264; PMID: 37415627) and the $PIK3CA_{R88Q}$ R**Q**LCDLRLF peptide (PMID: 37415627). The first two are confirmed to be restricted to HLA-A*24:02 and HLA-A*03:01, respectively. On the contrary, a discordance is observed for the $PIK3CA_{R88Q}$ peptide, which has been reported as restricted to HLA-A*24:02 while our analysis predicted a very strong binding to HLA-B*15:01 (14.01 nM) and a low binding to HLA-A*24:02 (495.74 nM) (Table 2).

All the predicted neoantigens are identified in mutations identified in a very low percentage of tumor samples, ranging from 0.21% ($PIK3CA_{E726K}$ **K**TQKVQMKF) to 0.79% ($TP53_{R248W}$ SSCMGGMN**W**). Only the $GNA11_{Q209L}$ FRMVDVGG**L** epitope, restricted to HLA-B*27:05 and B*39:01, is the most frequent mutation in uveal melanoma (42,50% of all cases reported in TCGA) (Additional file 2: Fig S1) (Table 3).

#### Neoantigen prediction from the frameshift mutations.

Similarly, neoantigen predictions from the 23 proteins with a frameshift mutation were carried out. The amino acid sequence was downloaded from UniProt for each of the proteins but, in this case, the selection of peptides for neoantigen prediction was different for the wt and mutated sequences. Indeed, as for the missense mutations, the prediction of wt-peptides was based on a 17mer peptide, centered around the mutated residue (from − 8 to + 8), and overlapping peptides were designed with the mutated residue at each of the 9 positions (Table 4). On the contrary, the prediction of the mutated-peptides was based on a sequence starting at position − 8 from the mutated amino acid residue and including the entire downstream protein sequence. The number of mutated peptides ranged from 4 to 62, according to the position of the newly generated stop codon along the shifted reading frame. The wt and mutated peptides were subjected to the prediction analysis, to assess the affinity to the 12 HLA-A and B alleles. The results on the 686 peptides analyzed showed that 103 mutated peptides (neoantigens) (15.01%) have an affinity < 400 nM (Fig. 6; Additional file 1: Table S4).

Of these, 40 have an affinity value to the HLA alleles < 100 nM (5.83%) and only 9 (1.31%) include the mutated residue from which the frameshift starts (Table 5). The remaining 31 mutated epitopes cover the new sequence generated by the alternative open

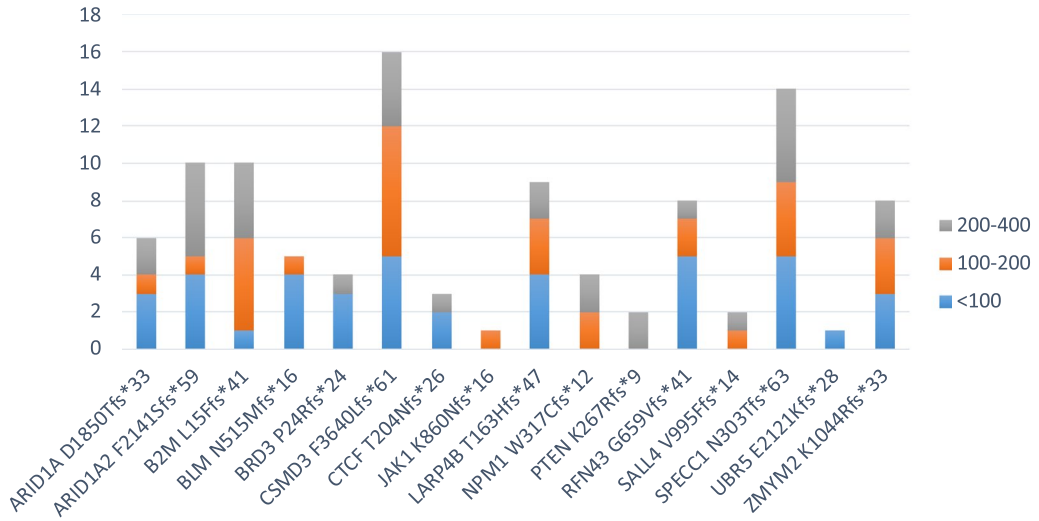


**Fig. 6** Number of predicted neoantigens from frameshift mutations. The number of predicted neoantigens for each frameshift mutations are reported. The predicted affinity of such neoantigens, expressed in nM, is indicated with color-code

reading frame. All of them can be considered optimal neoantigens given that the corresponding wt-epitopes either show very low affinity values to the HLA alleles (> 1000 nM), and are not antigenic, or are a completely different sequence and cannot be considered a "corresponding" epitope (Table 5). Only two of such neoantigens are strong binders to more than a single HLA allele: ($RFN43_{G659Vfs*41}$ TQLARFFPI) is a strong binder to three HLA alleles (A*02:01, B*08:02 and B*39:01); ($ARID1A_{D1850Tfs*33}$ WRIGGGTPL) is a strong binder to two HLA alleles (B*27:05 and B*39:01). All other epitopes are strong binders to a single HLA allele (Fig. 7A).

However, the "abnormal" mRNAs generated by the frameshift contain premature termination codons (PTCs), which are recognized and degraded by nonsense-mediated mRNA decay (NMD). [15, 16] Moreover, even when PTC-containing mRNAs escape from NMD, truncated proteins are not generated due to a translational repression [17]. Therefore, these epitopes have very low or no real chance to be presented by cancer cells, implying that only 9 neoantigens (1.31%) derived from InDels could be taken into consideration (Fig. 7B).

Indeed, the 23 hot spot InDel mutations analyzed in the present study generate a total of 40 predicted neoantigens (1.74 per InDel), which falls in the normal distribution of the expected values derived from the 6610 samples at TCGA with a confidence level of 99.99% (Fig. 8).

No neoantigens were predicted for HLA-A*01:01, A*02:01, A*03:01, A*24:02, A*26:02, B*40:01 and B*15:01. The HLA alleles with predicted neoantigens were HLA-B*07:02, B*08:02, B*27:05, B*39:01 and B*58:01.

None of the predicted epitopes derived from the frameshift mutations were found in the Immune Epitope Database & Tools (www.iedb.org), indicating that they have not been already described and validated in literature. Moreover, all the predicted neoantigens are identified in a very low percentage of tumor samples, ranging from 0.22% ($ARID1A2_{F2141Sfs*59}$ WLRGTAWQL, VPLQCRRAV and LATPPSAAW; $BLM_{N515Mfs*16}$ MKALISQEM, SQEMFSQAL, QEMFSQALL and KALISQEMF); to 1.23% ($RFN43_{G659Vfs*41}$ TQLARFFPI) (Table 6) (Additional file 1: Fig S2).

**HLA polymorphism and neoantigen prediction in cancers.**

The polymorphism of the HLA molecules taken into consideration in the present study greatly influences the array of peptides binding the HLA pocket.

Considering the missense mutations, the HLA-A*03:01 and B*58:01 alleles are predicted to bind and present 3 and 2 mutated neoantigens, respectively. The HLA-A*01:01, A*26:02, B*07:02 and B*40:01 alleles do not bind and present any mutated neoantigens. The remaining ones bind and present a single mutated neoantigen (Fig. 9A). Considering the frameshift mutations, the HLA-B*58:01 binds and presents 3 mutated neoantigens while the HLA-A*01:01, A*02:01, A*03:01,

**Table 5** Predicted neo-epitopes derived from frameshift mutations with an affinity value to the HLA alleles < 100 nM (green highlighted)

| PROTEIN | PEPTIDE | A*01:01 | A*02:01 | A*03:01 | A*24:02 | A*26:02 | B*07:02 | B*08:02 | B*27:05 | B*39:01 | B*40:01 | B*58:01 | B*15:01 | PASS |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| RNF43$_{659Wwt}$ | HPQRKRRG**G** | 44963,38 | 47423,84 | 43492,50 | 46303,32 | 45410,26 | 2570,48 | 3106,32 | 33262,16 | 38811,26 | 46667,95 | 44150,12 | 43589,07 | |
| RFN43$_{G659Vfs*41}$ | HPQRKRRG**V** | 39833,57 | 42190,54 | 39904,33 | 42817,33 | 38548,44 | 73,06 | 160,65 | 27511,63 | 19359,49 | 41886,24 | 42010,14 | 41463,75 | Y |
| RNF43$_{659Wwt}$ | **G**PSEPTPGS | 43027,24 | 41481,24 | 41074,38 | 47136,85 | 42298,86 | 8346,36 | 38918,06 | 41612,50 | 35229,14 | 40540,12 | 38908,38 | 41718,88 | |
| RFN43$_{G659Vfs*41}$ | **V***PPSPPLAL* | 35640,52 | 27552,15 | 34270,81 | 27048,82 | 31352,72 | 84,22 | 9163,54 | 34679,25 | 2734,68 | 24715,97 | 31419,27 | 27713,89 | Y |
| RFN43$_{G659Vfs*41}$ | *TQLARFFPI* | 20921,15 | 35,39 | 12274,17 | 429,33 | 19908,60 | 9730,88 | 48,47 | 561,27 | 87,73 | 1563,53 | 8403,09 | 449,87 | Y |
| UBR5$_{E2121Kfs*28}$ | *VQNQGHLLM* | 15941,7 | 7510,65 | 16071,19 | 8562,97 | 24230,62 | 13912,22 | 10767,75 | 5315,29 | 932,08 | 2256,74 | 7861,79 | 37,86 | Y |
| LARP4B$_{163wt}$ | VLKK**T**LEFC | 38160,8 | 5506,41 | 26725,33 | 32239,9 | 40718,16 | 34775,06 | 13611,76 | 35294,38 | 41055,3 | 40380,35 | 29582,33 | 15311,4 | |
| LARP4B$_{T163Hfs*47}$ | VLKK**H***WNSA* | 30030,9 | 5226,21 | 14160,07 | 26932,6 | 26226,9 | 6484,76 | 59,58 | 21503,61 | 26517,1 | 34760,4 | 34033,93 | 3090,36 | Y |
| LARP4B$_{T163Hfs*47}$ | *VTCILYHRW* | 14369,7 | 31471,33 | 25539,42 | 3274,92 | 27506,28 | 39284,03 | 29543,63 | 28032,67 | 41011,3 | 39219,88 | 12,22 | 19106,2 | Y |
| LARP4B$_{T163Hfs*47}$ | *YHRWIVTSM* | 27486,6 | 21328,21 | 27782,05 | 9453,38 | 17416,71 | 3029,32 | 2397,29 | 6956,11 | 71,78 | 15675,85 | 23967,54 | 3411,6 | Y |
| LARP4B$_{T163Hfs*47}$ | *TSMCQSQRW* | 13234,7 | 36351,27 | 29813,37 | 7442,94 | 18940,96 | 26276,9 | 21594,54 | 23793,12 | 33376,4 | 35661,72 | 10,43 | 10467,1 | Y |
| SPECC1$_{N303Tfs*63}$ | *HQAPQVAML* | 30751 | 1507,82 | 27522,35 | 15258,83 | 10767,75 | 9257,82 | 7887,26 | 7090 | 80,74 | 1528,14 | 17939,61 | 181,9 | Y |
| SPECC1$_{N303Tfs*63}$ | *QMLPTLSTL* | 33515,8 | 69,86 | 17421,61 | 5100,85 | 21890,24 | 7399,42 | 4421,92 | 12907,36 | 1868,46 | 13110,89 | 3917,56 | 259,59 | Y |
| SPECC1$_{N303Tfs*63}$ | *TPLRVQSVL* | 35867,6 | 31516,66 | 35150,32 | 33352,25 | 31727,41 | 26,51 | 1311,98 | 28466,07 | 1412,66 | 25361,26 | 22527,95 | 27254,8 | Y |
| SPECC1$_{N303Tfs*63}$ | *LRVQSVLLL* | 32094,4 | 21288,09 | 32058,34 | 18220,71 | 29443,75 | 23813,99 | 10944,53 | 46,92 | 223,95 | 15991,06 | 18453,44 | 16544,2 | Y |
| SPECC1$_{N303Tfs*63}$ | *LLLGVPQTA* | 30989,8 | 72,97 | 22955,34 | 35114,21 | 37978,38 | 26596,1 | 14334,88 | 26576,83 | 18500,6 | 31316,45 | 22885,9 | 8993,32 | Y |
| ARID1A$_{1850wt}$ | WRIGGG**D**TT | 42057,5 | 38113,79 | 44153,45 | 41690,02 | 36897,71 | 31857,44 | 30081,01 | 4367,24 | 1333,07 | 27622,89 | 39344,85 | 32946,6 | |
| ARID1A$_{D1850Tfs*33}$ | WRIGGG**T***PL* | 33262,9 | 14204,88 | 33402,81 | 23652,21 | 16911,61 | 4634,47 | 3507,53 | 51,15 | 6,08 | 3835,77 | 29509,75 | 5617,81 | Y |
| ARID1A$_{D1850Tfs*33}$ | *CLPGLTHPA* | 21540,4 | 94,71 | 24117,89 | 19089,29 | 15016,1 | 18130,65 | 9974,46 | 25225,25 | 12815,7 | 31471,33 | 32775,22 | 7590,47 | Y |
| CTCF$_{T204Nfs*26}$ | *NQKEQTALY* | 9354,88 | 33891,36 | 16460,46 | 29705,16 | 2770,99 | 34042,78 | 19198,24 | 17214,56 | 23179,2 | 21504,08 | 28402,99 | 58,38 | Y |
| CTCF$_{T204Nfs*26}$ | *QRCRCVCLR* | 36903,7 | 34797,63 | 22424,11 | 34456,32 | 38704,33 | 40297,42 | 31570,92 | 94,17 | 32247,3 | 40397,39 | 33857,29 | 38268,4 | Y |
| CSMD3$_{F3640Lfs*61}$ | *FLHLYLQDL* | 25302,6 | 34,8 | 31449,88 | 21888,35 | 23408,06 | 16595,48 | 214,57 | 3463,71 | 21527,6 | 5356,62 | 25624,11 | 31899,2 | Y |
| CSMD3$_{F3640Lfs*61}$ | *YLQDLDFIF* | 4977,9 | 61,21 | 24088,16 | 284,22 | 12343,42 | 18992,67 | 5753,82 | 105,48 | 8020,48 | 3955,42 | 9411,74 | 2225,99 | Y |
| CSMD3$_{F3640Lfs*61}$ | *SIQDVQFMK* | 23370,1 | 21553,47 | 89,24 | 31466,24 | 28839,33 | 35425,98 | 36299,4 | 27598,69 | 21073,1 | 39249,61 | 39395,53 | 25317,7 | Y |
| CSMD3$_{F3640Lfs*61}$ | *ITMAKQLLK* | 7748,88 | 22712,74 | 9,19 | 23341,54 | 27126,78 | 29749,87 | 20557,19 | 22085,33 | 9741,62 | 38770,96 | 38642,8 | 6751,89 | Y |
| CSMD3$_{F3640Lfs*61}$ | *CMTPTQSQW* | 23469,2 | 18926,62 | 22067,65 | 3424,25 | 22835,2 | 26377,73 | 23620,5 | 2285,61 | 21837,3 | 30404,61 | 32315,7 | 31,13 | Y |
| B2M$_{15wt}$ | AVLALLS**L**S | 33144 | 7768,27 | 12419,11 | 39422,83 | 27134,71 | 31004,58 | 31474,74 | 24480,96 | 39973,9 | 35457,43 | 16723,65 | 14974,9 | |
| B2M$_{L15Ffs*41}$ | AVLALLS**F***W* | 28532,1 | 17037,23 | 14041,83 | 7605,84 | 14377,44 | 32131,97 | 27812,12 | 23003,58 | 40835,5 | 35299,34 | 32,16 | 5276,1 | Y |
| BRD3$_{P24Rfs*24}$ | *SPTPASPAA* | 36619,7 | 30880,04 | 32935,55 | 42861,84 | 33705,59 | 82,17 | 13525,58 | 35074,33 | 5050 | 32530,4 | 33767,64 | 30248,1 | Y |
| BRD3$_{P24Rfs*24}$ | *SPAARPTSC* | 39338 | 38975,38 | 35591,96 | 42539,81 | 39192,73 | 58,76 | 1031,16 | 36128,56 | 11640,4 | 38384,05 | 36409,12 | 33866,8 | Y |
| BRD3$_{P24Rfs*24}$ | *TSCSTCRMW* | 19926,3 | 36883,75 | 31036,47 | 17277,63 | 30391,12 | 35618,92 | 32413,4 | 28924,01 | 39449,7 | 40500,66 | 30,06 | 14105,5 | Y |
| ARID1A2$_{2141wt}$ | LATPP**F**SRL | 26799,8 | 14123,19 | 27427,52 | 16424,71 | 16900,26 | 3941,71 | 11272,01 | 24362,86 | 11429,5 | 22055 | 1750,71 | 7519,19 | |
| ARID1A2$_{F2141Sfs*59}$ | LATPP**S***AAW* | 25302,6 | 28993,26 | 29795,95 | 14419,97 | 17309,44 | 7602,14 | 24297,04 | 28801,9 | 18275,8 | 27769,73 | 7,45 | 2881,67 | Y |
| ARID1A2$_{F2141Sfs*59}$ | *WPTWLRGTA* | 30949 | 31140,75 | 33196,01 | 38810,41 | 28260,77 | 40,77 | 2436,27 | 21143,94 | 4030,38 | 30553,35 | 30280,2 | 28520,6 | Y |
| ARID1A2$_{F2141Sfs*59}$ | *WLRGTAWQL* | 25460 | 65,54 | 18830 | 9975,21 | 16682,62 | 2052,27 | 226,76 | 7277,85 | 3186,59 | 13700,26 | 16825,46 | 1269,7 | Y |
| ARID1A2$_{F2141Sfs*59}$ | *VPLQCRRAV* | 38163,7 | 27665,36 | 32700,84 | 37369,79 | 37915,14 | 9,78 | 1049,08 | 24823,99 | 8803,18 | 33282,68 | 34020,68 | 30561,9 | Y |
| BLM$_{515wt}$ | **N**ESSYFPGN | 40349,8 | 41961,08 | 42937,04 | 42348,31 | 40843,48 | 43017,95 | 36150,45 | 33709,96 | 36061,4 | 22256,12 | 35694,92 | 35206,7 | |
| BLM$_{N515Mfs*16}$ | **M***KALISQEM* | 28856,5 | 23544,46 | 33726,03 | 28694,9 | 23141,63 | 11925,17 | 5438,91 | 7815,23 | 85,03 | 6667,54 | 16450,32 | 1612,63 | Y |
| BLM$_{N515Mfs*16}$ | *KALISQEMF* | 32033,7 | 28451,91 | 30910,47 | 6552,4 | 37016,07 | 23615,38 | 27026,01 | 23787,2 | 33954,1 | 26392,01 | 23,69 | 3627,61 | Y |
| BLM$_{N515Mfs*16}$ | *SQEMFSQAL* | 19555,9 | 5648,15 | 32409,2 | 18990 | 30963,02 | 10069,77 | 4312,17 | 13846,75 | 35,43 | 502,2 | 32799,71 | 665,11 | Y |
| BLM$_{N515Mfs*16}$ | *QEMFSQALL* | 22677,1 | 13243,2 | 29433,86 | 17547,22 | 22241,2 | 16224,59 | 14413,57 | 14435,26 | 2075,4 | 8,02 | 18951,82 | 5364,22 | Y |
| ZMYM2$_{K1044Rfs*33}$ | *YQDTSLMMI* | 2378,56 | 50,68 | 27599,28 | 6661,62 | 30416,46 | 21304,22 | 12623,68 | 8557,78 | 194,02 | 1272,41 | 13439,8 | 4033,96 | Y |
| ZMYM2$_{K1044Rfs*33}$ | *SLMMIVLTI* | 22112,6 | 20 | 6134,36 | 6061,39 | 25902,87 | 19120,09 | 838,48 | 13606,6 | 8797,27 | 20576,55 | 6866,75 | 1646,95 | Y |
| ZMYM2$_{K1044Rfs*33}$ | *RLYQDTSLM* | 19606,3 | 230,18 | 1441,58 | 9310,15 | 11138,17 | 3041,96 | 8865,79 | 3857,83 | 7592,6 | 12327 | 1962,68 | 32,62 | Y |

The neo-epitopes pass the validation only when the corresponding wt epitope is a poor binder

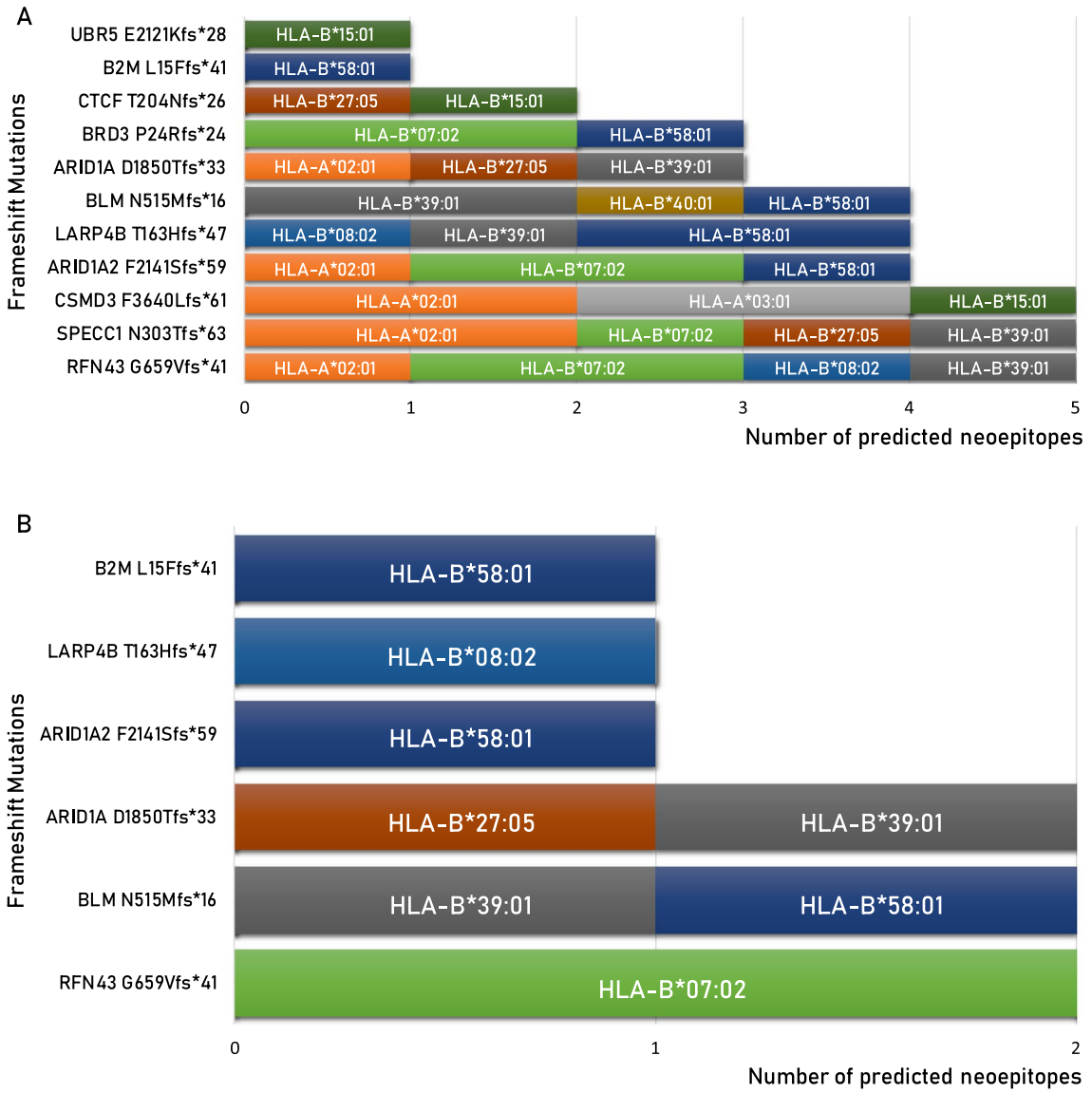


**Fig. 7** High-affinity predicted neoantigens from frameshift mutations and HLA restriction. The number of predicted neoantigens for each frameshift mutations are reported with indication of the HLA restriction, considering the total number of mutations (**A**) or only those not including the product of "abnormal" mRNA (**B**)

A*24:02, A*26:02, B*15:01 and B*40:01 alleles do not bind and present any mutated neoantigens. The remaining ones are predicted to bind 1 or 2 mutated neoantigens (Fig. 9B). Overall, considering both types of mutations, the HLA allele predicted to bind and present the highest number of mutated neoantigens is the B*58:01 (5 neoantigens), followed by the A*03:01 and B*39:01 (3 neoantigens). The HLA-A*01:01, A*26:02, B*40:01 alleles do not bind and present any mutated neoantigens. The remaining ones are predicted to bind 1 or 2 mutated neoantigens (Fig. 9C).

Furthermore, the HLA alleles do influence the mutated proteins for which neoantigens are predicted. Indeed, 50 out of the 62 top missense mutation (80.6%) as well as 12 out of the 23 top frameshift mutations (52.2%) are not predicted to include neoantigens sequences binding to the most frequent HLA alleles. Most importantly, none of the missense and frameshift mutations identified in a relevant percentage of a specific tumor type, is predicted to include neoantigens sequences (Table 7). Looking the other way around, the percentage of tumor cases characterized by missense or frameshift mutations, generating neoantigens in specific HLA alleles, is extremely variable,

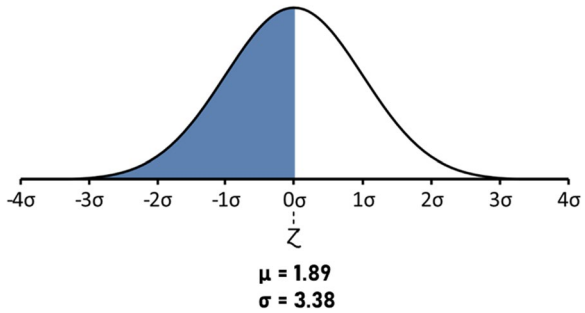


**Fig. 8** Z-score of the observed predicted neoantigens from the hot-spot InDel mutations. The normal distribution of the number of predicted neoantigens from InDel mutations in the 6610 samples present at TCGA. The Z-score of the observed predicted neoantigens from the hot-spot InDel mutations is indicated. It falls in the normal distribution of the expected values with a confidence level of 99.99% (p-value = 0.49)

ranging from 42.5% (eye) to 0.17% (hematopoietic) with an average of 6.97% and a median of 2.42%. Considering the so-called big killers, the percentage range from 18.8% (colon) to 0.4% (prostate). Furthermore, for those with a high-unmet medical need, the percentage is 2.4% for pancreatic ca and 1.78 for brain ca (Fig. 10A).

However, the alleles more prevalently associated to the predicted neoantigens are not from the A locus, which overall has a 60% frequency in the general population. Indeed, most of them are predicted to be linked to alleles of the B locus, in particular HLA-B*58:01, which are among the less frequent and not equally distributed in the global population (Fig. 10B).

## Discussion

The first 100 most frequent cancer mutations reported in the TCGA database were selected to predict shared mutated neoantigens that could be useful for developing off-the-shelf cancer vaccines and/or T cell therapies. Such a selection is significantly representative of all cancer mutations. Indeed, although the first 100 mutations represent a large minority of all somatic mutations in the database (100/193,061 = 0.005%), they cover 56.65% of all identified cancer mutations. Moreover, from the 100th mutation on, each of them is identified in a number of cases lower than 29/14,254 cases and, from the 19,000th mutation, in a single case.

The majority of mutations considered for the study are missense mutations (62%). The top 100 mutations contain the most prevalent ones in the different cancer types, including those with a high unmet medical need (e.g. brain ca, pancreas ca, stomach ca). Indeed, the $IDH1_{R132H}$ is the most prevalent mutation in brain tumors, the $KRAS_{G12D}$ in pancreatic cancer and the $ACVR2A_{K437Rfs*5}$ in gastric cancer, which have a 5 year relative survival rates of almost 36%, 12% and 33%, respectively. Therefore, if such mutations would generate shared tumor specific antigens (TSAs), they would be the optimal antigens for developing specific "off-the-shelf" immunotherapies for about one third of patients affected by these difficult-to treat cancers.

To perform the prediction analyses, the proteins present in the top 100 mutations were manually modified, according to the specific mutations. For the missense mutations, peptides were selected in order to have the mutated residue in each the nine positions ($P_1$ to $P_9$); for the frameshift mutations, peptides were selected also with the sequence downstream of the shifted reading frame. Consequently, while the 945 mutated peptides derived from the missense mutations diverged from the corresponding wt peptides only for a single amino acid, the 686 derived from the InDels included also peptides with a sequence completely different from the wt peptides.

The number of mutated peptides (neoantigens) with affinity < 400 nM to one of the 12 HLA alleles considered in the study is very low, 49 (5.18%) for the ones derived from the missense mutations and 103 (15.01%) for the ones derived from the frameshift mutations. However, the number significantly drops to 20 (2.11%) and 40 (5.83%), respectively, when considering a higher affinity of < 100 nM. Indeed, only peptides with a predicted affinity < 100 nM have been previously shown to have a 100% concordance with ex vivo binding assay [18]. Considering that a neoantigen can be classified as optimal only if the corresponding wt peptide is not antigenic, only 10 neoantigens (1.05%) are identified from the missense mutations. Likewise, also the number of neoantigens derived from the frameshift mutations with a real chance to be presented by cancer cells drops to 9 (1.31%) given that the "abnormal" mRNAs generated by the frameshift contain premature termination codons (PTCs) are recognized and degraded by nonsense-mediated mRNA decay (NMD) [15, 16]. Moreover, even when PTC-containing mRNAs escape from NMD, truncated proteins are not generated due to a translational repression [17].

Considering both types of mutations, the HLA alleles associated with the highest number of predicted neoantigens are from the B loci, namely the B*58:01 (5 epitopes), B*03:01 and B*39:01 (3 epitopes each), B*07:02, B*08:01 and B*27:05 (2 epitopes each). The HLA-A*02:01, A*24:02 and B*15:01 are associated with 1 epitope each. Of interest, three of the HLA alleles (HLA-A*01:01; A*26:02 and B*40:01) are predicted to present no mutated neoantigens. The HLA-A*02:01 and 24:02 are two of the most frequent alleles at global scale (about 40%), these findings imply that the vast majority of cancer patients at global

**Table 6** Predicted neo-epitopes with an affinity value to the HLA alleles < 100 nM, derived from missense mutations, are listed with selected information.

| **FRAMESHIFT** | | | **TOT FREQ** | **TOP FREQ** | **TUMOR** |
|---|---|---|---|---|---|
| **PROTEIN** | **PEPTIDE** | **A*02:01** | | | |
| RFN43 G659Vfs*41 | *TQLARFFPI* | 35.39 | 1.23% | N/A | N/A |
| SPECC1 N303Tfs*63 | *QMLPTLSTL* | 69.86 | 0.36% | N/A | N/A |
| SPECC1 N303Tfs*63 | *LLLGVPQTA* | 72.97 | 0.36% | N/A | N/A |
| ARID1A D1850Tfs*33 | *CLPGLTHPA* | 94.71 | 0.34% | N/A | N/A |
| CSMD3 F3640Lfs*61 | *FLHLYLQDL* | 34.80 | 0.28% | N/A | N/A |
| CSMD3 F3640Lfs*61 | *YLQDLDFIF* | 61.21 | 0.28% | N/A | N/A |
| ARID1A2 F2141Sfs*59 | *WLRGTAWQL* | 65.54 | 0.22% | N/A | N/A |
| **PROTEIN** | **PEPTIDE** | **A*03:01** | | | |
| CSMD3 F3640Lfs*61 | *SIQDVQFMK* | 89.24 | 0.28% | N/A | N/A |
| CSMD3 F3640Lfs*61 | *ITMAKQLLK* | 9.19 | 0.28% | N/A | N/A |
| **PROTEIN** | **PEPTIDE** | **A*24:02** | | | |
| N/A | | | N/A | N/A | N/A |
| **PROTEIN** | **PEPTIDE** | **B*07:02** | | | |
| RFN43 G659Vfs*41 | HPQRKRRG**V** | 73.06 | 1.23% | N/A | N/A |
| RFN43 G659Vfs*41 | **V***PPSPPLAL* | 84.22 | 1.23% | N/A | N/A |
| SPECC1 N303Tfs*63 | *TPLRVQSVL* | 26.51 | 0.36% | N/A | N/A |
| BRD3 P24Rfs*24 | *SPTPASPAA* | 82.17 | 0.24% | N/A | N/A |
| BRD3 P24Rfs*24 | *SPAARPTSC* | 58.76 | 0.24% | N/A | N/A |
| ARID1A2 F2141Sfs*59 | *WPTWLRGTA* | 40.77 | 0.22% | N/A | N/A |
| ARID1A2 F2141Sfs*59 | *VPLQCRRAV* | 9.78 | 0.22% | N/A | N/A |
| **PROTEIN** | **PEPTIDE** | **B*08:02** | | | |
| RFN43 G659Vfs*41 | *TQLARFFPI* | 48.47 | 1.23% | N/A | N/A |
| LARP4B T163Hfs*47 | VLKK**H***WNSA* | 59.58 | 0.43% | N/A | N/A |
| **PROTEIN** | **PEPTIDE** | **B*27:05** | | | |
| SPECC1 N303Tfs*63 | *LRVQSVLLL* | 46.92 | 0.36% | N/A | N/A |
| ARID1A D1850Tfs*33 | WRIGGG**T***PL* | 51.15 | 0.34% | N/A | N/A |
| CTCF T204Nfs*26 | *QRCRCVCLR* | 94.17 | 0.32% | N/A | N/A |
| **PROTEIN** | **PEPTIDE** | **B*39:01** | | | |
| RFN43 G659Vfs*41 | *TQLARFFPI* | 87.73 | 1.23% | N/A | N/A |
| LARP4B T163Hfs*47 | *YHRWIVTSM* | 71.78 | 0.43% | N/A | N/A |
| SPECC1 N303Tfs*63 | *HQAPQVAML* | 80.74 | 0.36% | N/A | N/A |
| ARID1A D1850Tfs*33 | WRIGGG**T***PL* | 6.08 | 0.34% | N/A | N/A |
| BLM N515Mfs*16 | **M***KALISQEM* | 85.03 | 0.22% | N/A | N/A |
| BLM N515Mfs*16 | *SQEMFSQAL* | 35.43 | 0.22% | N/A | N/A |
| **PROTEIN** | **PEPTIDE** | **B*40:01** | | | |
| BLM N515Mfs*16 | *QEMFSQALL* | 8.02 | 0.22% | N/A | N/A |
| **PROTEIN** | **PEPTIDE** | **B*58:01** | | | |
| LARP4B T163Hfs*47 | *VTCILYHRW* | 12.22 | 0.43% | N/A | N/A |
| LARP4B T163Hfs*47 | *TSMCQSQRW* | 10.43 | 0.43% | N/A | N/A |
| B2M L15Ffs*41 | AVLALLS**F***W* | 32.16 | 0.25% | N/A | N/A |
| BRD3 P24Rfs*24 | *TSCSTCRMW* | 30.06 | 0.24% | N/A | N/A |
| ARID1A2 F2141Sfs*59 | LATPP**S***AAW* | 7.45 | 0.22% | N/A | N/A |
| BLM N515Mfs*16 | *KALISQEMF* | 23.69 | 0.22% | N/A | N/A |
| **PROTEIN** | **PEPTIDE** | **B*15:01** | | | |
| UBR5 E2121Kfs*28 | *VQNQGHLLM* | 37.86 | 0.49% | N/A | N/A |
| CTCF T204Nfs*26 | *NQKEQTALY* | 58.38 | 0.32% | N/A | N/A |
| CSMD3 F3640Lfs*61 | *CMTPTQSQW* | 31.13 | 0.28% | N/A | N/A |

**Table 6** (continued)

The peptide sequences include the mutated aminoacid residue (bold & underlined) or the newly generated sequence downstream of the frameshift (italics). Values in the column of the haplotypes indicate the predicted affinity (nM). TOT FREQ: frequency in the TCGA database; TOP FREQ: top frequency in specific tumor; TUMOR: tumor type in which the top frequency is reported

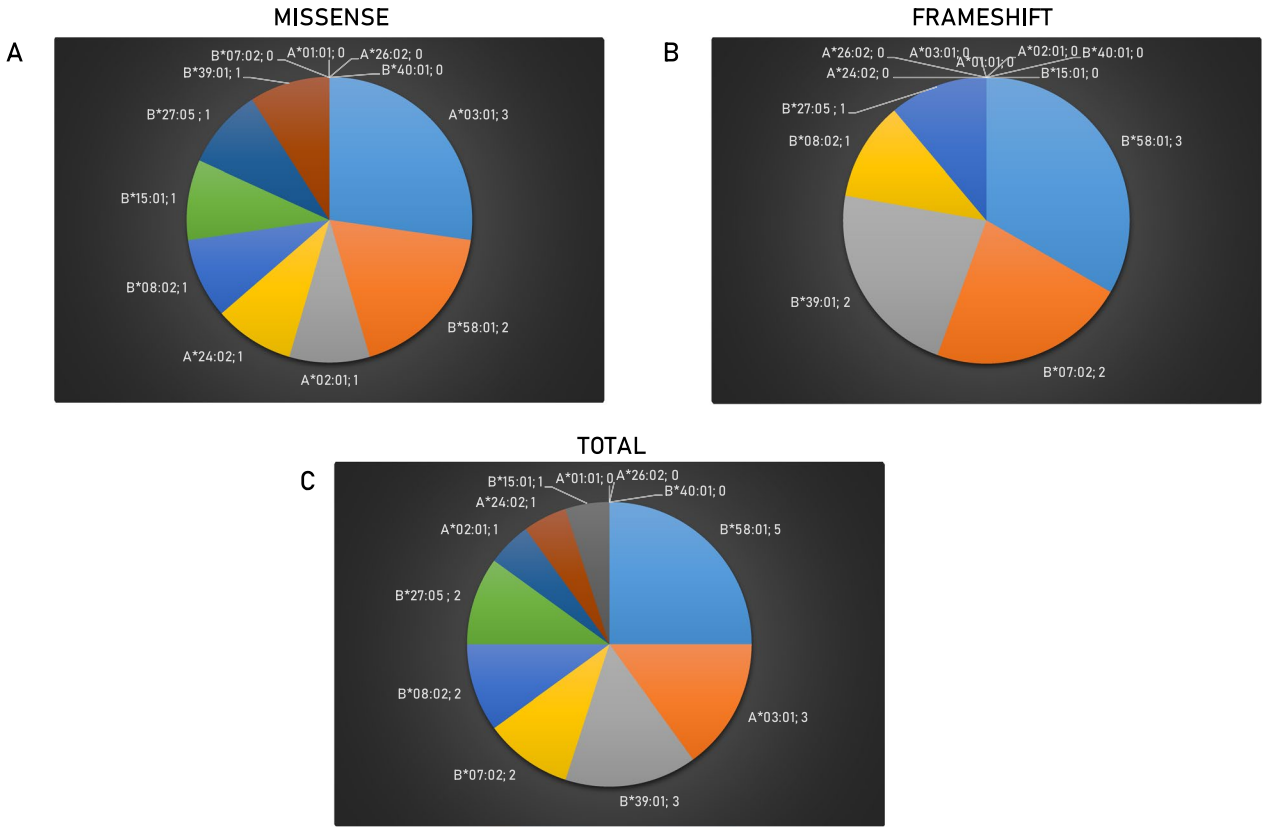


**Fig. 9** Number of predicted neoantigens for each haplotype. The number of predicted neoantigens is indicated for each of the 12 haplotypes taken into consideration. The numbers are indicated in a top-down listing in a clockwise direction. Neoantigens derived from missense mutations are listed in panel **A**; those derived from frameshift mutations are listed in panel **B**; the total neoantigens are listed in panel **C**

level (> 50%) cannot benefit from tumor-specific shared mutated neoantigens.

Overall, the percentage of tumor cases characterized by missense or frameshift mutations generating neoantigens in specific HLA alleles is low, variable and associated to low-frequent HLA alleles. Indeed, the average of tumor cases is 6.97% and a median of 2.42% with a wide range going from 42.5% (eye) to 0.17% (hematopoietic). 22 out of 31 tumors (71%) show a percentage of cases characterized by mutations generating neoantigens lower than 5% and most of the big killers (e.g. breast, lung, prostate, liver ca) as well as those with a high unmet medical need (e.g. pancreas and brain ca) are in the lower part of the list (< 5%). The number of observed predicted neoantigens from the hot-spot missense mutations is significantly lower than the expected ones. On the contrary, the number of observed predicted neoantigens from the hot-spot InDel mutations is perfectly comparable to the expected ones. This supports the hypothesis that, the first ones are selected by the immunological pressure, while the latter are not because they are not translated and not presented to the immune system.

However, also the few cancers with a relevant percentage of cases (> 10%) with mutations generating neoantigens, these are associated to low prevalent HLA alleles. The $GNA11_{Q209L}$ missense mutation, giving rise to the FRMVDVGG**L** epitope, is the most frequent mutation in uveal melanoma (UM) (42,50% of all cases reported in TCGA). Unfortunately, the clinical impact of this neoantigen appears to be very limited because UM is a rare tumor, with an average incidence rate of 5 per million globally [19] and the binding B*27:05 and B*39:01 HLA alleles are among the least frequent in the World. The two big killers colon and stomach cancers show 18.8% and 23.9% of cases which are characterized by mutations giving rise to shared neoantigens and could benefit from

**Table 7** Missense and frameshift mutations for which neoepitopes have not been predicted in any of the 12 haplotypes considered in the study

| MISSENSE | TOP FREQ | FRAMESHIFT | TOP FREQ |
|---|---|---|---|
| $BCOR_{N1459S}$ | | $ACVR2A_{K437Rfs*5}$ | Stomach |
| $BRAF_{V640E/V600E}$ | Colon, Skin, Thyroid | $APC_{T1556Nfs*3}$ | |
| $BRAF_{V640M}$ | | $BCORL1_{P1681Qfs*20}$ | |
| $ERBB2_{S310F}$ | | $GLI1_{G274Afs*6}$ | |
| $FBXW7_{R465C}$ | | $JAK1_{K860Nfs*16}$ | |
| $FBXW7_{R465H}$ | | $JAK1_{P430Rfs*2}$ | |
| $FBXW7_{R479Q}$ | | $NPM1_{W317Cfs*12}$ | |
| $FBXW7_{R505C}$ | | $PTEN_{K267Rfs*9}$ | |
| $FBXW7_{R505G}$ | | $PTEN_{T319*}$ | |
| $FGFR2_{S252W}$ | | $RPL22_{K15Rfs*5}$ | Corpus Uteri |
| $FGFR3_{S249C}$ | Bladder | $SALL4_{V995Ffs*14}$ | |
| $GNAQ_{Q209P}$ | | $ZMYM2_{K1044Rfs*33}$ | |
| $HRAS_{Q61R}$ | Adrenal gland | | |
| $IDH1_{R132C}$ | | | |
| $IDH1_{R132H}$ | Brain | | |
| $KRAS_{A146T}$ | | | |
| $KRAS_{G12A}$ | | | |
| $KRAS_{G12C}$ | Lung | | |
| $KRAS_{G12D}$ | Pancreas, Rectum | | |
| $KRAS_{G12R}$ | | | |
| $KRAS_{G12S}$ | | | |
| $KRAS_{G12V}$ | | | |
| $KRAS_{G13D}$ | | | |
| $KRAS_{Q61H}$ | | | |
| $NRAS_{G12D}$ | | | |
| $NRAS_{G13D}$ | | | |
| $NRAS_{G13R}$ | | | |
| $NRAS_{Q61K}$ | | | |
| $NRAS_{Q61L}$ | | | |
| $NRAS_{Q61R}$ | Hematopoietic | | |
| $PIK3CA_{C420R}$ | | | |
| $PIK3CA_{E542K}$ | | | |
| $PIK3CA_{E545K}$ | Cervix uteri, Larynx | | |
| $PIK3CA_{H1047R}$ | Breast | | |
| $POLE_{P286R}$ | | | |
| $PTEN_{R130G}$ | Uterus | | |
| $PTEN_{R130Q}$ | | | |
| $TP53_{G245S}$ | | | |
| $TP53_{H179R}$ | | | |
| $TP53_{H193R}$ | | | |
| $TP53_{I195T}$ | | | |
| $TP53_{R175H}$ | Ovary, Esophagus | | |
| $TP53_{R248Q}$ | | | |
| $TP53_{R248W}$ | Bones | | |
| $TP53_{R273C}$ | | | |
| $TP53_{R273H}$ | | | |
| $TP53_{R273L}$ | | | |
| $TP53_{R282W}$ | | | |
| $TP53_{V157F}$ | | | |
| $TP53_{Y220C}$ | | | |

**Table 7** (continued)

For those, which are reported as top frequent in a specific tumor, the tumor types are listed (TOP FREQ.)

cancer vaccines based on TSAs. Considering that they show and age-standardized rate (ASR) of 6.2 and 6.1, respectively, this would be a huge advancement in cancer therapy. Unfortunately, such neoantigens are mostly associated to the B loci of the HLA, that show a prevalence much lower than 10%, and even lower than 1%, in world populations. This drastically reduce the potential application of such therapy. The only exception is represented by the predicted neoantigens derived from $PIK3CA_{H1047L}$ (FMKQMNDAL) and LARP4BT163Hfs*47 (VLKKHWNSA), linked to HLA-B*08:01, as well as the one derived from RFN43G659Vfs*41 (HPQRKRRGV), linked to HLA-B*07:02. Indeed, these two alleles cover 21.8% and 20.5% of the European population and, therefore, could represent a great opportunity for providing an additional therapeutic opportunity to European patients affected by such deadly cancers.

In conclusions, the search for shared mutated tumor-specific neoantigens for developing off-the-shelf highly specific immunotherapies results in an unfortunate failure. The most frequent mutations, either missense or InDel, do not give rise to any predicted neoantigen with high affinity to the most frequent HLA-A and B alleles. Such evidence is likely to be the result of a very strong selection by the immune system in the very early stages of tumor development, which eliminates cancer cells expressing mutated immunogenic neoantigens. At that stage, the tumor cells characterized by mutations giving rise to highly antigenic non-self-mutated neoantigens would be efficiently targeted and eliminated. The result is the selection of cancer cells expressing only wild type self-antigens and, consequently, able to escape the immune control. Finally, they will form tumor lesions embedded in a very immune-suppressive microenvironment, which is difficult to be accessed by T cells (Fig. 11).

Therefore, cancer vaccines may only rely upon personalized mutated neoantigens, with all the caveats and limitations, or upon wild type over-expressed tumor-associated antigens (TAAs), which may suffer from immunological tolerance. In order to overcome the latter drawback, non-self-antigens mimicking the TAAs (molecular mimicry), and able to elicit cross-reactive T cells, should be actively searched (i.e. antigens derived from microorganisms). This will provide the essential tool for developing off-the-shelf vaccines with the optimal immunogenicity to elicit an efficient anti-tumor T cell immune response [20–23].

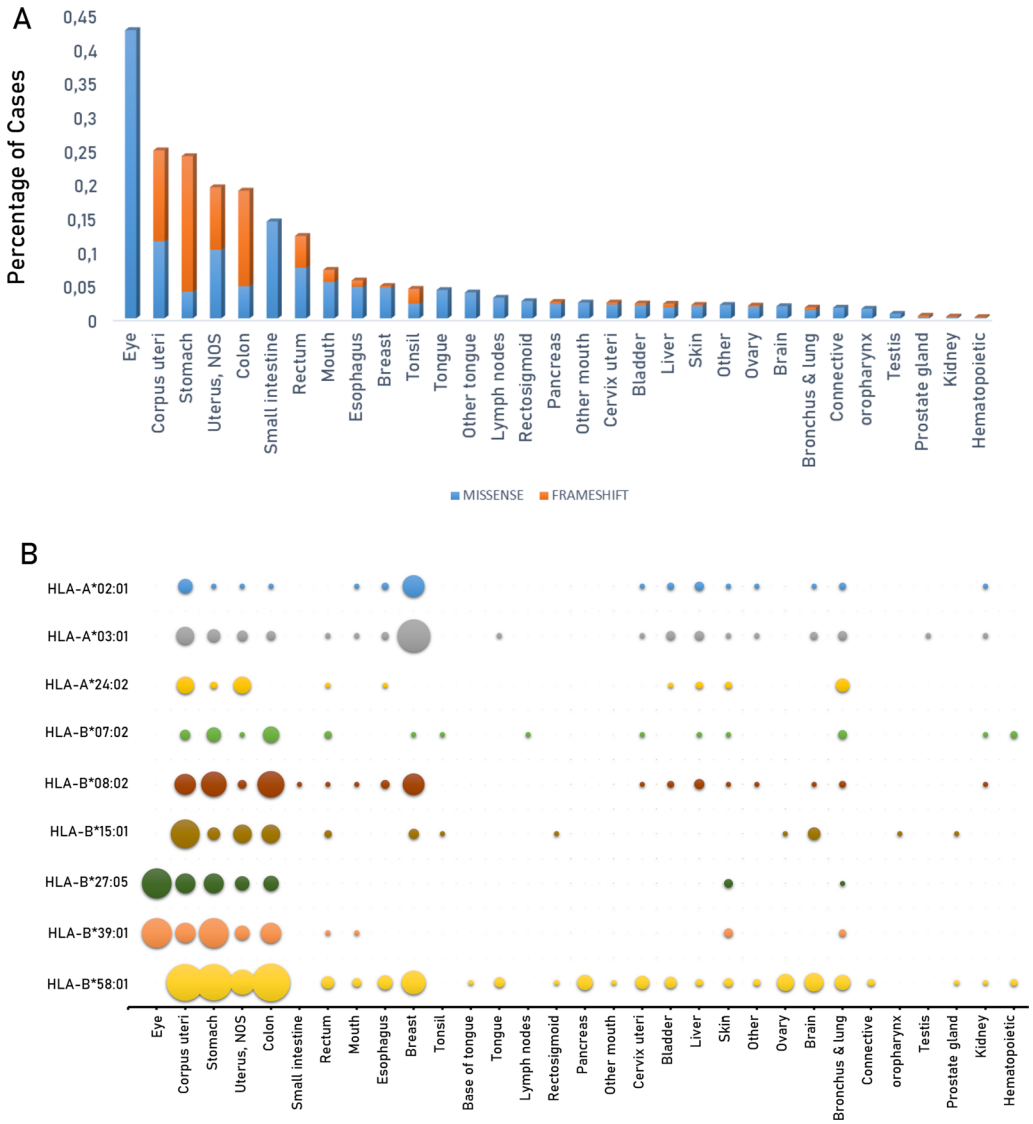


**Fig. 10** Predicted neoantigens in each tumor and haplotype. The percentage of cases with mutations predicting for neoantigens are indicated for each cancer (**A**); the percentage of neoantigens associated to each haplotype are indicated for each cancer (**B**)

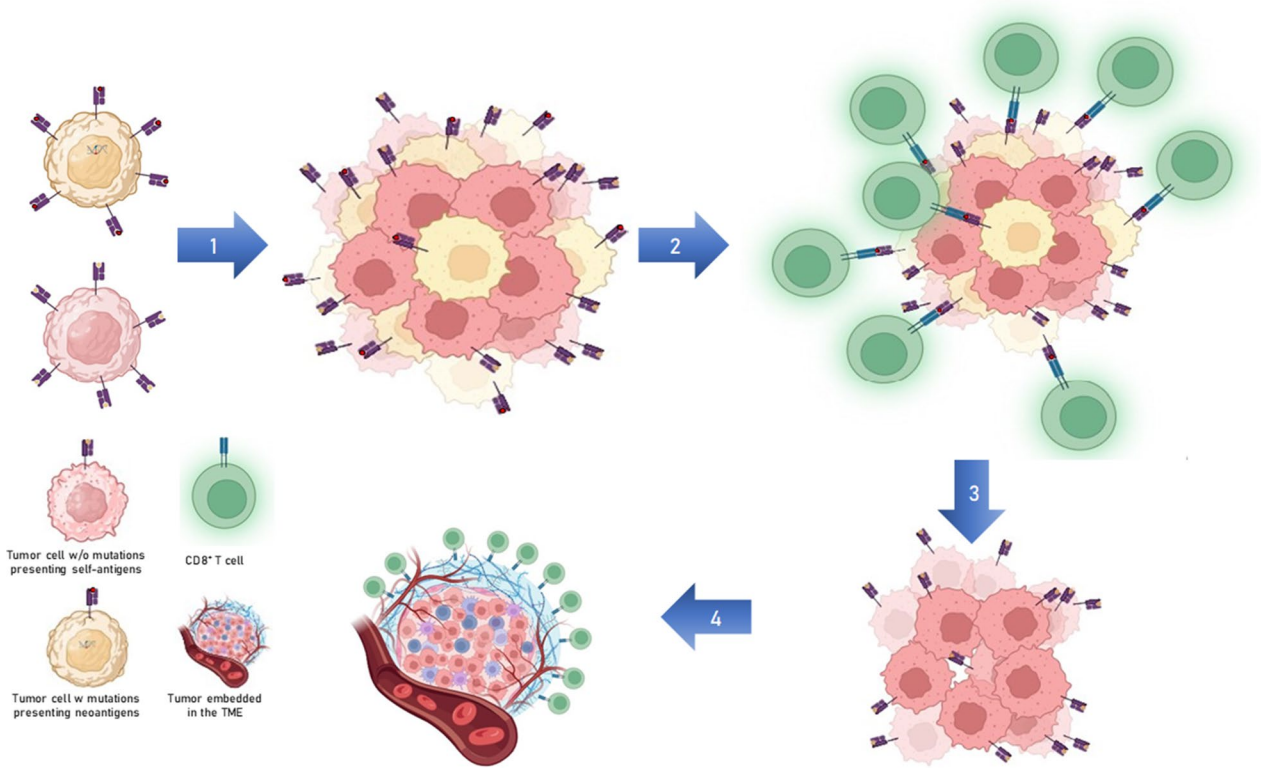


**Fig. 11** Prospected evolution of tumor lesions. Cells with mutations presenting shared neoantigens are eliminated in the very early stages of tumor lesions, when a full TME is not present. Cells not expressing shared neoantigens can outgrow without immunological control to form a tumor lesion embedded in the TME, difficult to be attacked by T cells

## Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12967-024-05110-0.

**Additional file 1: Table. S1.** List of the 100 most frequent mutations in the 14254 tumor cases reported in TCGA. **Table S2.** List of the 100 most frequent mutations reported in TCGA. **Table. S3.** List of all peptides from the wt and mutated sequences derived from proteins with missense mutations. **Table. S4.** List of all peptides from the wt and mutated sequences derived from proteins with frameshift mutations.

**Additional file 2: Fig. S1.** Percentage of mutated samples, for each tumor type, presenting the indicated missense mutation giving rise to neoantigens. **Fig. S2.** Percentage of mutated samples, for each tumor type, presenting the indicated frameshift mutation giving rise to neoantigens.

**Acknowledgements**
Not applicable.

**Author contributions**
CR performed 70% of all the antigen prediction analyses; BC and AM performed the remaining 30% (15% each) of the antigen prediction analyses. MT and LB designed the structure of the review article, supervised the analysis and drafted 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) (to LB).

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

## Declarations

**Ethics approval and consent to participate**
Not applicable.

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

**Competing interests**
The authors declare no potential competing interests.

Received: 26 February 2024 Accepted: 19 March 2024
Published online: 10 April 2024

**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. Castle JC, Uduman M, Pabla S, Stein RB, Buell JS. Mutation-derived neoantigens for cancer immunotherapy. Front Immunol. 2019;10:1856.
7. Snyder A, Makarov V, Merghoub T, et al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. N Engl J Med. 2014;371:2189–99.
8. Gubin MM, Zhang X, Schuster H, et al. Checkpoint blockade cancer immunotherapy targets tumour-specific mutant antigens. Nature. 2014;515:577–81.
9. Rizvi NA, Hellmann MD, Snyder A, et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. Science. 2015;348:124–8.
10. Puig-Saus C, Sennino B, Peng S, et al. Neoantigen-targeted $CD8^+$ T cell responses with PD-1 blockade therapy. Nature. 2023;615:697–704.
11. Bjerregaard AM, Nielsen M, Jurtz V, et al. An analysis of natural T cell responses to predicted tumor neoantigens. Front Immunol. 2017;8:1566.
12. Wells DK, van Buuren MM, Dang KK, et al. Key parameters of tumor epitope immunogenicity revealed through a consortium approach improve neoantigen prediction. Cell. 2020;183:818-834.e13.
13. Katsikis PD, Ishii KJ, Schliehe C. Challenges in developing personalized neoantigen cancer vaccines. Nat Rev Immunol. 2023;24(3):213.
14. Rojas LA, Sethna Z, Soares KC, et al. Personalized RNA neoantigen vaccines stimulate T cells in pancreatic cancer. Nature. 2023;618:144–50.
15. 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.
16. Wagner E, Lykke-Andersen J. mRNA surveillance: the perfect persist. J Cell Sci. 2002;115:3033–8.
17. 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.
18. Petrizzo A, Tagliamonte M, Mauriello A, et al. Unique true predicted neoantigens (TPNAs) correlates with anti-tumor immune control in HCC patients. J Transl Med. 2018;16:286.
19. Naseripoor M, Azimi F, Mirshahi R, Khakpoor G, Poorhosseingholi A, Chaibakhsh S. Global incidence and trend of uveal melanoma from 1943–2015: a meta-analysis. Asian Pac J Cancer Prev. 2022;23:1791–801.
20. Tagliamonte M, Cavalluzzo B, Mauriello A, et al. Molecular mimicry and cancer vaccine development. Mol Cancer. 2023;22:75.
21. Tornesello AL, Tagliamonte M, Tornesello ML, Buonaguro FM, Buonaguro L. Nanoparticles to improve the efficacy of peptide-based cancer vaccines. Cancers. 2020;12(4):1049.
22. Tornesello ML, Annunziata C, Buonaguro L, Greggi S, Buonaguro FM. TP53 and PIK3CA gene mutations in adenocarcinoma, squamous cell carcinoma and highgrade intraepithelial neoplasia of the cervix. J Trans Med. 2014;12(1):255
23. Pezzuto F, Izzo F, Buonaguro L, et al. Tumor specific mutations in TERT promoter and CTNNB1 gene in hepatitis B and hepatitis C related hepatocellular carcinoma. Oncotarget. 2016;7(34):54253–62.


## Publisher's Note