OPEN ACCESS Freely available online

Journal of Proteomics & Bioinformatics

Research Article

# ARDiTox: Platform for the Prediction of T-Cell Receptors (TCRs) Potential Off-Target Binding

**Victor Murcia Pienkowski[1*], Tamara Boschert[2], Piotr Skoczylas[1], Anna Sanecka-Duin[1], Maciej Jasinski[1], Bartlomiej Krol-Jozaga[1], Giovanni Mazzocco[1], Slawomir Stachura[1], Lukas Bunse[2,3], Jan Kaczmarczyk[1], Edward W. Green[2], Agnieszka Blum[1]**

*[1]Department of Immunology, Ardigen Research and Development Center, Krakow, Poland; [2]Department of Neuroimmunology and Brain Tumor Immunology, German Cancer Research Center (DKFZ), Heidelberg, Germany; [3]Department of Neurology, Heidelberg University, Mannheim, Germany*

## ABSTRACT

Cellular immunotherapies, such as those utilizing T lymphocytes expressing native or engineered T-cell Receptors (TCRs), have already demonstrated therapeutic efficacy. However, some high-affinity TCRs have also proved to be fatal due to off-target immunotoxicity. This process occurs when the immune system acts against epitopes found on both tumor cells and healthy tissues. Moreover, some TCRs can be cross-reactive to epitopes with highly dissimilar sequences. To address this issue, we developed ARDitox, a novel *in silico* method based on computational immunology and Artificial Intelligence (AI), for predicting and analyzing potential off-target toxicities. We tested the performance of ARDitox *in silico* on 4 different epitopes found in the literature where TCRs were used to target cancer-related antigens as well as on a set of TCR targeting a viral epitope. Two of them have a specific clinical outcome in which immunotoxicity was reported (*MAGEA3*112-120 and *MAGEA3*168-176 epitopes), one was tested using an X-scan approach (*AFP158*-166 epitope), and the last one with no cross-reactive epitopes identified in clinical trials (NY-ESO-1157-165 epitope). Overall, ARDitox has identified immunotoxic epitopes in line with the data available in the literature. In addition, we investigated a very promising TCR, which is still in development, against a peptide coded by the NLGN4X gene. For this epitope, we detected a cross-reactive peptide that otherwise would be difficult to detect *in vitro*. In conclusion, *in silico* approach is a powerful tool that accurately identifies off-target epitopes and should be considered in preclinical studies, as it can effectively complement the development of safer anti-cancer therapies.

**Keywords:** Off-target toxicity; Off-target binding; Cross-reactivity; Molecular mimicry

## INTRODUCTION

Recent advances in the field of immunotherapy are slowly changing the landscape of available treatments for cancer. This is especially applicable for hematological malignancies and some solid tumors [1-3]. Among the most promising, prospective therapies that boost the body's natural defenses, are adoptive cell therapies with T lymphocytes expressing native or engineered T-cell Receptors (TCRs) as well as TCR mimics (TCRm) antibodies [4,5]. There are two main approaches to TCR-based therapies: autologous (the expanded T-cells are obtained and administered to the same cancer patient) and allogeneic (the expanded T-cell clone is given from a donor to a patient) [6]. Unfortunately, like with any novel technologies, several difficulties need to be addressed before such therapies become widely used. One of the main issues, that in some proved to be life-threatening to treated patients, is off-target immunotoxicity [7]. The mechanisms of such toxicity include T-cells acting against both the cancer cells and healthy tissues.

T-cell immune surveillance consists of the TCR scanning of short peptides presented at the cell surface by receptors called Human Leukocyte Antigens (HLA) [8]. Importantly, only a small fraction of all peptides from the human proteome is presented by the HLA. In a multi-step process, a peptide is loaded onto the HLA

**Correspondence to:** Victor Murcia Pienkowski, Department of Immunology, Ardigen Research and Development Center, Krakow, Poland, E-mail: victor.murciapienkowski@gmail.com

**Received:** 22-Aug-2023, Manuscript No. JPB-23-26184; **Editor assigned:** 24-Aug-2023, PreQC No. JPB-23-26184 (PQ); **Reviewed:** 07-Sep-2023, QC No. JPB-23-26184; **Revised:** 14-Sep-2023, Manuscript No. JPB-23-26184 (R); **Published:** 25-Sep-2023, DOI: 10.35248/0974-276X.23.16.651.

**Citation:** Pienkowski VM, Boschert T, Skoczylas P, Sanecka-Duin A, Jasinski M, Krol-Jozaga B, et al. (2023) ARDiTox: Platform for the Prediction of T-Cell Receptors (TCRs) Potential Off-Target Binding. J Proteomics Bioinform.16:651

and exported to the cell surface [9]. For a given HLA of type I, it is estimated that the number of unique peptide sequences presented to T-cells is ~ 1,000-~ 25,000. However, considering all theoretical peptides that could appear in the cell (e.g. derived from cancer mutations or viral proteins), the number of putative epitopes increases to a set of >209 [10]. As such, assuming that in the human body there are ~ 108 unique TCRs, a T-cell must be able to interact with several peptides to cover the whole epitope set [11,12]. All this results in a situation where a non-autologous TCR applied to a patient against e.g. a certain tumor-associated epitope, might lead to an additional interaction between the TCR and an Off-Target Epitope (OTE) presented on healthy tissue. Moreover, the risk of cross-reactivity and potential immunotoxicity may be increased for the TCRs engineered specifically to enhance peptide-HLA (pHLA) TCR affinity. As previously shown, TCRs with high affinity towards a single target may exhibit increased cross-reactivity against other targets.

Off-target toxicity has already proven to be extremely dangerous to the patient's health as it resulted in the death of at least four people in two independent clinical trials [13]. Importantly, the sequences of the target epitope and the OTE do not necessarily have to be very similar. The clinically relevant cases of cross-reactivity showed that a minimum of 5 identical amino acids is sufficient for off-target toxicity to occur [14]. Unfortunately, using experimental methods to test all possible off-targets is costly and time-consuming thus has to be limited to a restricted subspace of potential off-target sequences.

Leveraging the recent advances in computational immunology and AI can augment these efforts, ultimately increasing the number and safety of available treatments. To this end, we introduce ARDitox, a novel method for predicting and analyzing off-target toxicity.

## MATERIALS AND METHODS

### The ARDitox pipeline

The pipeline of our method consists of 5 consecutive steps as discussed in the following sections (Figure 1).

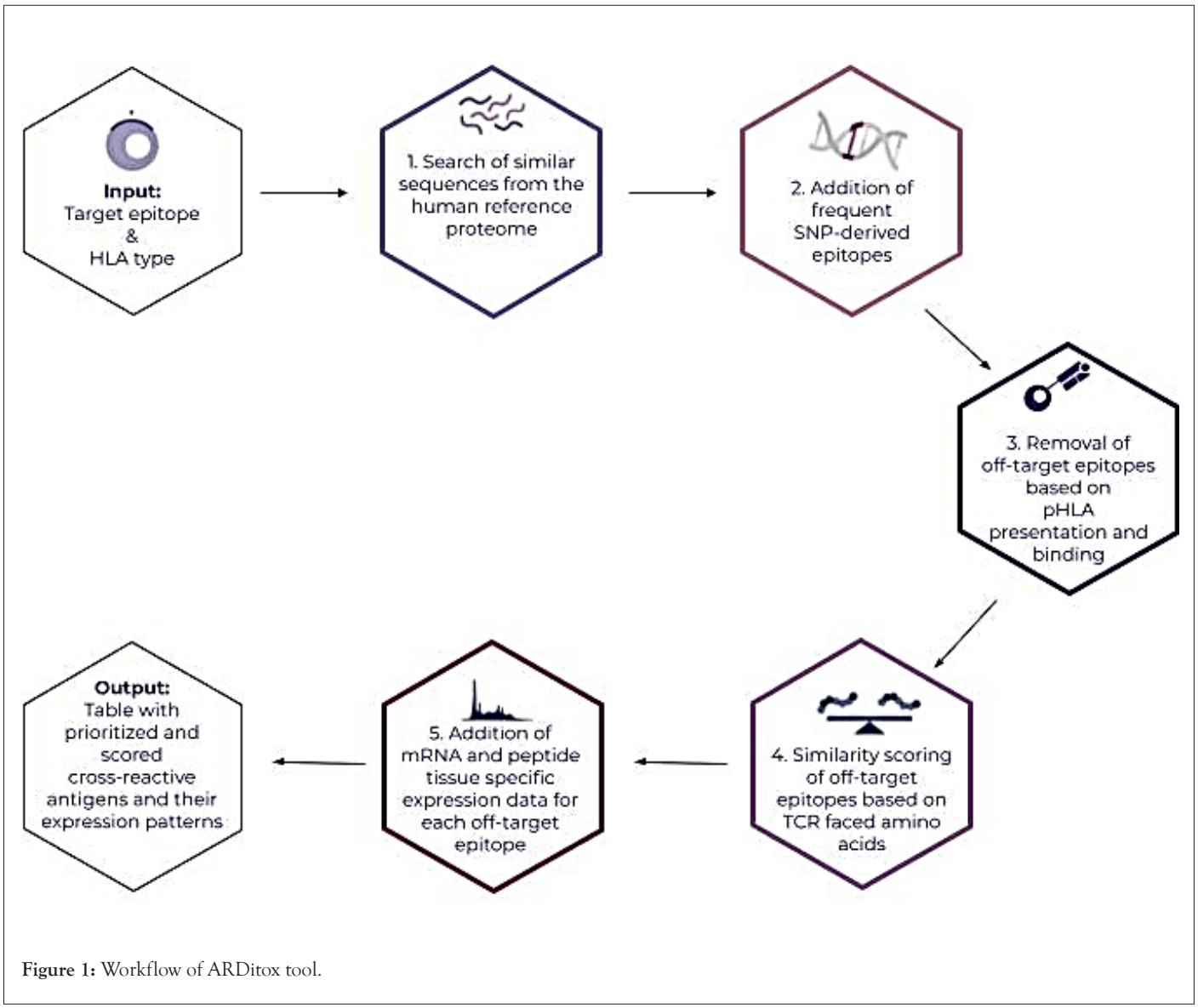


**Figure 1:** Workflow of ARDitox tool.

**Identification of all putative off-target sequences:** ARDitox takes as an input a target epitope from 8 to 11 amino acids long and its corresponding HLA-type. The first step of the algorithm consists of the identification of all epitopes that have at least 5 amino acids shared with the target epitope. Importantly, only OTEs of the same length are taken into account as it is considered rare for epitopes of different lengths to bind to the same TCR [15]. This step generates a large number of putative OTEs, e.g. for an 8 amino acid target epitope there are combinatorially 459360 possible putative OTEs. Each of these putative OTEs is then checked for presence in the human reference proteome. The reference for the above-mentioned search can be found on the UniProt website [16].

**Addition of single nucleotide variant epitopes:** Single Nucleotide Polymorphisms (SNPs) are a major source of OTEs as a given human genome contains ~ 7000 nonsynonymous germline SNPs. As such, SNPs can be a major source of novel off-target sites in TCR-based therapies. Unfortunately, both rare and frequent SNPs are not included in the UniProt proteome reference sequences. ARDitox tackles this problem by accounting for frequent nonsynonymous mutations (occurring in more than 1% of at least one studied population) from gnomAD database based on putative OTEs generated in the previous step [17]. Each identified nonsynonymous frequent SNP occurring in the sequence corresponding to the OTEs is taken into account through an additional putative off-target epitope. For the OTEs derived from SNP, their frequency in the largest human subpopulation is provided in the final results. This step further increases the number of OTEs to be analyzed, making the algorithm more sensitive.

**Selection of presented epitopes:** The previous steps generate an enormous number of potential OTEs. However, only a small fraction of them will be bound and presented via class I HLA. In order to limit the number of putative OTEs to the ones presented at the cell surface, we use an in-house developed presentation model [18]. The model is based on machine learning methods and trained on curated, publicly available datasets [19-21]. The datasets consist of the results of mass-spectrometry experiments conducted on monoallelic human cell lines. The presentation model is based on artificial neural networks and uses both the peptide sequence and the HLA type as separate inputs. Overall, this model can be used to generate predictions for any canonical class I HLA (i.e., A, B and C). The output consists of pHLA probability of being presented at the cell surface. Next, peptide-HLA binding is evaluated on the OTEs using MHCflurry [22]. Only OTEs with a probability of being presented >50% and binding affinity <2000 nM proceed to the next steps.

**Off-target epitopes ranking:** In the target epitope, amino acids at different positions can interact with the HLA and with the TCR. In order for these interactions to occur, the physico-chemical properties of the amino acids at certain positions must remain similar [23]. In this step, we establish the positions of TCR-faced residues, depending on the HLA type the epitope binds to. The positions are based on literature and database search that contain information, which residues are crucial for TCR binding [24-27]. The comparison between the target epitope and the putative OTE is performed based on the differences in physico-chemical properties of the TCR-facing residues. To this end, we consider physico-chemical properties most relevant to the pHLA:TCR interaction that are acquired from https://www.genome.jp/. All the matrices containing these properties were linearized into vectors. The distance between each target epitope $e$ and the putative OTE p can be computed.

**mRNA and peptide tissue specific expression:** Finally, mRNA expression in Transcripts Per Million (TPM) (Gene and Transcript Expression (GTEx)) and protein expression level (Human Protein Atlas (HPA)) of the putative off-target epitopes are added in an attempt to identify tissues sensitive to off-target toxicity [28,29].

## Implementation

The method was implemented using Python 3.7 (scikit-learn, pandas, and numpy) and R v4.0.2 (dplyr, ggplot2 and BSgenome) [30-35].

## *In vitro* validation

**mRNA electroporation for TCR expression:** TCRs were ordered from TWIST Biosciences in custom vectors. *In vitro* transcribed RNA was generated using T7 Scribe Standard RNA IVT Kit (Biozym #150404). RNA transfection was performed by electroporation using the 4D-Nucleofector electroporation system (Lonza).

**Cell culture:** Jurkat T cells (Leibnitz Institute DSMZ #ACC282) were cultured in RPMI 1640 Medium (Gibco, #61870143)+10% heat-inactivated FBS (PAN-Biotech) and 1% Penicillin-Streptomycin (Capricorn Scientific #PS-B). The commercially available EBV-immortalized BOLETH cell line expressing HLA-A*02:01 was used as antigen presenting cells. The BOLETH cells were cultured with 50 mM β-mercaptoethanol, 1 mM sodium pyruvate (Thermo Fisher Scientific #11360039), and 1X MEM non-essential amino acids (Gibco, #11140035). BOLETH and T cells were plated in a 1:3 ratio in a round-bottom 96-well plate and the respective peptides were added at a final concentration of 10 µM. As a positive control, Jurkat T cells were plated in a well pre-coated with CD28/CD3 monoclonal antibodies at a 1:400 dilutions. After 16 h of co-culture T cell activation was assessed.

**Flow cytometry:** Flow cytometry samples were diluted and washed with FACS buffer (PBS+2 mM EDTA+2% FBS) and centrifuged at 300 g for 5 min. The supernatants were removed and the cells were resuspended in 50 µl of staining solution containing the respective diluted antibodies as indicated in Table 1. Following 20 min incubation at 4°C cells were washed twice with FACS buffer, pelleted (300 g for 5 min), and resuspended in 100 µL FACS buffer (Table 1).

**Dataset preparation:** We selected three groups of peptides presented on either HLA-A*02:01 or HLA-A*01:01 for the evaluation of our methodology such as Tumor Associated Antigen (TAA) epitopes, known immunodominant viral epitopes and epitopes derived from frameshift mutations. TAA and virus epitopes were acquired from IEDB frameshift derived epitopes were obtained from a library of Neo Open Reading Frame peptides [36,37]. We used ARDisplay to predict frameshift epitopes presented by HLA-A*02:01, and down sampled the dataset to 16 epitopes. Importantly, we randomly selected 16 TAAs presented on HLA-A*02:01 in order to compare TAA *vs.* frameshift epitopes on equally abundant groups.

**Table 1:** List of the antibodies and their respective dilutions.

| Specificity | Fluorophore | Clone | Supplier | Product number | Dilution |
|---|---|---|---|---|---|
| CD2 | PerCP/Cy5 | RPA-2.10 | Biolegend | 300216 | 0.180555556 |
| CD3 | Unlabeled | OKT3 | Biolegend | 317325 | 0.319444444 |
| CD20 | PacificBlue | 2H7 | Biolegend | 302320 | 0.180555556 |
| CD28 | Unlabeled | E18 | Biolegend | 122022 | 0.319444444 |
| CD69 | PE/Cy7 | FN50 | Biolegend | 310912 | 0.180555556 |
| Mouse-TCR-β | PE | H57-597 | Biolegend | 109208 | 0.180555556 |

**Note:** TCR: T cell receptors; PE: Phycoerythrin

## RESULTS

ARDitox estimates the safety of the putative OTEs through a safety score that can vary from 0 to 14. A score close to 0 means that the putative OTE and the target epitope are predicted to have an almost identical pHLA:TCR interaction, and that cross-reactive binding is highly probable.

### Known cross-reactive epitopes

We tested ARDitox on 5 epitopes of TCRs targeting TAA epitopes and on 1 of a set of T cells targeting a viral peptide are described below [38-41]. Table 2 shows the target epitopes and their respective cross-reactive epitopes (Table 2).

Morgan, et al. [13] described a cell therapy based on TCR-engineered T-cells against the KVAELVHFL epitope derived from *MAGEA3*. The clinical study was performed on 9 HLA-A*02:01 positive patients with metastatic cancer expressing both *MAGEA3* and *MAGEA12*. Unfortunately, three patients developed severe side effects. Two patients went into a coma that ultimately resulted in their death. One patient developed Parkinson's disease-like symptoms that lasted for 4 weeks after the administration of the drug. Based on molecular assays the authors showed that genes from the same family as the target, specifically *MAGEA12* and to a lower degree MAGEA1, MAGEA8, and MAGEA9 are expressed in the brain and are probably responsible for the off-target toxicity effects. A re-evaluation of the OTEs that could be responsible for the neurotoxicity of this TCR was conducted by Martin, et al. [42]. They suggested that another epitope (SAAELVHFL) derived from *EPS8L2*, a gene highly expressed in the brain, might be responsible for the observed side effects.

ARDitox, when applied to analyze the initial target, has identified 294 putative OTEs with a high probability of being presented at the cell surface. Among them, 24 OTEs had a safety score below 3.0 as shown in Figure 2A. Specifically, putative OTE originating from *MAGEA12, MAGEA8, MAGEA9,* and *EPS8L2* genes were labeled by ARDitox as OTEs with a high potential to generate off-target toxicities (having the lowest possible score of 0). Furthermore, in *MAGEA12, MAGEA8,* and *EPS8L2* genes, the mRNA expression was observed in different parts of the brain.

Linette, et al. [7] described a off-target toxicity in a clinical study conducted on four patients diagnosed with myeloma and melanoma. The immunotherapy was directed against an epitope (*MAGEA3*, EVDPIGHLY) presented by HLA-A*01:01. The first two patients that received the therapy developed cardiogenic shock, which resulted in their death within the next few days. The off-target epitope was not identified prior to the clinical studies. Only after performing the experiments on cultured beating cardiomyocytes, a Titin (TTN) epitope was identified to be responsible for the toxicity effects.

Application of ARDitox allowed the identification of 84 potential OTEs, among which nine had a safety score below 3.0 as shown in Figure 2B. Importantly, the epitope originating from TTN was one of the top hits with a safety score of 0. The expression of TTN mRNA and the protein itself was found to be present in muscle and cardiac cells.

Stadtmaueret, et al. [38] performed a clinical study on 25 high-risk multiple myeloma patients using T cells engineered against NY-ESO-1 (SLLMWITQV), a Cancer-Testis Antigen (CTAs) with expression in multiple types of cancer. The TCR-engineered T-cells (TCR-T) against NY-ESO-1 are considered one of the most promising approaches for cancer immunotherapy with no adverse off-target toxicity detected and its high potential for the increase in patient survival.

This particular SLLMWITQV epitope is presented by HLA-A*02:01. For that peptide-HLA combination, ARDitox has detected 203 putative OTEs, off-target epitopes, however, out of all the OTEs only a single epitope derived from the *LRBA* gene (FLLMFIKQL) had a safety score below 3.0 as shown in Figure 2C.

Cai, et al. [39] tested pre-clinically engineered T cells against *AFP158* TAA(FMNKFIYEI). Based on an *in vitro* X-scan experiment, two off-target epitopes that could have activated the T-cell had been identified ENPP1436 and *RCL1*215. However, according to the authors, ENPP1436 is neither processed nor presented on the human cells.

Results provided by ARDitox showed 39 putative off-target epitopes for this target. As many as 8 of them were characterized by a safety score below 3.0. Among them, an experimentally identified epitope; *RCL1*215 was identified with a safety score of 2.47 indicating its high off-target toxicity potential as shown in Figures 2D and 2E. The mRNA expression of *RCL1* was found to be present across multiple tissues.

Rist, et al. [41,42] showed that a high proportion of CD8+ T cells against an EBV epitope from *BZLF1* (SELEIKRY) protein presented by HLA-B*18:01 cross-reacted with a human off-target epitope *CPSF3L* (DELEIKAY). The authors hypothesized that *BZLF1* is an example of molecular mimicry.

ARDitox identified 661 putative OTEs for SELEIKRY. Among them, DELEIKAY was found with a safety score of 3.94 and was defined as one of the top 15 cross-reactive peptides (Figures 2A-2E and Table 3).

**Table 2:** Target epitopes used for ARDitox validation, together with their properties and status obtained in previous studies.

| Case no. | Status | Targeted peptide | Off-target peptide(s) | HLA-type | Clinical status | Toxic side effects on the patients |
|---|---|---|---|---|---|---|
| 1 | Toxic | KVAELVHFL | KMAELVHFL | A*02:01 | Terminated after phase 2 | Mental status changes, comas, death |
| | | *MAGEA3* | MAGEA12 | | | |
| | | | SAAELVHFL *EPS8L2* | | | |
| 2 | Toxic | EVDPIGHLY | ESDPIVAQY | A*01:01 | Terminated after phase 2 | Myocardial damage, death |
| | | *MAGEA3* | *TTN* | | | |
| 3 | Safe | SLLMWITQV | None | A*02:01 | Completed phase 2 | No adverse effects |
| | | NY-ESO-1 | | | | |
| 4 | Unknown status | FMNKFIYEI | ILNKFIPDI | A*02:01 | Off-target based on X-scan | Not applicable |
| | | AFP | *RCL1* | | | |
| 5 | Mimicking | SELEIKRY | DELEIKAY | B*18:01 | Tested *in vitro* | Not applicable |
| | | Epstein-Barr *BZLF1* | *CPSF3L* | | | |
| 6 | Unknown status | NLDTLMTYV | SLDALITHV | A*02:01 | Preclinical studies | Not applicable |
| | | NLGN4X | *ADH1A* | | | |

**Note:** HLA: Human Leukocyte Antigen

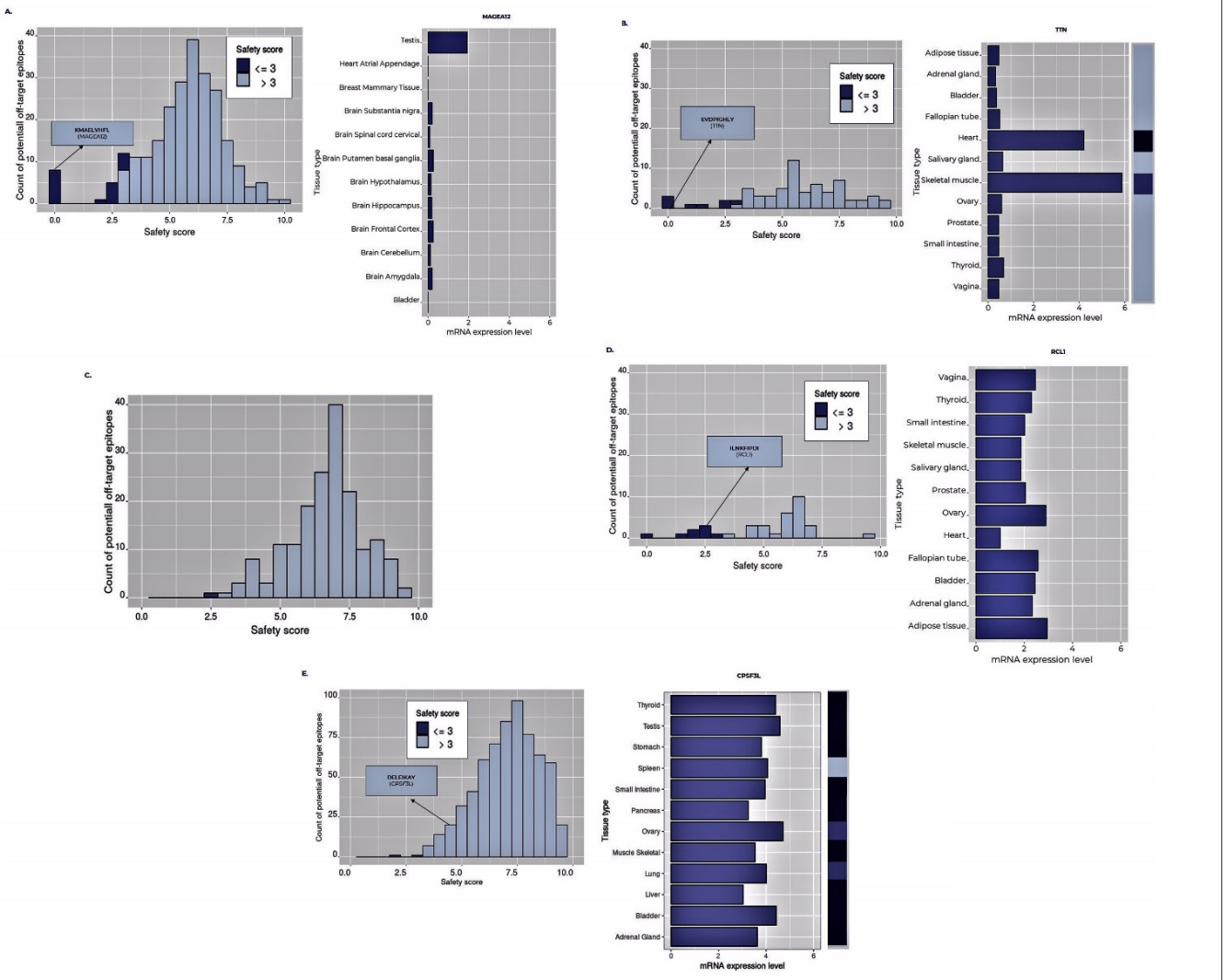


**Figure 2:** Distribution of ARDitox safety scores for putative off-target epitopes. (A): Case 1-*MAGEA3* (KVAELVHFL); (B): Case 2-*MAGEA3* (EVDPIGHLY); (C): Case 3-NY-ESO-1 (SLLMWITQV); (D): Case 4-AFP (FMNKFIYEI); (E): Case 5-Epstein-Barr *BZLF1* (SELEIKRY). If available mRNA (x=axis in logTPM) and peptide expression values of the genes that gave rise to the off-target epitope are presented. **Note:** (■): High peptide expression level; ( ): Medium peptide expression level; (■): Not detected

**Table 3:** Lack of presentation of epitope PLDPLITHV derived from *Adh1* a mouse ortholog gene of *ADH1A*.

| *Homo sapiens* | | | *Mus musculus* | | |
|---|---|---|---|---|---|
| **Presentation probability** | **Epitope sequence** | **Gene** | **Presentation probability** | **Epitope sequence** | **Gene** |
| 0.917 | NLDTLMTYV | *NLGN4X* | Not available | Not available | Not available |
| 0.929 | SLDALITHV | *ADH1A* | 0.138 | PLDPLITHV | *Adh1* |

Lastly, we validated ARDitox predictions in *in vitro* experiments by using NLGN4X131-139 (NLDTLMTYV) that has been reported as a promising recurrent TAA in glioblastoma. A more recent study reported a TCR targeting the NLGN4X epitope as part of IMA950 trial [43]. Thus, a prospective safety analysis was performed to evaluate the suitability of this epitope as a cell therapy target.

This particular epitope is presented by HLA-A*02:01. For that peptide-HLA combination, ARDitox was able to detect a single putative off-target epitope: NLGN4Y with a safety score below 3.0 and 16 additional putative off-target epitopes with a safety score below 5 as shown in Figures 3A-3D. It is worth noting that the on-target NLGN4X epitope itself is not reported as a target because it has a low presentation probability on healthy cells (<0.5) in contrast to its high expression on tumor cells. All the epitopes with a safety score <5 were further verified *in vitro* as described in the method section. This resulted in the identification of an off-target epitope with a score of 4.87 (SLDALITHV; *ADH1A*) that weakly activated the examined TCR (Figures 3A-3D).

## OTE trends

**TAA epitopes *vs.* virus epitopes:** We employed ARDitox to analyze 148 epitopes from TAA and viruses presented either by HLA-A*02:01 or HLA-A*01:01. As expected, due to the evolutionary distance between the tested peptides, the number of putative OTEs for TAA epitopes was higher (31854) than OTE peptides of viral origin (22279). The t-test conducted on the safety scores suggested a significant difference between the distributions (mean TAA=7.39, mean virus=7.62, p-value<2.2e-16). On the other hand, Cohen's d (-0.1514) suggested a rather negligible difference [44]. The above values, together with the similar shapes of distributions of TAA and virus OTE as shown in Figure 4A suggest that the distribution of safety scores is comparable between the two groups. However, if only OTEs with safety scores below 3 are considered, we see a 6.4-fold enrichment of TAA *vs.* viral epitopes.

**TAA epitopes *vs.* frameshift epitopes:** Lastly, ARDitox was tested on 16 frameshift epitopes and 16 randomly subsampled TAA peptides, which were predicted by our model to be presented by HLA-A*02:01. Importantly, only 336 putative OTEs were identified for the frameshift-derived fragments, while as many as 3911 putative OTEs were found in the TAA group. The t-test conducted on the safety scores was significant with a p-value=2.544e-06 (mean TAA=7.21, mean frameshift=7.67), while Cohen's d value was low (0.28) as shown in Figure 4B. Interestingly, no OTE derived from frameshift variants with a score below 3 was found (Figures 4A and 4B).

# DISCUSSION

Cellular cancer immunotherapy (e.g., TILs, Engineered TCRs) is a promising alternative or a complementary approach to surgery, radiotherapy, and chemotherapy in cancer treatment [45]. However, before it is fully embraced as a form of cancer treatment several shortcomings need to be addressed. One of the main issues is the adverse effect caused by off-target toxicity [46]. To help mitigate this problem we have introduced ARDitox, a novel method for analyzing potential cross-reactivity for a given pHLA, that includes the identification of off-target epitopes that differ significantly from the targeted epitope.

So far very few computational algorithms for predicting off-target binding in TCR-based cellular cancer immunotherapies have been proposed. Similarly, to ARDitox, Expitope and iVax approaches start with querying human proteome for peptides homologous to the target with a predefined number of mismatches allowed [47]. However, iVax does not consider the presentation of recognized putative OTEs in the process of ranking them and Expitope estimates peptide presentation on HLA with a proxy by the combination of the proteasomal cleavage probability and transporter associated with antigen processing as well as HLA binding. This means that ARDitox is the first available method leveraging a model for predicting peptide presentation by HLA molecules and unlike Expitope includes recognition of the TCR facing residues in order to evaluate the safety of the cross-reactive epitopes. Overall, ARDitox is a novel approach to OTEs identification with a unique pipeline that includes an in-house AI trained presentation model, a unique scoring function focused on physico-chemical properties of the TCR-facing residues and an extended search of peptides derived from frequent mutations.

We tested ARDitox on four clinically validated TCRs targeting TAA epitopes, one TCR in a preclinical stage and one virus epitope as well as on two datasets such as TAA *vs.* Virus epitopes and TAA *vs.* frameshift derived epitopes. In all analysis with reported side effects, ARDitox correctly identified the OTEs that caused the toxicity in the treated patients as shown in Table 4. Moreover, in case, where no risky OTEs were found experimentally, ARDitox found only one cross-reactive epitope with a safety score <3, which was in accordance with clinical trials. We have successfully identified an OTE that might lead to autoimmunity as a result of molecular mimicry after EBV infection, showing ARDitox's potential usefulness in the development of vaccines, while taking into account molecular mimicry. Lastly, we have experimentally found the *ADH1A* epitope, an off-target that would not be identified in mouse models due to the lack of its presentation, of the orthologue epitope derived from Adh1 (PLDPLITHV) making it impossible to find this OTE with this model. Importantly, *ADH1A* had a safety score close to 5, which corresponded with a weak binding of the TCR. Early identification of this OTE is valuable as additional safety measures can be considered to ensure that activation of the T cell against *ADH1A* will not occur during clinical trials (Table 4).

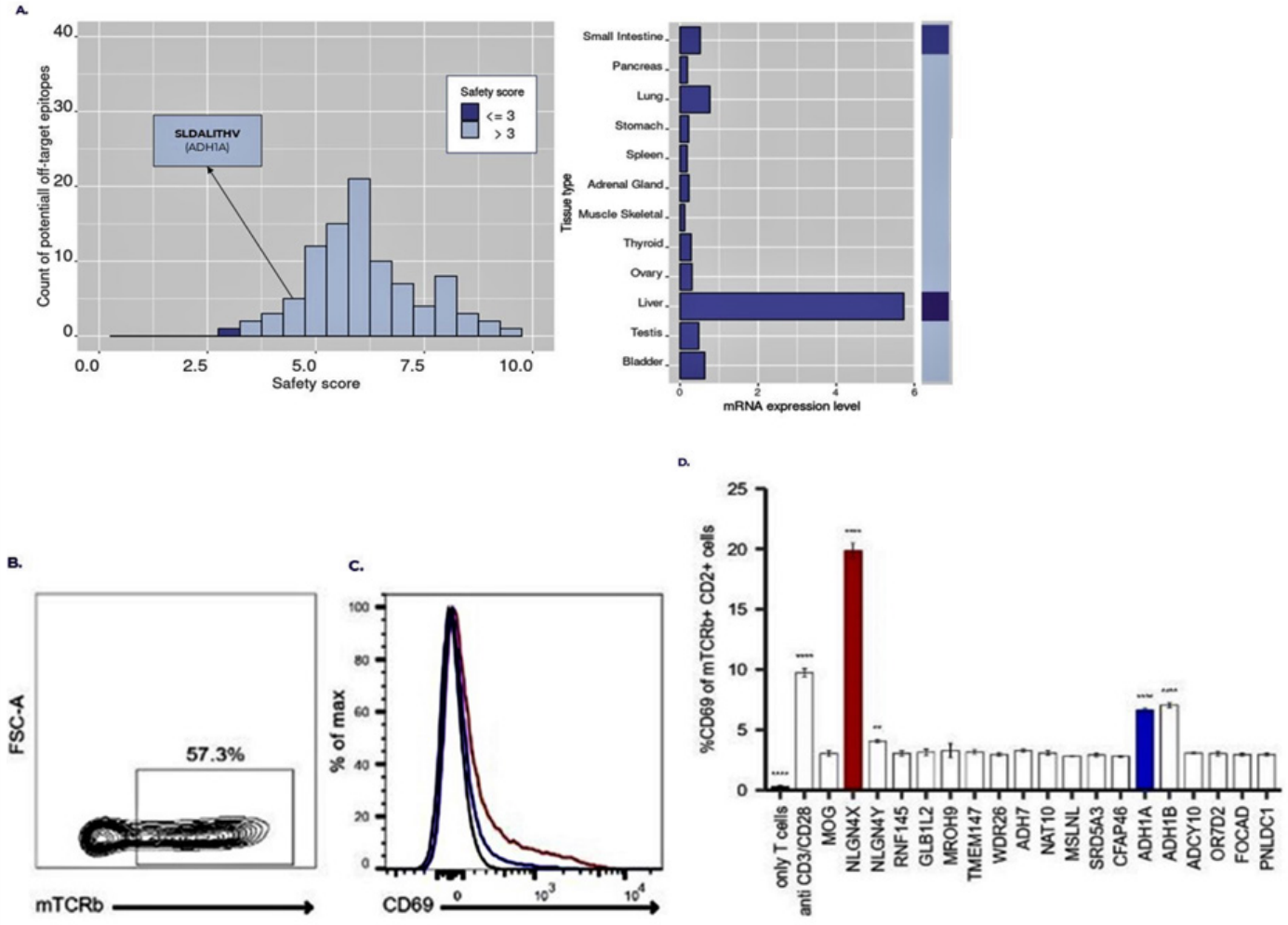


**Figure 3:** Results for case 6-NLGN4X (NLDTLMTYV). (A): Distribution of ARDitox safety scores for putative OTEs and expression plots for *ADH1A* OTE; (B): Representative flow cytometry density plot depicting transfection efficiency of murine TCR (mTCRb) in Jurkat cells; (C): Histogram depicting CD69 of NLGN4X TCR Jurkat T cells co-cultured with NLGN4X (red) or *ADH1A* peptide (blue)-pulsed presenter cells are overlayed on only T cell control (black); (D): CD69 levels of mTCRb+ Jurkat T cells (CD2+) after co-culture with presenter cells loaded with the indicated putative OTEs. Represented is the mean with SD of n=3 technical replicates. Significance calculated with one-way ANOVA in comparison to the Myelin Oligodendrocyte Glycoprotein (MOG) control peptide. **Note:** (■): High peptide expression level; (■): Medium peptide expression level; (□): Not detected

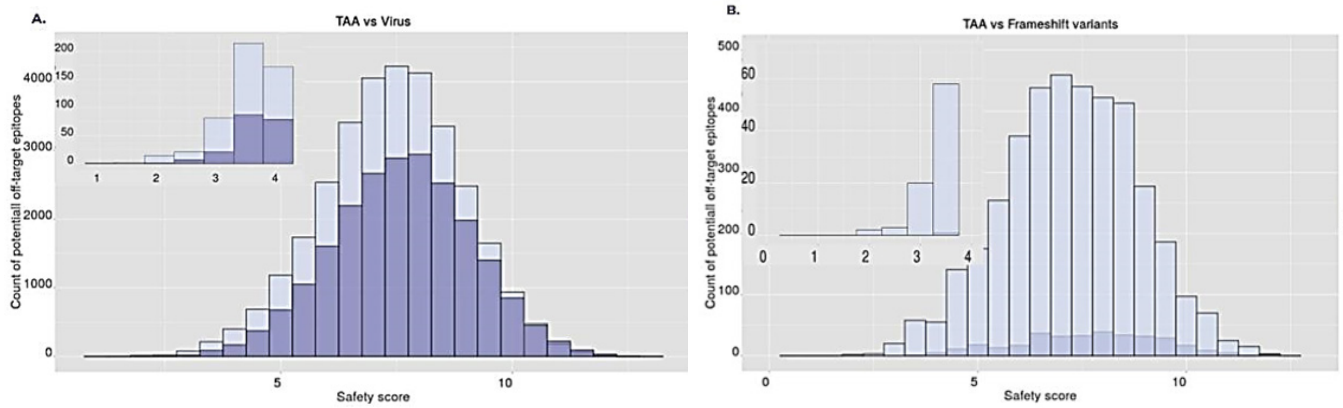


**Figure 4:** (A): Distribution of ARDitox safety scores of OTEs from 148 TAA epitopes and 148 virus epitopes presented on HLA-A*02:01 or HLA-A*01:01; (B): Distribution of ARDitox safety scores of OTEs from 16 TAA epitopes and 16 frameshift epitopes presented on HLA-A*02:01. **Note:** (■): TAA virus; (■): Frameshift variant TAA

**Table 4:** Estimation of targeted peptide toxicity by ARDitox based on three main variables.

| Experimental | | | ARDitox | | |
|---|---|---|---|---|---|
| **Case no. and status** | **Targeted peptide** | **OTE** | **No. OTE with safety scores <3** | **No. of all OTE** | **RNA expression of the off-target peptide** |
| 1) Toxic | KVAELVHFL<br>*MAGEA3* | KMAELVHFL<br>MAGEA12<br>SAAELVHFL *EPS8L2* | 18 | 294 | Low expression in the brain |
| 2) Toxic | EVDPIGHLY<br>*MAGEA3* | ESDPIVAQY<br>TTN | 8 | 84 | High expression in muscle and heart |
| 3) Safe | SLLMWITQV<br>NY-ESO-1 | NA | 1 | 203 | Not applicable |
| 4) Preclinical studies | FMNKFIYEI<br>AFP | ILNKFIPDI<br>*RCL1* | 8 | 39 | Equally expressed across all tissues |
| 5) Probably leads to autoimmune disease | SELEIKRY<br>Epstein-Barr *BZLF1* | DELEIKAY<br>*CPSF3L* | 2 | 661 | Equally expressed across several tissues |
| 6) Probably safe | NLDTLMTYV<br>NLGN4X | KLDSLMTLL<br>*ADH1A?* | 3 | 97 | Equally expressed across several tissues. Protein expressed in small intestine, liver and duodenum |

**Note:** RNA: Ribonucleic Acid

When using ARDitox to assess the risk of the therapy, we strongly recommend to check three variables addressed by the software such as number of all OTEs, number of OTEs with a safety score <3 and expression of OTEs with a safety score <3. Furthermore, we highly recommend testing *in vitro* all OTEs with a score <5 as some weak TCR off-target interactions might have negative consequences for the patient's health. It should be stressed out that the importance of the variables mentioned above should not be neglected as exemplified by TTN's OTE in use. The number of all putative OTEs was considered as moderate as only 84 were identified. Based on this variable, the TCR against this target epitope seemed to be very promising, however, when the distribution of the safety scores was verified, it turned out that ~ 10% of the putative OTEs had a score <3. The cross-reactivity of each of these OTEs should be checked experimentally because low safety scores indicate that the TCR may bind to both the target and the OTE in a similar fashion. Lastly, checking the expression status of each putative OTE with a safety score <3 should indicate which tissue types are of particular interest for the experimental verification. This would have been important, as during the preclinical *in vitro* studies no toxicity towards the tested heart muscle cell line was detected, as TTN protein is expressed only in contracting cardiac myocytes. Identifying TTN as OTE upfront could have enforced the addition of appropriate cell lines to the test panel. The potential shortcoming of our model is that currently, for some less frequent HLA-types, incorrect amino acids may be scored as the ones facing the TCR. However, this problem is minor as it occurs only for HLA-types that have generally low frequency. Furthermore, it can be mitigated as more data regarding TCR-faced amino acid positions for rare HLAs becomes available.

In order to assess the effectiveness of the proteome search for OTEs and the proposed scoring methodology, we compared the analysis performed on TAAs and viral epitopes. The dataset used was composed of an equal number of TAA and viral derived epitopes, presented by either HLA-A*02:01 or HLA-A*01:01. As expected, we saw fewer hits from viral epitopes in comparison to the number of OTE found for TAA epitopes, because viral proteins differ substantially from proteins present in the human reference genome. This indicates that the first step of the ARDitox pipeline works efficiently. On the other hand, the overall safety score distribution is similar between both groups, with a Cohen's d equal to -0.15. As expected, the mean safety score is lower for TAA epitopes, but the difference between means is negligible (0.23). However, if only OTEs with safety scores <3 are considered, we see a 6.4-fold enrichment of TAA (64 TAA OTEs *vs.* 10 Viral OTEs) which is a much higher ratio when compared to 1.43 (31854 TAA OTEs/22279 Viral OTEs) obtained when results with all values of the safety score are concerned. These results are in line with our previous suggestion regarding the interpretation of the ARDitox results. The main focus, when assessing the risk of the target causing off-target toxicity, should be emphasized on the verification of the number of putative OTEs with a safety score <3.

Lastly, we wanted to check whether frameshift mutations are promising targets for immunotherapeutic strategies, since they give a rise to multiple, out-of-frame, random protein products that should not map to the reference proteome [48]. Furthermore, frameshift derived epitopes usually do not share functional domains with other genes found in the human genome and as such the general number of putative OTEs should be both

lower and with higher safety scores. In order to verify this, we used a database composed of 16 frameshift-derived neoepitopes that were predicted as presented by our presentation model. When compared to 16 TAA epitopes, the number of OTEs from frameshift neoepitopes was 10X lower (frameshift OTEs: 336 *vs.* TAA OTEs: 3911). Furthermore, none of the putative frameshift OTEs had a safety score <3. As such, ARDitox results strongly confirm that, as long as nonsense-mediated decay does not occur for a particular variant, frameshift neoepitopes are safe and promising alternatives to TAA epitopes in TCR therapies.

## CONCLUSION

In conclusion, we have developed a method for the identification of off-target toxicity that can be successfully applied in the development of cellular immunotherapies. Our tool, ARDitox, takes into account peptide processing, pHLA binding, pHLA presentation probability, determination and similarity of TCR-faced amino acids, frequent variants as a source of off-target epitopes and gene mRNA and protein expression levels. The potential shortcoming of our model is that currently, for some less frequent HLA types, incorrect amino acids may be scored as the ones facing the TCR, however, this problem is minor as it occurs only for HLA types that have generally low frequency. Furthermore, with time, this issue can be mitigated as more data regarding TCR-faced amino acid positions for rare HLAs becomes available. Most importantly, the application of our platform, ARDitox, to process data from several use case studies allowed efficient identification of OTEs, which proves its applicability in the development of TCR-based cancer immunotherapies.

## REFERENCES

1. Koebel CM, Vermi W, Swann JB, Zerafa N, Rodig SJ, Old LJ, et al. Adaptive immunity maintains occult cancer in an equilibrium state. Nature. 2007;450(7171):903-907.
2. Thomas R, Al-Khadairi G, Roelands J, Hendrickx W, Dermime S, Bedognetti D, et al. NY-ESO-1 based immunotherapy of cancer: Current perspectives. Front Immunol. 2018;9:947.
3. Dong S, Ghobrial IM. Immunotherapy for hematological malignancies. J Life Sci. 2019;1(1):46.
4. Farkona S, Diamandis EP, Blasutig IM. Cancer immunotherapy: The beginning of the end of cancer? BMC Med. 2016;14(1):1-8.
5. Duan Z, Ho M. T-cell receptor mimic antibodies for cancer immunotherapy. Mol Cancer Ther. 2021;20(9):1533-1541.
6. Riddell SR. Engineering antitumor immunity by T-cell adoptive immunotherapy. ASH Educ. 2007;2007(1):250-256.
7. Linette GP, Stadtmauer EA, Maus MV, Rapoport AP, Levine BL, Emery L, et al. Cardiovascular toxicity and titin cross-reactivity of affinity-enhanced T cells in myeloma and melanoma. Blood. 2013;122(6):863-871.
8. Coles CH, McMurran C, Lloyd A, Hock M, Hibbert L, Raman MC, et al. T cell receptor interactions with human leukocyte antigen govern indirect peptide selectivity for the cancer testis antigen MAGE-A4. J Biol Chem. 2020;295(33):11486-11494.
9. Cruz-Tapias PC, Anaya JM. Major histocompatibility complex: Antigen processing and presentation. 2013.
10. Newey A, Griffiths B, Michaux J, Pak HS, Stevenson BJ, Woolston A, et al. Immunopeptidomics of colorectal cancer organoids reveals a sparse HLA class I neoantigen landscape and no increase in neoantigens with interferon or MEK-inhibitor treatment. J Immunother Cancer. 2019;7(1):309.
11. de Greef PC, Oakes T, Gerritsen B, Ismail M, Heather JM, Hermsen R, et al. The naive T-cell receptor repertoire has an extremely broad distribution of clone sizes. Elife. 2020;9:e49900.
12. Qi Q, Liu Y, Cheng Y, Glanville J, Zhang D, Lee JY, et al. Diversity and clonal selection in the human T-cell repertoire. Proc Natl Acad Sci. 2014;111(36):13139-13144.
13. Morgan RA, Chinnasamy N, Abate-Daga DD, Gros A, Robbins PF, Zheng Z, et al. Cancer regression and neurologic toxicity following anti-MAGE-A3 TCR gene therapy. J Immunother. 2013;36(2):133-151.
14. Dhanik A, Kirshner RJ, MacDonald D, Thurston G, Lin CH, Murphy AJ, et al. In-silico discovery of cancer-specific peptide-HLA complexes for targeted therapy. BMC Bioinform. 2016;17:286.
15. Bateman A, Martin MJ, Orchard S, Magrane M, Agivetova R, Ahmad S, et al. UniProt: The universal protein knowledgebase in 2021. Nucleic Acids Res. 2021;49(D1):D480-D489.
16. Kwak SH, Chae J, Choi S, Kim MJ, Choi M, Chae JH, et al. Findings of a 1303 Korean whole-exome sequencing study. Exp Mol Med. 2017;49(7):e356.
17. Chen S, Francioli LC, Goodrich JK, Collins RL, Kanai M, Wang Q, et al. A genome-wide mutational constraint map quantified from variation in 76,156 human genomes. BioRxiv. 2022:2022-2303.
18. Mazzocco G, Niemiec I, Myronov A, Skoczylas P, Kaczmarczyk J, Sanecka-Duin A, et al. AI aided design of epitope-based vaccine for the induction of cellular immune responses against SARS-CoV-2. Front Genet. 2021;12:602196.
19. Abelin JG, Keskin DB, Sarkizova S, Hartigan CR, Zhang W, Sidney J, et al. Mass spectrometry profiling of HLA-associated peptidomes in mono-allelic cells enables more accurate epitope prediction. Immunity. 2017;46(2):315-326.
20. Di Marco M, Schuster H, Backert L, Ghosh M, Rammensee HG, Stevanovic S. Unveiling the peptide motifs of HLA-C and HLA-G from naturally presented peptides and generation of binding prediction matrices. J Immunol. 2017;199(8):2639-2651.
21. Sarkizova S, Klaeger S, Le PM, Li LW, Oliveira G, Keshishian H, et al. A large peptidome dataset improves HLA class I epitope prediction across most of the human population. Nat Biotechnol. 2020;38(2):199-209.
22. O'Donnell TJ, Rubinsteyn A, Bonsack M, Riemer AB, Laserson U, Hammerbacher J. MHCflurry: Open-source class I MHC binding affinity prediction. Cell Syst. 2018;7(1):129-132.
23. Moise L, Gutierrez AH, Bailey-Kellogg C, Terry F, Leng Q, Abdel Hady KM, et al. The two-faced T cell epitope: Examining the host-microbe interface with JanusMatrix. Hum Vaccin Immunother. 2013;9(7):1577-1586.
24. Bijen HM, van der Steen DM, Hagedoorn RS, Wouters AK, Wooldridge L, Falkenburg JF, et al. Preclinical strategies to identify off-target toxicity of high-affinity TCRs. Mol Ther. 2018;26(5):1206-1214.
25. Border EC, Sanderson JP, Weissensteiner T, Gerry AB, Pumphrey NJ. Affinity-enhanced T-cell receptors for adoptive T-cell therapy targeting MAGE-A10: Strategy for selection of an optimal candidate. Oncoimmunology. 2019;8(2):e1532759.
26. Cameron BJ, Gerry AB, Dukes J, Harper JV, Kannan V, Bianchi FC, et al. Identification of a titin-derived HLA-A1 presented peptide as a cross-reactive target for engineered MAGE A3 directed T cells. Sci Trans Med. 2013;5(197):197ra103.

27. Karapetyan AR, Chaipan C, Winkelbach K, Wimberger S, Jeong JS, Joshi B, et al. TCR fingerprinting and off-target peptide identification. Front Immunol. 2019;10:2501.

28. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The genotype-tissue expression (GTEx) project. Nat Genet. 2013;45(6):580-585.

29. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Tissue-based map of the human proteome. Science. 2015;347(6220):1260419.

30. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. J Mach Learn Res. 2011;12:2825-30.

31. Reback J, McKinney W, jbrockmendel, Bossche J Van den, Augspurger T, Cloud P, et al. Pandas-dev/pandas. 2021.

32. Harris CR, Millman KJ, Van Der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. Nature. 2020;585(7825):357-362.

33. Wickham H. Data analysis. Springer. 2016.

34. A grammar of data manipulation. 2023.

35. BSgenome: Software infrastructure for efficient representation of full genomes and their SNPs version 1.58.0 from bioconductor. 2023.

36. Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, et al. The immune epitope database (IEDB): 2018 update. Nucleic Acids Res. 2019;47(D1):D339-D343.

37. Koster J, Plasterk RH. A library of Neo Open Reading Frame peptides (NOPs) as a sustainable resource of common neoantigens in up to 50% of cancer patients. Sci Rep. 2019;9(1):6577.

38. Stadtmauer EA, Faitg TH, Lowther DE, Badros AZ, Chagin K, Dengel K, et al. Long-term safety and activity of NY-ESO□1 spear T cells after autologous stem cell transplant for myeloma. Blood Adv. 2019;3(13):2022-2034.

39. Cai L, Caraballo Galva LD, Peng Y, Luo X, Zhu W, Yao Y, et al. Preclinical studies of the off-target reactivity of *AFP158*-specific TCR engineered T cells. Front Immunol. 2020;11:607.

40. Dutoit V, Herold-Mende C, Hilf N, Schoor O, Beckhove P, Bucher J, et al. Exploiting the glioblastoma peptidome to discover novel tumour-associated antigens for immunotherapy. Brain. 2012;135(4):1042-1054.

41. Rist MJ, Hibbert KM, Croft NP, Smith C, Neller MA, Burrows JM, et al. T cell cross-reactivity between a highly immunogenic EBV epitope and a self-peptide naturally presented by HLA-B* 18:01+ cells. J Immunol. 2015;194(10):4668-4675.

42. Martin AD, Wang X, Sandberg ML, Negri KR, Wu ML, Warshaviak DT, et al. Re-examination of MAGE-A3 as a T-cell therapeutic target. J Immunother. 2021;44(3):95-105.

43. Hilf N, Kuttruff-Coqui S, Frenzel K, Bukur V, Stevanovic S, Gouttefangeas C, et al. Actively personalized vaccination trial for newly diagnosed glioblastoma. Nature. 2019;565(7738):240-245.

44. Sawilowsky SS. New effect size rules of thumb. J Mod Appl Stat Methods. 2009;8(2):597-599.

45. Hunter P. The fourth pillar: Despite some setbacks in the clinic, immunotherapy has made notable progress toward becoming an additional therapeutic option against cancer. EMBO Rep. 2017;18(11):1889-1892.

46. García-Fernandez C, Saz A, Fornaguera C, Borros S. Cancer immunotherapies revisited: State of the art of conventional treatments and next-generation nanomedicines. Cancer Gene Ther. 2021;28(9):935-946.

47. Jaravine V, Mosch A, Raffegerst S, Schendel DJ, Frishman D. Expitope 2.0: A tool to assess immunotherapeutic antigens for their potential cross-reactivity against naturally expressed proteins in human tissues. BMC Cancer. 2017;17(1):892.

48. Spaanderman IT, Peters FS, Jongejan A, Redeker EJ, Punt CJ, Bins AD. Framing the potential of public frameshift peptides as immunotherapy targets in colon cancer. PloS One. 2021;16(6):e0251630.