AI-based tools for target identification foster the generation of novel TCR hits against solid tumor antigens

Anna Sanecka-Duin, Paweł Biernat, Weronika Czarnocka, Krystyna Dudaniec, Oskar Gniewek, Łukasz Grochowalski, Maciej Jasiński, Małgorzata Kamuda, Bartłomiej Król-Józaga, Joanna Marczyńska-Grzelak, Giovanni Mazzocco, Natalia Ochocka, Victor Murcia Pienkowski, Piotr Skoczylas, Sławomir Stachura, Rafał Szatanek, Jan Kaczmarczyk, Agnieszka Blum

Ardigen, Kraków, Poland

Introduction

T-cells are an essential component of adaptive immunity capable of recognizing pathogens and malignant cells. They use T cell receptors (TCRs) to screen antigens presented by HLA molecules on a cell surface. Upon recognition of a tumor-derived antigen, such as neoantigen encompassing a mutation site or peptide expressed at an abnormally high level, T cells can detect ongoing tumorigenesis and trigger an anti-tumor immune response. This natural mechanism is utilized in the development of cancer vaccines and TCR-based immunotherapies for cancer. The key to successful therapy, however, is identifying the right target antigen.

Here we present the **ARDentify platform**, a set of highly accurate in silico predictive tools for selecting clinically relevant cancer antigens [Fig. 1]. The platform allows identification population-wide or patient-specific targets and estimation of the number of patients who may benefit from a given targeted therapy. It consists of the **ARDisplay presentation model,** which calculates the probability of antigen presentation in complex with HLA class I or class II, with more than 2 times higher average precision compared to other approaches. In addition, the **ARDitox module** allows the exclusion of candidates with a high risk of off-target toxicity. The presented solution aims to streamline the development of therapies for potential candidates, thereby minimizing both time and cost.

Methodology

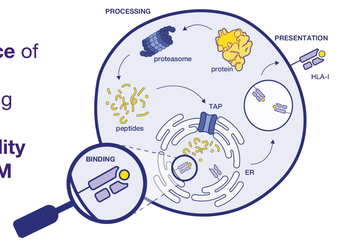
We applied the ARDentify platform to identify population-wide antigen candidates expressed in solid tumors using publicly available datasets. We derived neoantigens (tumor specific antigens, TSAs) from 25 TCGA cohorts (8197 patients) and tumor associated antigens (TAA) derived from proteins with abnormal expression in tumor tissue (50 cohorts from 164 histological studies) [Fig. 2a]. We shortlisted the antigen candidates using our ARDisplay and ARDitox modules and **designed a panel of 34 peptide HLA I complexes (pHLAs)** [Fig. 2b]. To identify TCRs that recognize selected antigens, we employed oligo-barcoded multimers of the selected pHLAs. Using the multimers, we isolated antigen-binding cells from healthy donors CD8$^+$ T cells and performed single-cell sequencing of TCRs and their associated antigen barcodes [Fig. 3]. This allowed us to identify pHLA:TCR pairs in a high-throughput manner and obtain TCR hits for all pHLA targets tested, yielding a total of **4140 full-length TCR sequences** [Fig. 4 and 5].

1 ARDentify platform for safe cancer immunotherapies

ARDisplay model: prediction of peptide presentation via HLA-I & -II

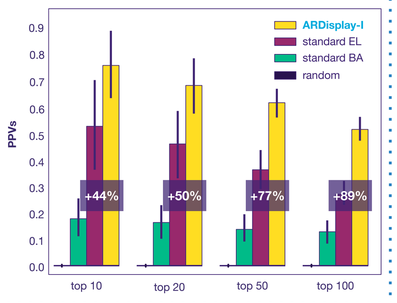
Distinctive features of ARDisplay

- Overall **2x higher predictive performance** of our models as compared to standard solutions for eluted ligand (EL) and binding affinity (BA) prediction
- Models are trained on selected **high-quality sources** of MS EL data (including over **2M unique peptides**, **182 HLA** alleles)



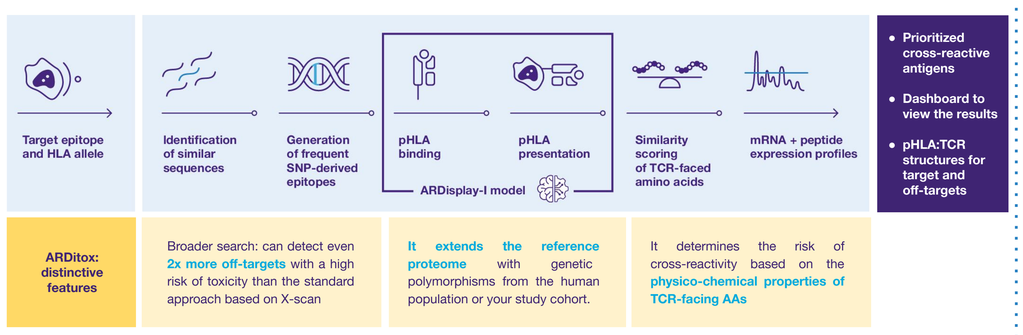
Ardigen's HLA-I presentation model compared to standard solutions for presentation prediction on **peptides from intracellular proteins.**

Positive predictive values (PPVs) at four distinct thresholds, ie. top-10, 20, 50, & 100 pHLA pairs selected by each method. Random resampling on observations from the Löffler dataset was done for 100 times to assess metrics variability, i.e. get an average performance and its' standard deviations. Benchmark on 22 cancer patients[1] with CRC (colorectal cancer, adenocarcinomas). It represents a wide range of HLA alleles and contains high-quality MS EL data.

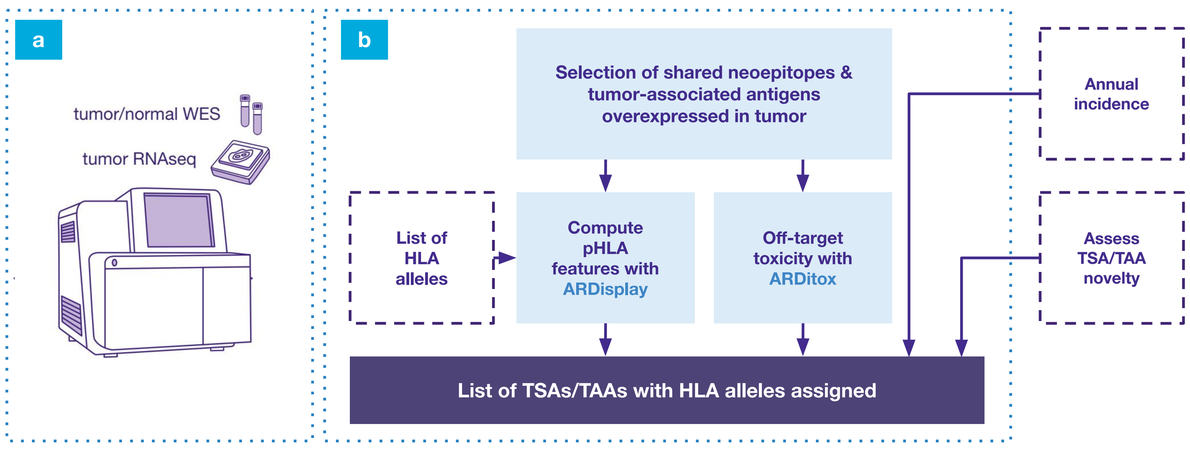


ARDitox model: deep in silico screening of potential cross-reactive epitopes

- **Experimentally validated** in collaboration with DKFZ. Results from Ardigen's ARDitox platform were validated in vitro, confirming the potential to generate immunotoxicity by the epitope detected as a putative off-target[2].
- Patent pending (EP22461636).



2 ARDentify platform was used to select promising population-wide TCR therapy targets for solid tumors



GENE ZBTB20 (ZBTB20_N10)		MUTATION UNDISCLOSED	PEPTIDE UNDISCLOSED	HLAs A02:01[1]
Incidence	15 411[2]	**Zinc Finger and BTB Domain Containing 20**, a transcription factor that acts as a transcriptional repressor. It is involved in processes such as neurogenesis, glucose homeostasis, and postnatal growth.		
Indication	CRC	**Role in carcinogenesis:** It has been identified as a potential oncogene in liver cancer - it interacts with PPARG to upregulate the WNT/CTNNB1 signaling pathway, leading to tumorigenesis [ref]. In gastric cancer, (...) [ref].		
ARDisplay-I presentation probability	0.86	**Patents:** *Adoptive cell therapy with ZBTB20 suppression* [ref] [3]		
Driver mutation	No	**Clinical trials:** None		
Fraction of patients eligible	0.90	[1] target can be presented via various HLA alleles, only selected ones are shown [2] easily assess the number of potential therapy recipients yearly; optionally additional breakdown by USA, Europe, and Asia can be provided [3] not related to T-cell therapies [ref] UNDISCLOSED		
ARDitox putative off-targets	1			

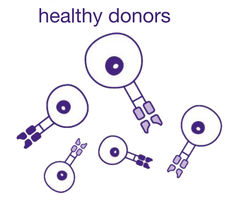
25 TCGA cohorts (8197 patients) screened for TSAs
50 cohorts from 164 histological studies screened for TAAs → **Selected: 26 TAAs and 8 TSAs**

NGS data → **ARDentify platform** (incl. ARDisplay & ARDitox models) → **Clinically relevant annotations** → **pHLA targets** with summary tables

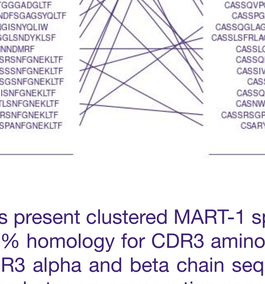
3 Experimental set-up for selection of specific TCRs from healthy donors

1. pHLA targets on HLA dextramers® (by ***Immudex***)

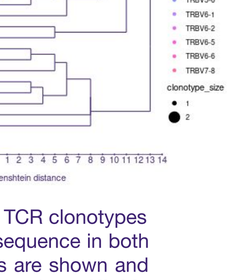
2. CD8$^+$ T-cells from healthy donors



3. TCR:pHLA pairs sorting



4. single-cell sequencing of T-cells (10x Genomics)

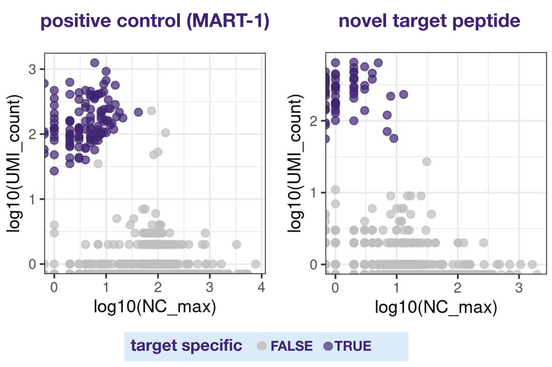


1. Synthesis pHLA dextramers® for 34 selected targets. Each dextramer is conjugated to a fluorophore and a unique oligonucleotide barcode.
2. CD8$^+$ T cells collected from 13 healthy individuals expressing 4 major HLAs of interests: A02:01, A03:01, A11:01 and A24:02.
3. FACS sorting of T cells recognizing pHLAs selected for particular donors based on their HLA repertoire.
4. single cell RNA sequencing of sorted T cells on 10x Genomics platform.

4 TCRs hits against 34 cancer epitopes found among CD8$^+$ T-cells from 13 healthy donors

	Gene target	TCR hits	HLA allele(s)
top population-wide targets (incl. novel peptides)	#1	342	A02:01, A03:01, A24:02
	#2	278	A02:01, A24:02
	#3	600	A03:01, A11:01
	#4	52	A02:01
	#5	406	A02:01, A03:01
	#6	11	A02:01
	#7	2	A02:01
other	13 other antigens	1 305	A02:01, A03:01, A11:01, A24:02
positive control	MART-1	1 144	A02:01
	Total	**4 140**	

5 Identifying T-cells targeting tumor-specific antigens: multistep verification of binding specificity



The scatterplots showing relation of number of unique molecular identifier (UMI) associated with the target dextramer (UMI_count) andthe dextramer of the negative control (max. of any of the 4 negative controls - NC_max). Each dot represents a single cell. UMI threshold for **TRUE** T-cells that specifically bound target pHLA were identified based on **UMI_count >15**, and **UMI count > 5 x NC_max**.

High clonotypes similarity between identified MART-1-binding TCR clones

The paired dendrograms present clustered MART-1 specific TCR clonotypes with high similarity (>80% homology for CDR3 amino acid sequence in both TRA and TRB). The CDR3 alpha and beta chain sequences are shown and the Levenshtein distance between consecutive sequences is given. **Highly similar clusters** were found **among rare clonotypes**. The high homology clusters are 7 fold more frequent within single specificity than across different specificities.

CONCLUSIONS:

We selected **34 tumor-specific pHLA antigens** (including novel peptides) derived from 20 gene targets annotated as **clinically relevant** on a **population-wide** basis. The database of **4140 full-length TCR sequences** that **bind specifically** to our pHLA targets was generated.

The **ARDentify platform** has demonstrated a high success rate in the experimental identification of TCRs for selected antigenic targets. We have shown that the **identification of novel TCRs** for tumor-specific antigens among rare CD8$^+$ T cells clonotypes **from healthy donors** is feasible and allows the generation of a large number of pHLA:TCR pairs that can be used in the development of cancer immunotherapy.

Bibliography:

1. Löffler MW, Kowalewski DJ, Backert L, et al. *Mapping the HLA Ligandome of Colorectal Cancer Reveals an Imprint of Malignant Cell Transformation*. Cancer Res. 2018;78(16):4627-4641. doi:10.1158/0008-5472.CAN-17-1745.
2. Murcia Pienkowski V, Boschert T, Skoczylas P, Sanecka-Duin A et al. *ARDitox: platform for the prediction of TCRs potential off-target binding* BioRXiv 2023 doi:https://doi.org/10.1101/2023.04.11.536336

Acknowledgements:

This work was partially finance by the following grants: "Creation of an innovative TCRact based on AI (artificial intelligence) IN SILICO technology to launch a NEW SERVICE of designing and optimizing T lymphocyte receptors (TCR) for use in cancer immunotherapy", co-financed by the European Regional Development Fund as part of Smart Growth Operational Programme 2014-2020, Submeasure 1.1.1.: Industrial research and development work implemented by enterprises, grant no: POIR.01.01.01-00-0019/20

European Funds Smart Growth · Republic of Poland · NCBR National Centre for Research and Development · European Union European Regional Development Fund

Contact:

Anna Sanecka-Duin, PhD
Lead Scientist Immunology
anna.sanecka-duin@ardigen.com

Agnieszka Blum, MD, PhD
Executive Vice President,
General Director of Immunology Unit
agnieszka.blum@ardigen.com