Streamlining design of safe and effective TCR therapies with AI

Mikołaj Mizera[1], Anna Sanecka-Duin[1], Oskar Gniewek[1], Łukasz Grochowalski[1], Maciej Jasiński[1], Paulina Król[1], Giovanni Mazzocco[1], Victor Murcia Pieńkowski[1], Oleksandr Myronov[1,2], Iga Niemiec[1], Piotr Skoczylas[1], Sławomir Stachura[1], Piotr Stępniak[1], Daniel Wojciechowski[1], Jan Kaczmarczyk[1], Agnieszka Blum[1]

1. Ardigen, Kraków, Poland
2. Faculty of Mathematics and Information Science, Warsaw University of Technology, Warsaw, Poland

contact: **jan.kaczmarczyk@ardigen.com**

See additional **interactive data visualizations** tcract.ardigen.com

OUR CHALLENGE

To develop a robust, self-explanatory predictive model for the rational design of pHLA-specific TCRs and validate it with clinical data using our oncological pHLA:TCR database.

BACKGROUND

Adoptive cell therapies with T-cells expressing engineered TCRs are one of the most promising approaches to cancer therapy [1]. Predicting TCR binding to a target antigen and TCR off-target toxicity [2] can provide insights supporting the development of safe therapies.

We see the opportunity to address the problem of pHLA:TCR binding prediction by biologically inspired modeling that incorporates domain knowledge about conservation of TCR sequential motifs and general patterns of interactions at the interface of peptide, HLA, and TCR:

- peptide N- and C-termini are located in the pockets A and F of HLA-I, respectively [3, 4],
- TCR **α**- and **β**-chain are located closer to the peptide N- and C-termini, respectively [5, 6],
- TCR **α**- and **β**-chain are fixed over **α**2 and **α**1 of HLA, respectively [6, 7],
- CDR1 & CDR2 loops interact mostly with HLAs while CDR3 interacts with peptides [7, 8].

METHODS

Building the oncological pHLA:TCR database

- We collect samples in observational clinical trial NCT04994093 and sequence them
- Immunogenic epitopes are predicted by Ardigen's ArdImmune Vax platform [9,10] and tested in vitro
- T-cells binding to immunogenic epitopes are used to generate TCR single-cell sequencing data with paired **α**- and **β**-chains
- After completing the clinical study, we will use our oncological database to externally validate and further improve our model of pHLA:TCR interaction

Dataset and TCR clustering

- We use pHLA:TCR binding dataset with 50k+ **α**-**β** paired TCR clonotypes and information on binding to 44 pHLAs [11]
- The dataset is curated to extract diverse observations and efficiently train a pHLA:TCR interaction model
- We cluster TCRs by properties of paired and unpaired **α**- and **β**-chains
- We used several clustering strategies to group similar TCRs by conserved properties:
  - Strategy A: use CDR3 length and V, J alleles as clustering criteria
  - Strategy B: use CDR3 length and V allele as clustering criteria
  - Strategy C: use CDR3 length as clustering criterion
- We characterize clusters by number of distinct pHLAs to which TCRs bind

Structural priors

- We curate 134 crystal structures of pHLA:TCR complexes
- We compute probability distributions of close-contacts between peptide, HLA, and TCR and analyze discovered patterns of these interactions

pHLA:TCR interaction modeling and validation

- We built a sequential pHLA:TCR interaction model regularized by structural information
- We use TCR clustering results to train and validate the model to generalize to diverse groups of TCRs and peptides

RESULTS

Dataset and TCR clustering

- Restrictive clustering identifies groups of TCRs that posses conserved properties and share pHLA binding specificity (Figure 2, strategy A)
- Inclusive clustering strategy results in clusters of TCRs that bind to more diverse sets of pHLAs (Figure 2, strategy C)
- Both CDR3a and CDR3b are important for TCR's binding specificity (Figure 2, strategy A-C)
- We trained sequence-based tree models to predict pHLA:TCR binding for specific pHLAs
- Validation strategy based on clusters of TCRs results in lower, but more realistic and less biased metric values than baseline random validation (Figure 3)

Structural priors

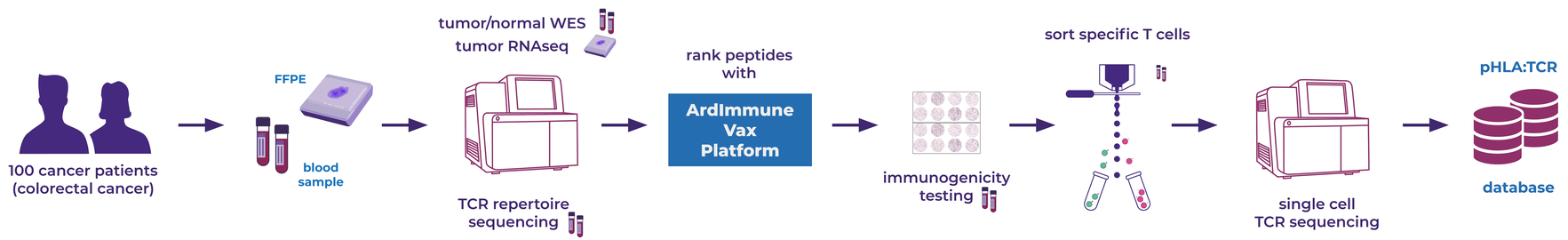
- Probability of interaction in pHLA:TCR complex is the highest for HLA and peptides with peaks at terminal positions (Figure 4, violet distribution)
- CDR3α and CDR3β chains interact with peptides at positions closer to N-terminal and C-terminal respectively (Figure 4, violet and red distributions)

pHLA:TCR interaction modeling and validation

- We trained pHLA:TCR interaction convolutional model using model regularization according to structural priors and TCR conservation patterns
- The model shows good performance (ROC AUC ~0.6) in pHLA:TCR binding prediction for a diverse set of viral- and cancer-origin antigens (Figure 5)
- In comparison with published models, our AI system shows superior robustness in the binding prediction of novel pHLA:TCR complexes [12-14]

Scan the QR code to explore the interactive data visualizations and learn more about our pHLA:TCR interaction model!

Figure 1. Sequences for oncological pHLA:TCR database can be collected more efficiently thanks to use of ArdImmune Vax Platform for identification of immunogenic epitopes.



Get familiar with our robust AI-based

validated model of pHLA:TCR interaction

that streamlines the rational design of novel TCR therapies

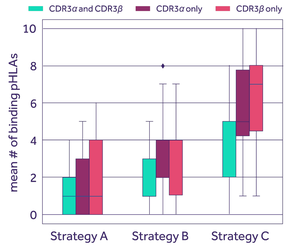


Figure 2. TCRs with conserved properties show similar pHLA binding profiles, i.e. TCRs in a cluster bind more specifically.

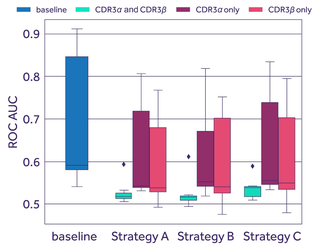


Figure 3. Our validation strategy splits significantly different TCRs between training and validation sets. The procedure results in inferior, but more accurate, statistical characteristics of the model's predictive performance.

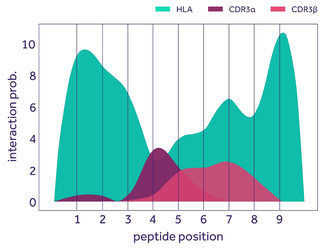


Figure 4. Peaks in probability distribution show conservation of favored interaction interfaces between peptide and HLA, CDR3α, CDR3β.

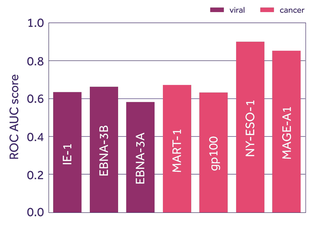


Figure 5. The model can generalize learned pHLA:TCR interaction patterns to predict the binding probability of novel pHLA:TCR complexes based on their sequences.

REFERENCES

1. S. Farkona, et al., *BMC Med.* 2016, *14*, 73.
2. V. A. Murcia Pienkowski, et al. *Cytotherapy* 2021, *23*, S96.
3. K. C. Garcia, et al. *Nat. Immunol.* 2009, *10*, 143.
4. E. T. Abualrous, et al. *PLoS One* 2015, *10*, e0135421.
5. D. K. Cole, et al. *J. Clin. Invest.* 2016, *126*, 3626.
6. C. H. Coles, et al. *The Journal of* 2020.
7. C. Szeto, et al. *Int. J. Mol. Sci.* 2020, *22*, 68.
8. L. Yin, et al. *Immunol. Rev.* 2012, *250*, 49.
9. P. Stepniak, et al. *JOURNAL FOR IMMUNOTHERAPY OF CANCER*, BMC CAMPUS, 4 CRINAN ST, LONDON N1 9XW, ENGLAND, 2019.
10. G. Mazzocco, et al. *Front. Genet.* 2021, *12*, 602196.
11. 10x Genomics, *Tech. rep* 2019.
12. A. Weber, et al. *Bioinformatics* 2021, *37*, i237.
13. I. Springer, et al. *Front. Immunol.* 2020, *11*, 1803.
14. V. I. Jurtz, et al. *bioRxiv* 2018, 433706.

We acknowledge funding through the project "Creating an innovative AI-based (Artificial Intelligence) IN SILICO TECHNOLOGY TCRact to launch a NEW SERVICE for designing and optimizing T-cell receptors (TCR) for use in cancer immunotherapies" cofunded by European Regional Development Fund (ERDF) as part of Smart Growth Operational Programme 2014-2020