AACR # 1447

Predicting immunogenic neoepitopes with biology-aware machine learning

ardigen

Giovanni Mazzocco[1], Oleksandr Myronov[1,2], Piotr Skoczylas[1], Jan Kaczmarczyk[1], Iga Niemiec[1], Katarzyna Gruba[1,2], Anna Sanecka-Duin[1] and Piotr Stepniak[1]
1. Ardigen S.A., ul. Podole 76, 30-394 Krakow, Poland, 2. Faculty of Mathematics and Information Science, Warsaw University of Technology, ul. Koszykowa 75, 00-662 Warsaw, Poland
Contact: giovanni.mazzocco@ardigen.com

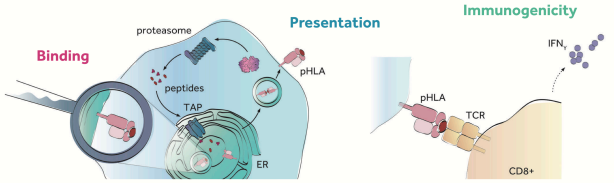
TRUE IMMUNOGENICITY

Cancer vaccines show encouraging results for personalized patient treatment, especially in combination with complementary immune-modulating therapies. The key challenge in cancer vaccine design remains the selection of truly immunogenic Tumor-Specific neoEpitopes (TSE).

Binding affinity and stability of pHLA complexes have been widely used for TSE selection, even though they represent imperfect proxies of immunogenicity. On the other hand, facets of the TSE presentation and immunogenicity such as pHLA:TCR interaction remain relatively underexploited, offering a significant margin of improvement to TSE selection methods. Including such features and training the model on immunogenicity data (instead of binding or presentation data) reduces the gap to the prediction of true immunogenicity.

To this end, we introduce **ArdImmune Rank**: a novel AI-based method for TSE immunogenicity prediction. Benchmarks performed on experimentally-validated immunogenicity data show that ArdImmune Rank outperforms other solutions.

DATA



Immunogenicity of TSEs requires a sequence of biological events including: (i) pHLA **binding**; (ii) pHLA **presentation** on the cell surface; (iii) pHLA:TCR recognition; (iv) IFNγ release.

We present results on Datasets A, B, C:

- Datasets A and B contain peptides, which are already **presented** on the cell surface;
- Dataset C contains cancer peptides with unknown status of presentation.

Name	Data description	Motivation
Dataset A: *Chowell et al. [1] (restricted to human peptides)*	Positives: **presented** (elution, MS) and **immunogenic** (T-cell reactivity). Negatives: **presented** self-derived peptides.	• **Binding** ≠ **immunogenicity**; • **Immunogenicity** model outperforms **binding** models.
Dataset B: *Curated set of presented peptides [2] (Ardigen)*	Curated collection of **presented** and **immunogenic** / non-immunogenic peptides.	• **Binding** ≠ **immunogenicity**; • **Immunogenicity** model outperforms **binding** models.
Dataset C: *Bjerregaard et al. [3] (restricted to HLA-A*02:01)*	Dataset of **immunogenic** / non-immunogenic peptides from 15 cancer patients (SKCM, HGSC, NSCLC, CLL).	Benchmarking of the model on peptides from cancer patients.

ARDIMMUNE RANK

Our immunogenicity model includes the following components:

- **peptide:** trained on representation of peptide sequence;
- **HLA:** trained on representation of HLA pseudo-sequence;
- **interaction:** trained on properties of the pHLA interaction.

The integration of Post-Translational Modification descriptors was tested, but it did not significantly enhance the performance of ArdImmune Rank. Features encompassing the interaction between pHLA and TCRs are currently under development, initial results are shown below.

Models: Deep Neural Networks with multiple inputs corresponding to the components above.

Benchmarks: Model performance was compared to netMHCpan 4.0 [4], netMHCstab 1.0a [5], and Repitope [6]. We also show performance of selected combinations of components of ArdImmune Rank.

Dataset A: Chowell *et al.*

Component	ROC AUC
affinity	0.51
stability	0.46
peptide	0.75
peptide+TCR	0.74
peptide+interaction	0.73
peptide+HLA+interaction	0.60

Table 1. Results on dataset A show that binding affinity and stability do not provide accurate approximation of immunogenicity of HLA-presented peptides. By construction, all non-immunogenic peptides in dataset A are self-derived peptides, which reduces the utility of this dataset for model testing purposes (in [1] it was used only for model training).

Dataset B: presented peptides

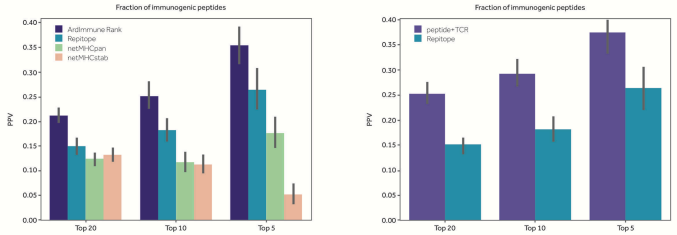


Figure 1. Model performance on Dataset B: a curated collection of peptides presented on the cell surface with known immunogenicity. We show positive predictive value (PPV): fraction of immunogenic peptides in the top 20, top 10, and top 5 of the rank list of peptides obtained using: (left) ArdImmune Rank, Repitope, netMHCpan, and netMHCstab; (right) Ardigen's method for encoding the peptide:TCR interaction (peptide+TCR) and Repitope. The immunogenic peptides are subsampled to obtain a 1:10 ratio. Error bars in both figures represent confidence interval of the mean PPV (from subsampling).

Dataset C: cancer peptides

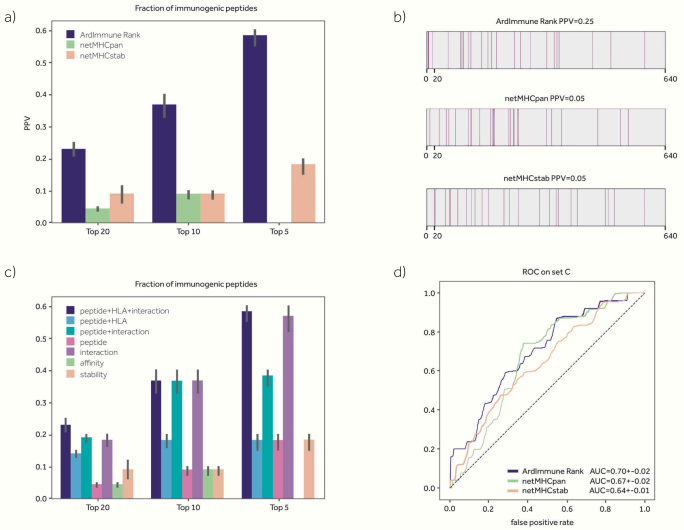


Figure 2. Model performance on dataset C: a collection of peptides from 15 cancer patients. (a-c) positive predictive value (PPV): fraction of immunogenic peptides in the top 20, top 10, and top 5 of the rank list of peptides obtained using selected methods: (a, b) ArdImmune Rank, netMHCpan, and netMHCstab, (c) components of ArdImmune Rank. (d) Receiver operating characteristic (ROC) for the selected scoring methods. Error bars in (a, c) represent confidence interval of the mean PPV.

CONCLUSIONS

- Our results provide clear evidence that there is a significant gain to achieve when moving away from pHLA binding affinity and stability models as predictors of immunogenicity;
- Despite the limited amount of testing data and a strong bias towards pHLA binding (most datasets contain peptides preselected by binding affinity), ArdImmune Rank displays better, more stable performance than other available solutions, representing a valuable tool for cancer vaccine design;
- Current work on ArdImmune Rank is focused on the improvement of components modeling pHLA:TCR interactions. Moreover, the model will be further experimentally validated by the upcoming dataset of peptides from cancer patients.

REFERENCES

1. Chowell D, et al. Proc Natl Acad Sci USA. 2015 Apr 7; 112(14):E1754-62.
2. Vita R, et al. Nucleic Acids Res. 2018 Oct 24; D339-D343.
3. Bjerregaard AM, et al. Front Immunol. 2017 Nov 15; 8:1566.
4. Jurtz V, et al. J Immunol. 2017 Nov 1; 199(9):3360-3368.
5. Rasmussen M, et al. 2016 Aug 15; 197(4):1517-24.
6. Ogishi M and Hiroshi Y, bioRxiv. 2018; 155317.