AI platform provides an EDGE and enables state-of-the-art identification of peptide-HLAs for the development of T cell inducing vaccines and beyond

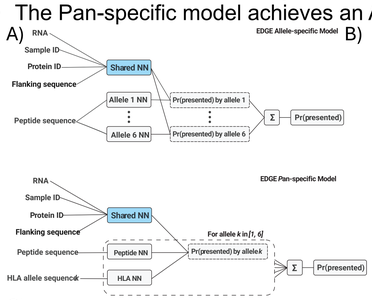
Joshua Klein, Daniel Sprague, Monica Lane, Meghan Hart, Olivia Petrillo, Italo Faria do Valle, Matthew Davis, Andrew Ferguson, Andrew Allen, Karin Jooss, Ankur Dhanik

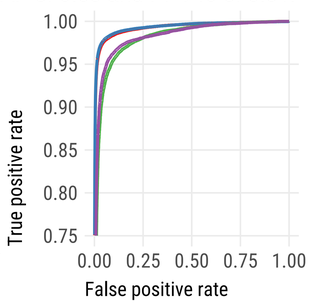
Background

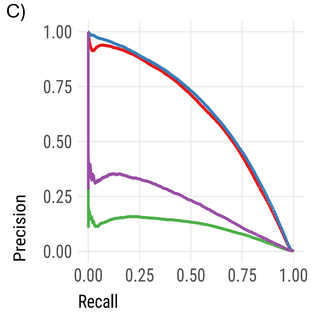
- Accurate prediction of T-cell targets is critical to the development of potent, T-cell inducing vaccines.
- Computational prediction of peptide-HLA complexes which serve as T cell targets, is highly desired.
- EDGE™ (Epitope Discovery for GEnomes) is an artificial intelligence-based platform that provides state-of-the-art models for predicting the cell-surface peptide presentation by HLA molecules.
- A model for predicting peptide presentation by HLA Class I with a PPV40 of 0.53 (10-fold improvement vs. standard tools) was published in Nature Biotech in 2018 [1].
- EDGE is now comprised of improved models trained for predicting peptides presented by HLA Class I or HLA Class II and geared toward oncology or infectious disease applications.
- The models are also predictive of CD8 and CD4 T-cell immunogenicity.

EDGE Models: HLA Class I (Oncology)

- EDGE models for HLA Class I and oncology application are trained using human mono-allelic cell lines and multi-allelic tissue Class I immunopeptidomics data.
- The models consistently outperform MHCFlurry 2.0 [9].
- Allele-specific model is an improved version of published 2018 EDGE model that predicts for 116 HLA alleles with available inhouse training data [1].
- The Allele-specific model achieves an Average Precision (AP) of 0.63 and Positive Predictive Value at 40% Recall (PPV40) of 0.79.
- The Pan-specific model uses HLA allele's sequence while training and is therefore applicable to any Class I allele with known sequence.
- The Pan-specific model achieves an AP of 0.65 and PPV40 of 0.81.







Model	AP	PPV40
EDGE (allele-specific)	0.63	0.79
EDGE (pan-specific)	0.65	0.81
MHCFlurry (binding affinity)	0.10	0.14
MHCFlurry (presentation)	0.21	0.27

Figure 1A. Schematic of the EDGE model. **Figure 1B.** ROC curves comparing performance of EDGE models vs. MHCFlurry **Figure 1C**. PR curves comparing performance of EDGE models vs MHCFlurry

EDGE Predicts Immunogenicity

- EDGE does not specifically model immunogenicity but performs well at predicting immunogenic peptides.
- When ranking mutations from 80 cancer patients [2,3,4] based on immunogenicity, EDGE consistently performs two-fold better than MHCFlurry 2.0 [9].
- In five patients that were treated with Gritstone's personalized cancer vaccines (NCT03639714) containing the top 20 mutations selected by EDGE, ≥ 50% of mutations per patient generated a detectable CD8 response [5].

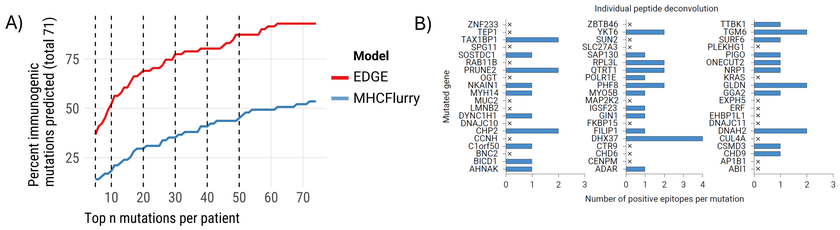


Figure 2A. EDGE consistently ranks twice as many or more immunogenic mutations in the top n ordered by sum of non-self epitope scores. **Figure 2B.** Immunogenic mutations are assessed in each personalized cancer vaccine using ELISPOT-based individual peptide deconvolution.

EDGE Model: HLA Class I (Infectious Diseases)

- Infectious diseases EDGE model was trained using a mixture of human immunopeptidomics dataset and infectious disease binding affinity data [6].
- The model was tested on a collection of publicly available datasets from infectious diseases (HIV [6,7], Influenza A [6], and SARS-CoV-2 [6,8]) held out during training.
- Optimizing EDGE for use on infectious diseases results in improved performance.
- On the HIV and Influenza A datasets, it performs better than the MHCFlurry 2.0; on SARS-CoV-2 dataset, their performance is comparable.

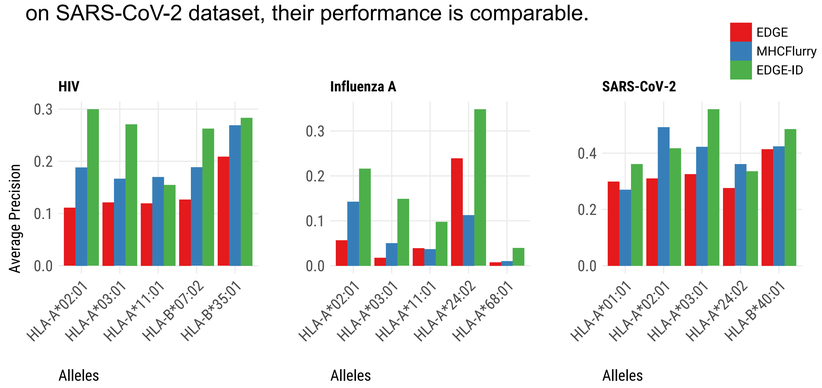


Figure 3. Performance benchmarking of EDGE model for infectious diseases (EDGE-ID) as compared to EDGE for oncology and MHCFlurry 2.0 This benchmark was done on a per-allele basis for each HLA allele in the dataset. Alleles were chosen based upon the number of positives available. The benchmarking datasets were withheld from our training and validation. MHCFlurry was trained with the data from the benchmarking dataset.

EDGE Model: HLA Class II (Oncology)

- EDGE model for Class II, termed EDGE-II, uses a pretrained protein language model, a novel learned HLA allele-deconvolution strategy, and in-house immunopeptidomics training data.
- EDGE-II achieves a test set AP of 0.92 and outperforms NetMHCIIpan and BERTMHC [10,11] on an externally curated validation set [10] with an AP = 0.71.
- CD4 immunogenicity in a personalized cancer vaccine context [13] is better predicted by EDGE-II than NetMHCIIpan and MARIA [10,12].

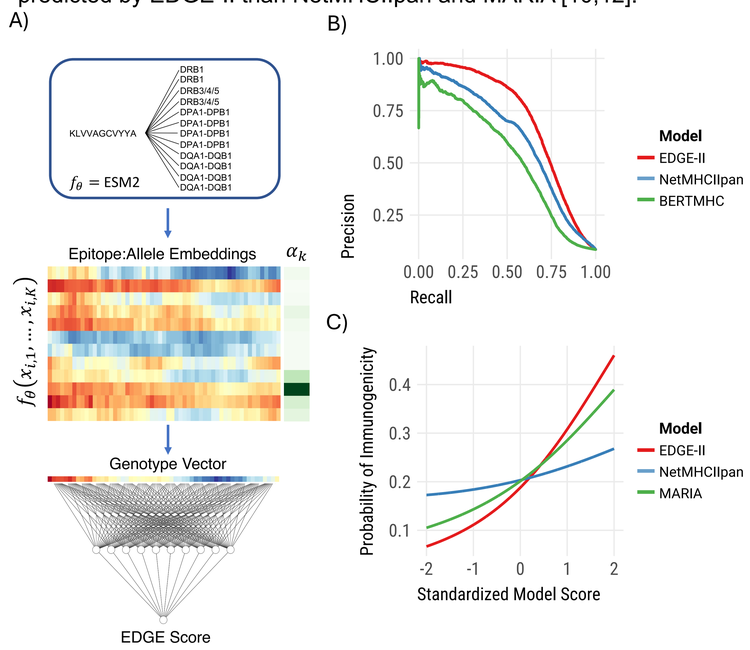


Figure 4A. Schematic of the EDGE-II model. **Figure 4B.** Performance of EDGE-II and public tools on an external validation dataset. **Figure 4C.** Correlation of model scores with CD4 immunogenicity on the Ott *et al.* personalized cancer vaccine dataset [13].

Conclusions

- EDGE is a state-of-the-art AI platform for predicting T-cell targets for oncology and infectious diseases applications.
- For oncology, it contains two models (allele-specific and pan-specific) for predicting Class I and one model (EDGE-II) for predicting Class II antigens.
- A model for predicting Class I antigens for infectious diseases (EDGE-ID) benefits from training on binding affinity data in addition to the immunopeptidomics data.
- Superior performance and association with immunogenicity is achieved by all models.
- EDGE could enable development of vaccines that can elicit a robust T-cell response resulting in durable patient response.

References

[1] Bulik-Sullivan, B., …, Yelensky, R. (2018). *Nature Biotechnology;* [2] Malekzadeh, P., … Deniger, D. C. (2020). *Clinical Cancer Research*, *26*(6); [3] Gros, A., …, Rosenberg, S. A. (2016). *Nature Medicine*, *22*(4); [4] Tran, E., …, Rosenberg, S. A. (2015). *Science*, *350*(6266)
[5] Palmer, C. D., … Jooss, K. (2022). *Nature Medicine*, *28*(8); [6] Vita, R., …, Peters,B. (2018). *Nucleic Acids Research*;[7] Editors: Mamrosh, J. L., …, Walker, B. D. (2023) *HIV Molecular Immunology;* [8] Tarke, A., …, Sette, A. (2021). *Cell Reports Medicine*, *2*(2)
[9] O'Donnell, T. J., …, & Laserson, U. (2020). *Cell Systems*, *11*(1); [10] Reynisson, B., …, Nielsen, M. (2020). *Nucleic Acids Research;* [11] Cheng, J., …, Malone, B. (2021). Bioinformatics, 37(June)
[12] Chen, B., … Alizadeh, A. A. (2019). *Nature Biotechnology;* [13] Ott, P. A., …, Wu, C. J. (2017). *Nature*, *547*(7662)