

ML-based prediction of pairwise antibody epitope binning

ADIMAB

Akila Katuwawala, Elizabeth Champney, Elizabeth McGurk, Bingjie Ouyang, Cameron Henkel, Michael Brown, Arvind Sivasubramanian (Adimab LLC, Lebanon, NH, and Mountain View, CA, USA)

ABSTRACT

Antibody binning assays play a crucial role in therapeutic antibody discovery by providing information on epitope diversity¹. We propose a computationally efficient Machine Learning-based (ML) workflow ("ML-Binning") to predict the pairwise binning competition for large panels of antibodies, including IgGs and HCABs (Heavy Chain Antibodies). The workflow utilizes Fv-sequence-derived input features and experimental data for a subset of the pairwise interactions (5-10%) to train a custom ML model that predicts pairwise binning for the remaining interactions. In the initial benchmarking, using simulated training/test data from distinct IgG and HCAB panels, the ML-Binning method accurately predicted the pairwise binning (i.e. competitor or non-competitor status) with a typical AU-ROC (Area Under the Receiver Operating Curve) > 0.8. Subsequently, in a real-world blind test using a panel of N=69 target-specific IgGs, the model accurately predicted the pairwise competition with an AU-ROC of 0.83 and recapitulated the experimentally determined clustering of epitope bins. We have incorporated this workflow into an end-to-end software solution as a resource- and time-saving option that is competitive with the standard experimental approaches for pairwise antibody binning.

METHODS

ML input features



Figure 1: A 1D feature vector of length 268 (typical maximum VH+VL length for an IgG) is generated based on the structure-aware sequence alignment for each antibody pair. (Illustrated for select VH residues)

ML model training

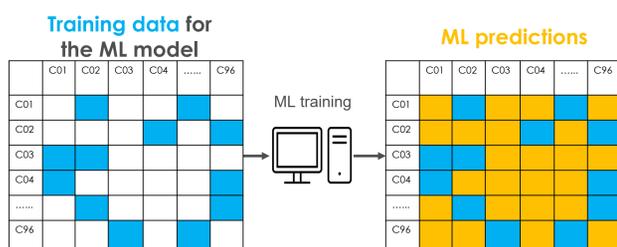


Figure 2: The ML model uses experimentally determined (Octet BLI) competition data for a random subset of $N \times m$ antibody pairs and subsequently predicts competition between remaining $N \times (N-m)$ pairs.

ML-Binning system architecture

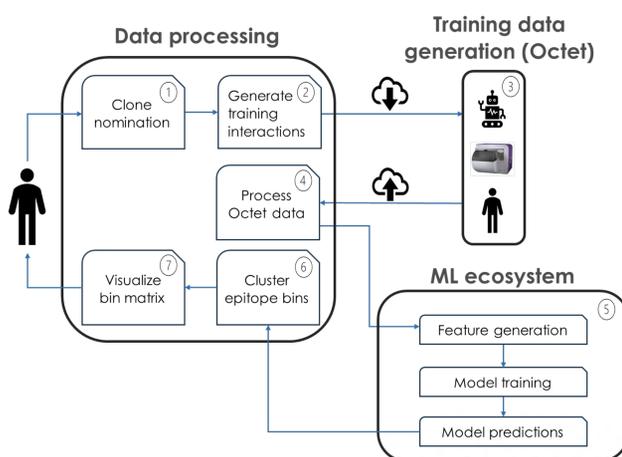


Figure 3: The end-to-end software solution executes the *in silico* steps, such as feature generation, model training, and prediction, as well as the handoffs between the *in silico* and experimental aspects of the workflow.

RESULTS

Benchmarking with simulated training and test datasets establishes the robustness of the proposed approach.

The initial method development used SPR (Carterra) experimental competition data for 8 IgG panels and 1 HCAB panel from distinct discovery campaigns.

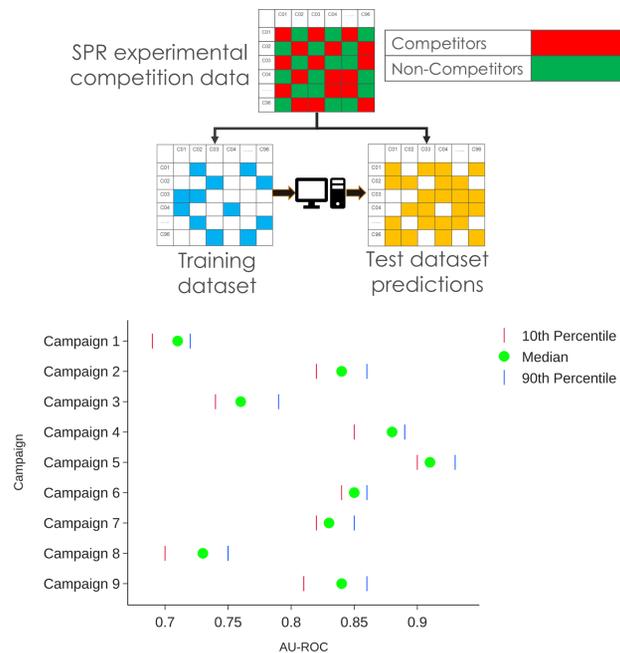


Figure 4: Across 500 random training/test splits, the model returns a median AU-ROC > 0.8 for 6 out of 9 campaign datasets.

Alternative approaches for generating training data and selecting input features

Cluster antibodies by sequence similarity, then assemble the training set of size (a) $N \times$ (cluster centers) (OR) (b) (cluster centers) \times (cluster centers)

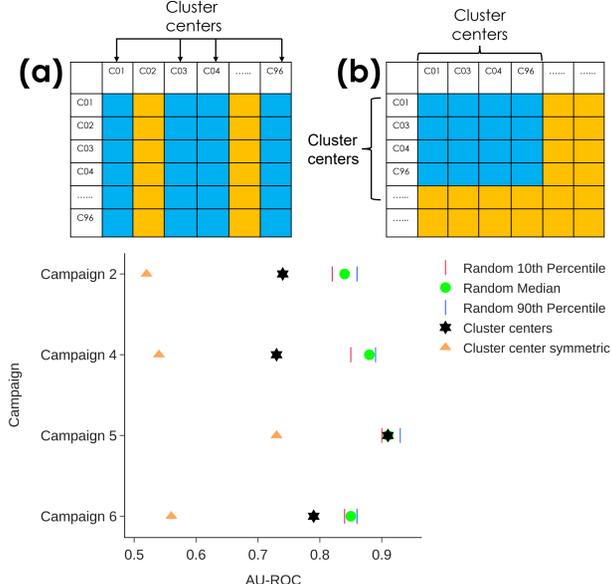


Figure 5: Training datasets based on random $N \times m$ pairs consistently return higher test dataset AU-ROC.

VH and VL germline assignments as extra features

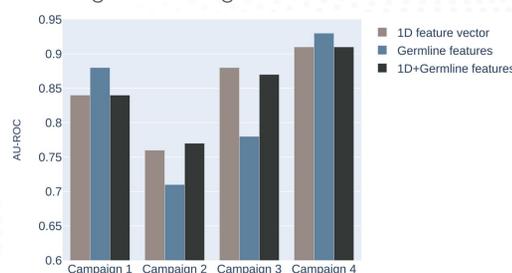


Figure 6: The 1D feature vector outperforms germline features in 2 out of 4 campaigns.

The proposed workflow provides accurate predictions in a real-world campaign.

Panel: 69 IgGs, 4761 pairwise interactions

Training: Octet binning data for 345 pairs (7%)

Prediction: Remaining 4461 interactions

Validation: Compare ML predictions to the Carterra competition data for the full panel.

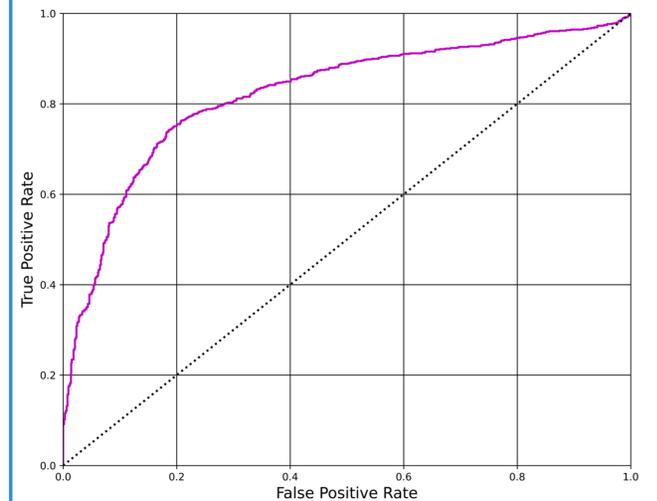


Figure 7: AU-ROC value of 0.83 on the blind dataset.

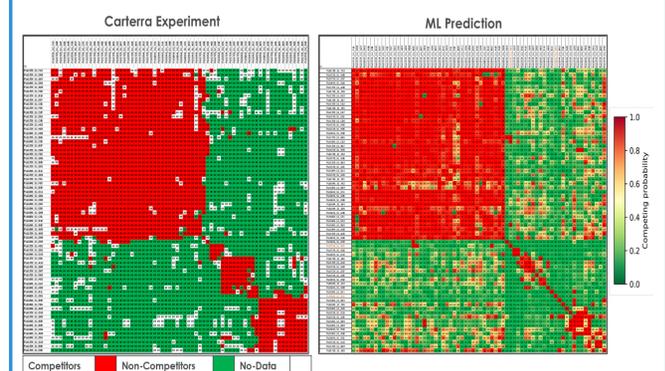


Figure 8: ML-binning predicted bins agree with those determined using Carterra.

CONCLUSIONS

Factor	Carterra	ML-binning	Savings
Antibody consumption	0.36 nmol	0.11 nmol	70%
Antigen consumption	7.5 nmol	0.5 nmol	93%
Chip/Sensor cost	\$940	\$550	42%
Instrument time	3 days	1 day	66%
Attrition/Mispredictions	~11%	~10%	on par

Table 1: For 96 x 96 pairs, ML-binning saves cost, time, and reagent use compared to complete pairwise binning on Carterra.

REFERENCES

- Sivasubramanian A, Estep P, Lynaugh H, et al. Broad epitope coverage of a human in vitro antibody library. *MAbs*. 2017;9(1):29-42. doi:10.1080/19420862.2016.1246096

ACKNOWLEDGEMENTS

We thank all members of Adimab's High-throughput Expression, Protein Analytics, Computational Biology, Antibody Engineering, and Legal groups for their valuable contributions to this project.

