nature machine intelligence

Article

https://doi.org/10.1038/s42256-023-00694-6

Deep neural networks predict class I major histocompatibility complex epitope presentation and transfer learn neoepitope immunogenicity

In the format provided by the authors and unedited

Supplementary Information

IM-Positives / EL -Negatives

To estimate the ability of methods to discriminate between presented and immunogenic epitopes, we evaluated all methods using all mass-spec identified epitopes as negatives and immunogenic epitopes as positives (IM+/EL-).

To interpret the results, we considered the composition of the training sets of each method. BigMHC IM and the final version of BigMHC EL were trained using all mass-spec identified epitopes as positives, and BigMHC IM was further fine-tuned using transfer learning with the IM training set. MHCflurry-2.0, MixMHCpred-2.2, and HLAthena were also trained using most of the EL examples as positives. TransPHLA, MHCnuggets-2.4.0, and PRIME-2.0 were not trained with these examples.

We considered that these differences in training set composition might be related to method performance, acting as confounders. To quantitatively measure the confounding effect, we used a correlation test to compare each methods performance on the IM+/EL- data with performance on EL+/rand- data (random negatives), which were evaluated on all methods. We found that there was a strong and statistically significant negative correlation between IM+/EL- performance and EL+/rand- performance. Methods that do not use the EL data as a positive class in their training sets outperform those that do. Because of this correlation, we conclude that the ability of methods to discriminate between mass-spec identified and immunogenic epitopes cannot be correctly evaluated with the data available to us. A detailed description of our analysis can be found below.

We computed AUROC and AUPRC for each of these methods, using both sources of IM+ data: the Neoepitope test set, the IEDB Infectious Disease test set. This yielded 15 AUROC values and 15 AUPRC values for each source of IM+ data, presented in Supplementary Table 1.

Supplementary Table 1

Source	Neoepitopes		IEDB Infectious Disease Antigens	
Metric	AUROC	AUPRC	AUROC	AUPRC
BigMHC EL	0.1113	0.0003	0.1674	0.0030
BigMHC IM	0.3397	0.0005	0.3715	0.0040
BigMHC ELIM	0.2965	0.0004	0.3096	0.0035
PRIME-2.0 Scores	0.2087	0.0004	0.3236	0.0036
PRIME-2.0 Ranks	0.2141	0.0029	0.3174	0.0100
MixMHCpred-2.2 Scores	0.1936	0.0004	0.2391	0.0032
MixMHCpred-2.2 Ranks	0.1899	0.0004	0.2392	0.0076
MHCflurry-2.0 Scores	0.1632	0.0003	0.2245	0.0032
MHCflurry-2.0 Ranks	0.1632	0.0003	0.2245	0.0032
TransPHLA	0.3675	0.0005	0.3585	0.0039
MHCnuggets-2.4.0	0.4736	0.0006	0.3815	0.0039
NetMHCpan-4.1 Scores	0.2490	0.0004	0.3137	0.0035
NetMHCpan-4.1 Ranks	0.2382	0.0004	0.2851	0.0034

HLAthena Scores	0.2284	0.0004	0.2308	0.0032
HLAthena Ranks	0.2289	0.0004	0.2308	0.0035

Next, we compared these values to the AUROC and AUPRC when all of the mass-spec identified epitopes were used as positives and the negatives were random. Because these performance statistics were stratified by allele, we took the mean AUROC and AUPRC for all alleles, so that each method was represented by a single AUROC and AUPRC value.

Finally, we computed the Spearman Correlations in Supplementary Table 2.

Supplementary Table 2

Spearman Correlation of All Method Performances on all EL alleles (EL-pos/random-neg) and the IM-pos/EL-neg Performances				
Metric	Neoepitope Test Set		IEDB Infectious Disease Test Set	
	AUROC	AUPRC	AUROC	AUPRC
Correlation	-0.5179	-0.8393	-0.7964	-0.7536
p-value	4.8×10^{-2}	9.137×10^{-5}	3.8×10^{-4}	1.2×10^{-3}

AUROC and AUPRC Metrics

Our test sets are imbalanced, so high AUROC does not imply high AUPRC, nor vice versa; therefore, these results do not demonstrate inconsistencies. In imbalanced evaluation frameworks, neither AUROC nor AUPRC alone is sufficient for describing the model performance as they measure different aspects of a predictor; namely, these curves demonstrate the change in sensitivity over: specificity (ROC) and precision (PRC). If one observes a high AUROC and a low AUPRC, then the model is predicting true negatives better than true positives. Conversely, if one observes a low AUROC and high AUPRC, then the model is more precise at the cost of specificity.

BigMHC IM Generalization Across Alleles

Because data is limited it is difficult to assess the true generalizability across different alleles. However, we computed the BigMHC IM AUROC and AUPRC for each allele in the neoepitope test set and the IEDB infectious disease test set in Supplementary Table 3 and Supplementary Table 4 respectively. We also analyzed whether BigMHC IM performance is skewed toward a few common alleles. Although there is substantial variability in performance across alleles, we do not see a trend (average Spearman correlation magnitude is 0.1032), where BigMHC IM performance is better on the alleles with more evaluation data.

Supplementary Table 3

BigMHC IM Neoepitope Immunogenicity Evaluation				
Allele	Negatives	Positives	AUROC	AUPRC
HLA-A*01:01	25	19	0.4484	0.4740
HLA-A*02:01	139	25	0.4213	0.1216
HLA-A*03:01	30	9	0.4889	0.2231
HLA-A*11:01	10	4	0.3250	0.4076
HLA-A*23:01	5	5	0.2000	0.3408
HLA-A*30:02	2	21	0.5714	0.9488
HLA-A*31:01	9	5	0.2444	0.2416
HLA-A*68:02	9	11	0.2727	0.4177
HLA-B*07:02	47	17	0.6233	0.3024
HLA-B*08:01	15	15	0.3911	0.4878
HLA-B*15:01	24	3	0.3750	0.0913
HLA-B*18:01	13	22	0.5035	0.6891
HLA-B*35:01	14	7	0.7143	0.7039
HLA-B*40:01	10	3	0.3667	0.1751
HLA-B*40:02	1	1	1.0000	1.0000
HLA-B*44:02	21	1	0.5714	0.0500
HLA-B*49:01	16	3	0.3750	0.1371
HLA-B*53:01	12	3	0.6944	0.5940
HLA-B*55:01	6	3	0.5000	0.5323
HLA-B*57:01	8	2	0.7500	0.2875
HLA-C*03:03	17	5	0.4706	0.2271
HLA-C*04:01	9	1	1.0000	1.0000
HLA-C*05:01	44	2	0.4659	0.0378
HLA-C*06:02	10	1	0.8000	0.1667
HLA-C*07:04	7	1	0.1429	0.0714

Supplementary Table 4

BigMHC IM IEDB Infectious Disease Immunogenicity Evaluation				
Allele	Negatives	Positives	AUROC	AUPRC
HLA-A*01:01	28	90	0.4861	0.7805
HLA-A*02:01	79	575	0.5820	0.9077
HLA-A*02:05	2	3	0.3333	0.6556
HLA-A*03:01	74	23	0.6839	0.3586
HLA-A*11:01	138	54	0.6116	0.3575
HLA-A*23:01	2	3	0.6667	0.7639
HLA-A*24:02	61	108	0.4399	0.5794
HLA-A*26:01	1	39	0.3077	0.9712
HLA-A*29:02	2	39	0.6282	0.9776
HLA-A*30:01	3	20	0.6333	0.9145
HLA-A*33:03	3	2	0.5000	0.6625
HLA-A*34:02	4	5	0.6500	0.7685
HLA-A*68:02	16	3	0.3958	0.1228
HLA-A*74:01	1	3	0.6667	0.9028
HLA-B*07:02	84	103	0.4473	0.5224
HLA-B*08:01	26	44	0.5087	0.6701
HLA-B*13:02	1	2	1.0000	1.0000
HLA-B*14:02	2	2	0.2500	0.3333
HLA-B*15:01	16	35	0.3696	0.5939
HLA-B*15:03	6	3	0.1111	0.1960
HLA-B*15:10	7	2	0.8571	0.7083
HLA-B*27:01	1	1	0.0000	0.2500
HLA-B*27:02	6	7	0.4762	0.6469
HLA-B*27:05	6	14	0.4048	0.7050
HLA-B*35:01	6	69	0.5097	0.9435
HLA-B*40:01	9	30	0.7889	0.8981
HLA-B*41:02	10	2	0.3500	0.1269
HLA-B*42:01	1	1	1.0000	1.0000
HLA-B*44:02	2	36	0.7361	0.9800
HLA-B*44:03	2	19	0.5789	0.9444
HLA-B*45:01	1	1	0.0000	0.2500
HLA-B*49:01	2	1	1.0000	1.0000
HLA-B*51:01	1	36	0.0000	0.8976
HLA-B*53:01	5	8	0.8000	0.8406
HLA-B*58:01	2	1	0.5000	0.2500
HLA-C*04:01	3	2	0.6667	0.4167
HLA-C*07:01	6	7	0.1667	0.3651
HLA-C*07:02	8	16	0.5078	0.6646
HLA-C*16:01	9	2	0.9444	0.7917