Machine Learning Competition in Immunology – Prediction of HLA class I molecules

Guang Lan Zhang[a], [redacted] **Phil Bradley**[c], [redacted] **Nebojsa Jojic**[f], **Yohan Kim**[g], **Oliver Kohlbacher**[h], **Claus Lundegaard**[i], **Craig A. Magaret**[c], **Morten Nielsen**[i], [redacted] **Chen Yanover**[c], **Shanfeng Zhu**[e], **Michael T. Rock**[m], **James E. Crowe Jr.**[m], [redacted], **Włodzisław Duch**[o,p], **Vladimir Brusic**[a*]

[a]CVC, Dana-Farber Cancer Institute, Boston, MA, USA
[redacted]
[c]Fred Hutchinson Cancer Research Center
[redacted]
[e]SKEEP, Fudan University, Shanghai, China
[f]Microsoft Research, Redmond, WA, USA
[g]La Jolla Institute for Allergy and Immunology, La Jolla, Ca, USA
[h]University of Tuebingen, Germany
[i]CBS, Danish Technical University, Lyngby, Denmark
[redacted]
[m]Vanderbilt University, Nashville, TN, USA
[redacted]
[o]Nicolaus Copernicus University, Toruń, Poland
[p]SCE, Nanyang Technological University, Singapore.

1. Introduction

Experimental studies of immune system and related applications such as characterization of immune responses against pathogens, vaccine design, or optimization of therapies are combinatorially complex, time-consuming and expensive. The main methods for large-scale identification of T-cell epitopes from pathogens or cancer proteomes involve either reverse immunology or high-throughput mass spectrometry (HTMS). Reverse immunology approaches involve pre-screening of proteomes by computational algorithms, followed by experimental validation of selected targets (Mora et al., 2006; De Groot et al., 2008; Larsen et al., 2010). HTMS involves HLA typing, immunoaffinity chromatography of HLA molecules, HLA extraction, and chromatography combined with tandem mass spectrometry, followed by the application of computational algorithms for peptide characterization (Bassani-Sternberg et al., 2010). Hundreds of naturally processed HLA class I associated peptides have been identified in individual studies using HTMS in normal (Escobar et al., 2008), cancer (Antwi et al., 2009; Bassani-Sternberg et al., 2010), autoimmunity-related (Ben Door et al, 2010), and infected samples (Wahl et al, 2010).

Computational algorithms are essential steps in high-throughput identification of T-cell epitope candidates using both reverse immunology and HTMS approaches. Peptide binding to MHC molecules is the single most selective step in defining T cell epitope and the accuracy of computational algorithms for prediction of peptide binding, therefore, determines the accuracy of the overall method. Computational predictions of peptide binding to HLA, both class I and class II, use a variety of algorithms ranging from binding motifs to advanced machine learning techniques (Brusic et al., 2004; Lafuente and Reche, 2009) and standards for their assessments have been developed. The assessments of computational servers that predict peptide binding to several common HLA class I alleles have been performed by different groups (see Peters et al., 2006; Lin et al., 2008; Gowthaman et al., 2010). Some of these models were reported to be highly accurate while others need improvement.

2. The competition

Computational methods offer key support for collection, processing, and analysis of high-throughput data in immunology. This issue of the Journal of Immunological Methods focuses on machine learning aspects of high-throughput methods for identification of T cell epitopes. Here we report the results of machine learning competition where the target was to benchmark the performance of computational methods of peptide binding to three HLA class I molecules, HLA-A*01:01, -A*02:01, and B*07:02, using newly generated sets of 9-mer and 10-mer peptides. There were 20 contestants, including both existing and newly developed methods. For new contestants, the pre-processed training data were available from the DFMRLI web site (Zhang et al, 2011a, this issue) and also from other public sources such as IEDB, SYFPEITHI, MHCPEP, ANTIJEN, and MHCBN (Vita et al., 2010; Schuler et al, 2007; Brusic et al, 1997; Toseland et al., 2005; Lata et al., 2005). The target data are described in (Rock et al., 2011, this issue). These target data include 144 binders and 651 non-binders (9-mers) as well as 86 binders and 487 non-binders (10-mers) (Table 1). The competition was held in conjunction with the 19th International Conference on Artificial Neural Networks, held on 14-17 September 2009, Limassol, Cyprus.

The experimental measurement of peptide binding was identified using iTopia system (Wulf et al, 2009). The performance of 20 prediction systems in the competition and two benchmark predictors BIMAS (www-bimas.cit.nih.gov/molbio/hla_bind/ Parker et al; 1994) and SYFPEITHI (www.syfpeithi.de, Schuler et al., 2007) was compared with experimental results. The predictive performance was assessed in accordance to criteria defined in (Lin et al., 2008), where classification into binders and non-binders was performed, followed by prediction of binding affinity. The classification performance problem (see Table 2) – prediction of binders *vs*. non-binders – indicated that 15 (75%) of predictors and both benchmark predictors showed excellent performance in classification of HLA-A*01:01 9-mer binders. Nineteen (95%) predictors and both benchmark predictors showed excellent performance in prediction of HLA-A*02:01 9-mer binders. 16 (80%) of the predictors and both benchmark predictors showed excellent performance

in prediction of HLA-B*07:02 9-mer binders. 11 (50%) predictors and one benchmark predictor showed excellent performance in prediction of HLA-A*01:01 10-mer binders. 13 (65%) predictors and both benchmark predictors showed excellent performance in prediction of HLA-A*02:01 10-mer binders. Only 6 (30%) predictors and no benchmark predictors showed excellent performance in prediction of HLA-B*07:02 10-mer binders. The number of predictors that showed better performance than benchmark servers were twelve (A*01:01 9-mers), none (A*02:01 9-mers), fifteen (B*07:02 9-mers), eight (A*01:01 10-mers), one (A*02:01 10-mers), and sixteen (B*07:02 10-mers). In addition 13 and 4 predictors, respectively, predicted equally well as benchmark predictors in classification of A*02:01 9-mers and 10-mers. In summary, the existing benchmark predictors (BIMAS and SYFPEITHI) showed excellent classification performance for HLA*A*02:01, for both 9-mer and 10-mer peptides. The majority of modern predictors showed equal or improved predictions as compared to the benchmark performance. Modern predictors showed marked improvements over classification performance for A*01:01 and B*07:02. The A_{ROC} values of best predictors range from 0.96 to 1.00 showing nearly perfect classification performance. The analysis of overall performance was performed by ranking the average performance of each predictor for 9-mers and also for 10-mers (Table 3). The results show that the leading modern predictors have improved prediction performance relative to the benchmark predictors. The greatest improvement was shown for HLA-B*07:02 10-mers. The analysis of prediction of binding affinity results (Table 4) shows that predictor performance, measured by correlation coefficient, ranges from r=0.663 to r=0.931. This represents a marked improvement over the benchmarks (r=0.455 to r=0.775).

The purpose of this competition was to stimulate the development of improved methods in the field, engage more contributors from the machine learning community, and define new benchmarks in the field. The best performing methods are described elsewhere in this issue. The best classifier for A*01:01 9-mers H00001 is described in (Hu et al., 2011, this issue), the best classifier for A*02:01 10-mers PepMHC-I in (Vider-Shalit and Louzoun, 2011, this issue). The overall winner N00003 (NetMHCcons), together with N00001 (netMHC) and N00002 (netMHCpan) are described in (Lundegaard et al., 2011, this issue). An additional application of N00002 is described in (Zhang et al, 2011b, this issue). Other well-performing predictors are also described: I00001 (Kim et al., 2011, this issue), and imaginary1-3 (Huang and Jojic, 2011, this issue). Previous benchmark predictors, BIMAS and SYFPEITHI showed surprisingly robust performance that has only recently been surpassed due to availability of much larger data sets.

3. Conclusion

This competition clearly established new benchmarks in the field of HLA binding prediction. The results are consistent with previous comparison results (Lin et al., 2008; Gowthaman et al., 2010) where similar performance of best predictors was reported. Further improvements in the field should focus in expanding the range of HLA alleles that have excellent predictive models as well as inclusion and benchmarking of predictors

for 8-mers, 10-mers, and 11-mers. In addition, there is a need for predictors for classification of naturally processed peptides that will significantly reduce the number of binders that are readily identified in binding assays, but are functionally irrelevant. This will be achieved by combining existing data with new data generated by HTMS with existing set and definition of patterns that define naturally processed peptides. We must note, however that because of the selection bias, the target data for this competition tend to favor high-affinity binders. Because of this issue the prediction results are suitable for comparisoin of predictor performance, but may represent an overestimate of accuracy if applied to in silico scaning of complete proteins or proteomes.

Acknowledgements

This competition was kindly supported by the ICANN 2009 Conference, by the International Immunomics Society, and by the European Neural Network Society.

References

Antwi, K., Hanavan, P.D., Myers, C.E., Ruiz, Y.W., Thompson, E.J., Lake, D.F., 2009. Proteomic identification of an MHC-binding peptidome from pancreas and breast cancer cell lines. Mol. Immunol. 46(15), 2931.

Bassani-Sternberg, M., Barnea, E., Beer, I., Avivi, I., Katz, T., Admon, A., 2010. Soluble plasma HLA peptidome as a potential source for cancer biomarkers. Proc Natl Acad Sci USA 107(44), 18769.

Ben Dror, L., Barnea, E., Beer, I., Mann, M., Admon, A., 2010. The HLA-B*2705 peptidome. Arthritis Rheum. 62(2), 420.

Brusic, V., Bajic, V.B., Petrovsky, N., 2004. Computational methods for prediction of T-cell epitopes - a framework for modelling, testing, and applications. Methods 34(4), 436.

Brusic, V., Rudy, G., Harrison, L.C., 1998. MHCPEP, a database of MHC-binding peptides: update 1997. Nucleic Acids Res., 26(1):368-71.

De Groot, A.S., Rivera, D.S., McMurry, J.A., Buus, S., Martin, W., 2008. Identification of immunogenic HLA-B7 "Achilles' heel" epitopes within highly conserved regions of HIV. Vaccine 26(24), 3059.

Escobar, H., Crockett, D.K., Reyes-Vargas, E., Baena, A., Rockwood, A.L., Jensen, P.E., Delgado, J.C., 2008. Large scale mass spectrometric profiling of peptides eluted from HLA molecules reveals N-terminal-extended peptide motifs. J. Immunol. 181(7), 4874.

Gowthaman, U., Chodisetti, S.B., Parihar, P., Agrewala, J.N., 2010. Evaluation of different generic in silico methods for predicting HLA class I binding peptide vaccine candidates using a reverse approach. Amino Acids 39(5), 1333.

Hu, X., Mamitsuka, H., Zhu, S., Ensemble approaches for improving HLA Class I-peptide binding prediction. J. Immunol. Methodsc 2011 (in press).

Modeling major histocompatibility complex binding by nonparametric averaging of multiple predictors and sequence encodings.

Huang, J.C., Jojic, N., 2011. Modeling major histocompatibility complex binding by nonparametric averaging of multiple predictors and sequence encodings. J. Immunol. Methods (in press).

Kim, Y., Sette, A., Peters, B., 2011. Applications for T-cell epitope queries and tools in the Immune Epitope Database and Analysis Resource. J. Immunol. Methods. 2011 (in press).

Lafuente, E.M., Reche, P.A., 2009. Prediction of MHC-peptide binding: a systematic and comprehensive overview. Curr. Pharm. Des. 15(28), 3209.

Larsen, M.V., Lelic, A., Parsons, R., Nielsen, M., Hoof, I., Lamberth, K., Loeb, M.B., Buus, S., Bramson, J., Lund, O., 2010, Identification of CD8+ T cell epitopes in the West Nile virus polyprotein by reverse-immunology using NetCTL. PLoS One 5(9), e12697.

Lata, S., Bhasin, M., Raghava, G.P., 2009. MHCBN 4.0: A database of MHC/TAP binding peptides and T-cell epitopes. BMC Res. Notes 2, 61.

Lin, H.H., Ray, S., Tongchusak, S., Reinherz, E.L., Brusic, V., 2008. Evaluation of MHC class I peptide binding prediction servers: applications for vaccine research. BMC Immunol. 9, 8.

Lundegaard, C., Lund, O., Nielsen, M., 2011. Prediction of epitopes using neural network based methods. J. Immunol. Methods (in press).

Mora, M., Donati, C., Medini, D., Covacci, A., Rappuoli R., 2006. Microbial genomes and vaccine design: refinements to the classical reverse vaccinology approach. Curr Opin Microbiol. 9(5), 532.

Parker, K.C., Bednarek, M.A., Coligan, J.E., 1994. Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. J. Immunol. 152(1), 163.

Peters, B., Bui, H.H., Frankild, S., Nielson, M., Lundegaard, C., Kostem, E., Basch, D., Lamberth, K., Harndahl, M., Fleri, W., Wilson, S.S., Sidney, J., Lund, O., Buus, S., Sette, A., 2006. A community resource benchmarking predictions of peptide binding to MHC-I molecules. PLoS Comput. Biol. 2(6), e65.

Purcell, A.W., McCluskey, J., Rossjohn, J., 2007. More than one reason to rethink the use of peptides in vaccine design. Nat. Rev. Drug Discov. 6(5), 404.

Rock MT, McKinney BA, Yoder SM, Prudom CE, Wright DW, Crowe JE Jr. Identification of potential human respiratory syncytial virus and metapneumovirus T cell epitopes using computational prediction and MHC binding assays. J Immunol Methods. 2011 (in press).

Schuler, M.M., Nastke, M.D., Stevanović, S., 2007. SYFPEITHI: database for searching and T-cell epitope prediction. Methods Mol. Biol. 409, 75.

Vider-Shalit, T., Louzoun, Y., 2011. MHC-I prediction using a combination of T cell epitopes and MHC-I binding peptides. J. Immunol. Methods (in press).

Swets, J.A.,1988. Measuring the accuracy of diagnostic systems. Science 240,1285.

Toseland, C.P., Clayton, D.J., McSparron, H., Hemsley, S.L., Blythe, M.J., Paine, K., Doytchinova, I.A., Guan, P., Hattotuwagama, C.K., Flower, D.R., 2005. AntiJen: a quantitative immunology database integrating functional, thermodynamic, kinetic, biophysical, and cellular data. Immunome Res. 1(1), 4.

Vita, R., Zarebski, L., Greenbaum, J.A., Emami, H., Hoof, I., Salimi, N., Damle, R., Sette, A., Peters, B., 2010. The immune epitope database 2.0. Nucleic Acids Res. 38(Database issue), D854.

Wahl, A., Schafer, F., Bardet, W., Hildebrand, W.H., 2010. HLA class I molecules reflect an altered host proteome after influenza virus infection. Hum Immunol. 71(1), 14.

Wulf, M., Hoehn, P., Trinder, P., 2009. Identification of human MHC class I binding peptides using the iTOPIA- epitope discovery system. Methods Mol Biol. 524, 361.

Zhang, G.L., Deluca, D.S., Keskin, D.B., Chitkushev, L., Zlateva, T., Lund, O., Reinherz, E.L., Brusic, V., MULTIPRED2: A computational system for large-scale identification of peptides predicted to bind to HLA supertypes and alleles. J. Immunol. Methods 2011a (in press).

Zhang, G.L., Lin, H.H., Keskin, D.B., Reinherz, E.L., Brusic, V., Dana-Farber repository for machine learning in immunology. J. Immunol. Methods. 2011b (in press).

Table 1. The number of test peptides in each studied group.

	9-mers		10-mers	
	Binder	**Non-binder**	**Binder**	**Non-binder**
A*01:01	25	240	14	177
A*02:01	76	189	62	129
B*07:02	43	222	10	181
Total	144	651	86	487

Table 2. The prediction systems assessment for classification into binders and non-binders using the area under the ROC curve (see Lin et al., 2008; Swets, 1988). The values of $A_{ROC}>0.9$ indicate excellent, $0.8<A_{ROC}<0.9$, and $A_{ROC}<0.8$ indicate excellent, moderate, and poor classification performance respectively. The best performing predictors are indicated by shaded fields.

	Area under the ROC curve (A_{ROC})					
	A*01:01 9mer	**A*02:01 9mer**	**B*07:02 9mer**	**A*01:01 10mer**	**A*02:01 10mer**	**B*07:02 10mer**
BIMAS	0.91	**0.99**	**0.92**	**0.92**	**0.99**	**0.85**
SYFPEITHI	**0.92**	0.98	0.72	0.69	0.96	0.82
Bunsen	0.94	0.99	0.94	0.88	0.98	0.89
Frags	0.77	0.93	0.90	0.92	0.93	0.86
FudanCS	0.95	0.99	0.95	0.95	0.99	0.88
H00001	0.97	0.99	0.96	0.96	0.99	0.90
hnp	0.95	0.98	0.95	0.89	0.98	0.97
I00001	0.96	0.99	0.95	0.97	0.98	0.89
imaginary1	0.96	0.99	0.96	0.81	0.41	0.73
imaginary2	0.91	0.94	0.94	0.98	0.98	0.95
imaginary3	0.96	0.99	0.96	0.95	0.95	0.89
lpp729	0.91	0.99	0.91	0.81	0.94	0.83
MHChackers1	0.46	0.58	0.52	0.53	0.51	0.53
N00001	0.96	0.99	0.95	0.99	0.99	0.94
N00002	0.96	0.99	0.96	0.98	0.99	0.97
N00003	0.96	0.99	0.96	0.99	0.99	0.95
P00001	0.93	0.99	0.92	0.95	0.98	0.91
PepMHC-I	0.88	0.99	0.94	0.85	1.00	0.89
SBS	0.93	0.99	0.95	0.73	0.97	0.89
St+	0.86	0.91	0.90	0.92	0.83	0.83
SuperMHC	0.90	0.98	0.94	0.92	0.97	0.87
SuperMHCR	0.92	0.98	0.94	0.74	0.91	0.87
Maximum value	0.97	0.99	0.96	0.99	1.00	0.97

Table 3. Average classification performance of each predictor. The best prediction results are shaded.

	AVE	Rank AVE	AVE 9-mers	Rank AVE 9-mgers	AVE 10-mers	Rank AVE 10-mers
BIMAS	0.93	(12)	0.94	(14)	0.92	(11)
SYFPEITHI	0.85	(19)	0.87	(19)	0.82	(19)
Bunsen	0.94	11	0.96	8	0.92	11
Frags	0.89	16	0.87	19	0.90	14
FudanCS	0.95	6	0.96	8	0.94	9
H00001	0.96	4	0.97	1	0.95	5
hnp	0.95	6	0.96	8	0.95	5
I00001	0.96	4	0.97	1	0.95	5
imaginary1	0.81	19	0.97	1	0.65	19
imaginary2	0.95	6	0.93	17	0.97	3
imaginary3	0.95	6	0.97	1	0.93	10
Ipp729	0.90	15	0.94	14	0.86	15
MHChackers1	0.52	20	0.52	20	0.52	20
N00001	0.97	2	0.97	1	0.97	3
N00002	0.98	1	0.97	1	0.98	1
N00003	0.97	2	0.97	1	0.98	1
P00001	0.95	6	0.95	12	0.95	5
PepMHC-I	0.93	12	0.94	14	0.91	13
SBS	0.91	14	0.96	8	0.86	15
St+	0.88	18	0.89	18	0.86	15
SuperMHC	0.93	12	0.94	14	0.92	11
SuperMHCR	0.89	16	0.95	12	0.84	18
Maximum value	0.98		0.97		0.98	

Table 4. Prediction of binding affinity performance for competitors and benchmark predictors.

	CORRELATION COEFFICIENT					
	A0101 9mer	**A0201 9mer**	**B0702 9mer**	**A0101 10mer**	**A0201 10mer**	**B0702 10mer**
BIMAS	0.293	0.448	**0.513**	**0.455**	0.496	**0.576**
SYFPEITHI	**0.553**	**0.775**	0.324	0.174	**0.750**	0.453
Bunsen	0.609	0.848	0.688	0.408	0.855	0.526
Frags	0.384	0.642	0.539	0.490	0.697	0.447
FudanCS	0.377	0.675	0.492	0.267	0.535	0.314
H00001	0.406	0.706	0.555	0.332	0.641	0.353
hnp	0.632	0.900	0.728	0.448	0.803	0.514
I00001	0.541	0.773	0.626	0.498	0.810	0.442
imaginary1	0.739	0.888	0.764	0.454	0.196	0.216
imaginary2	0.557	0.826	0.656	0.603	0.818	0.567
imaginary3	0.193	0.311	0.232	0.376	0.718	0.299
Ipp729	0.471	0.856	0.542	0.307	0.733	0.437
MHChackers1	0.016	0.153	0.031	0.054	0.015	0.001
N00001	0.766	0.931	0.784	0.7607	0.930	0.603
N00002	0.731	0.916	0.785	0.749	0.924	0.663
N00003	0.757	0.925	0.790	0.7606	0.929	0.640
P00001	0.509	0.844	0.622	0.525	0.851	0.536
PepMHC-I	0.476	0.773	0.638	0.389	0.857	0.415
SBS	0.592	0.826	0.675	0.514	0.786	0.530
St+	0.494	0.454	0.565	0.681	0.498	0.258
SuperMHC	0.552	0.873	0.690	0.516	0.795	0.466
SuperMHCR	0.605	0.905	0.700	0.293	0.669	0.459
Maximum value	0.766	0.931	0.790	0.761	0.930	0.663