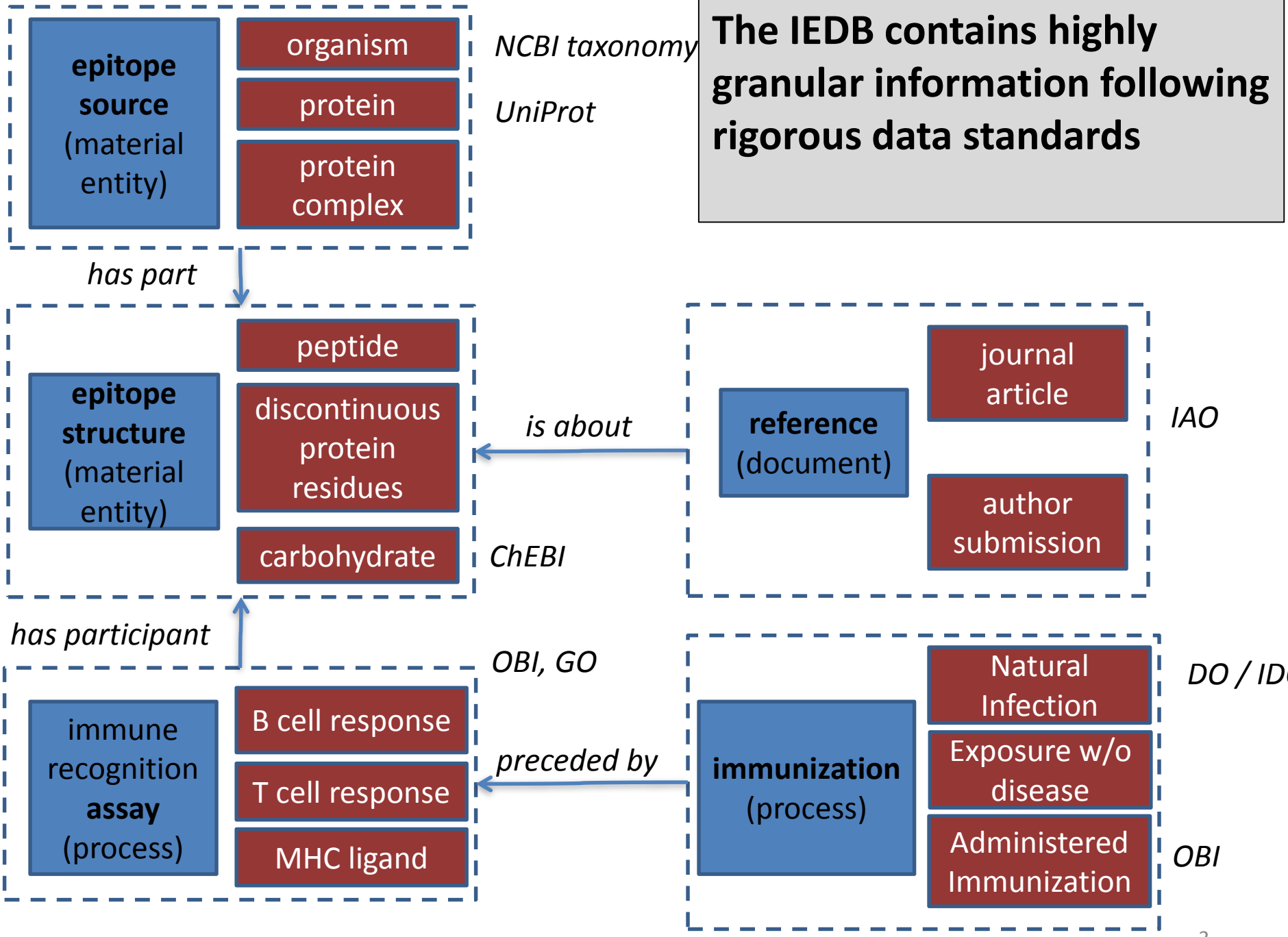
MHC binding and immunogenicity of eluted ligands; benchmarking and predictions

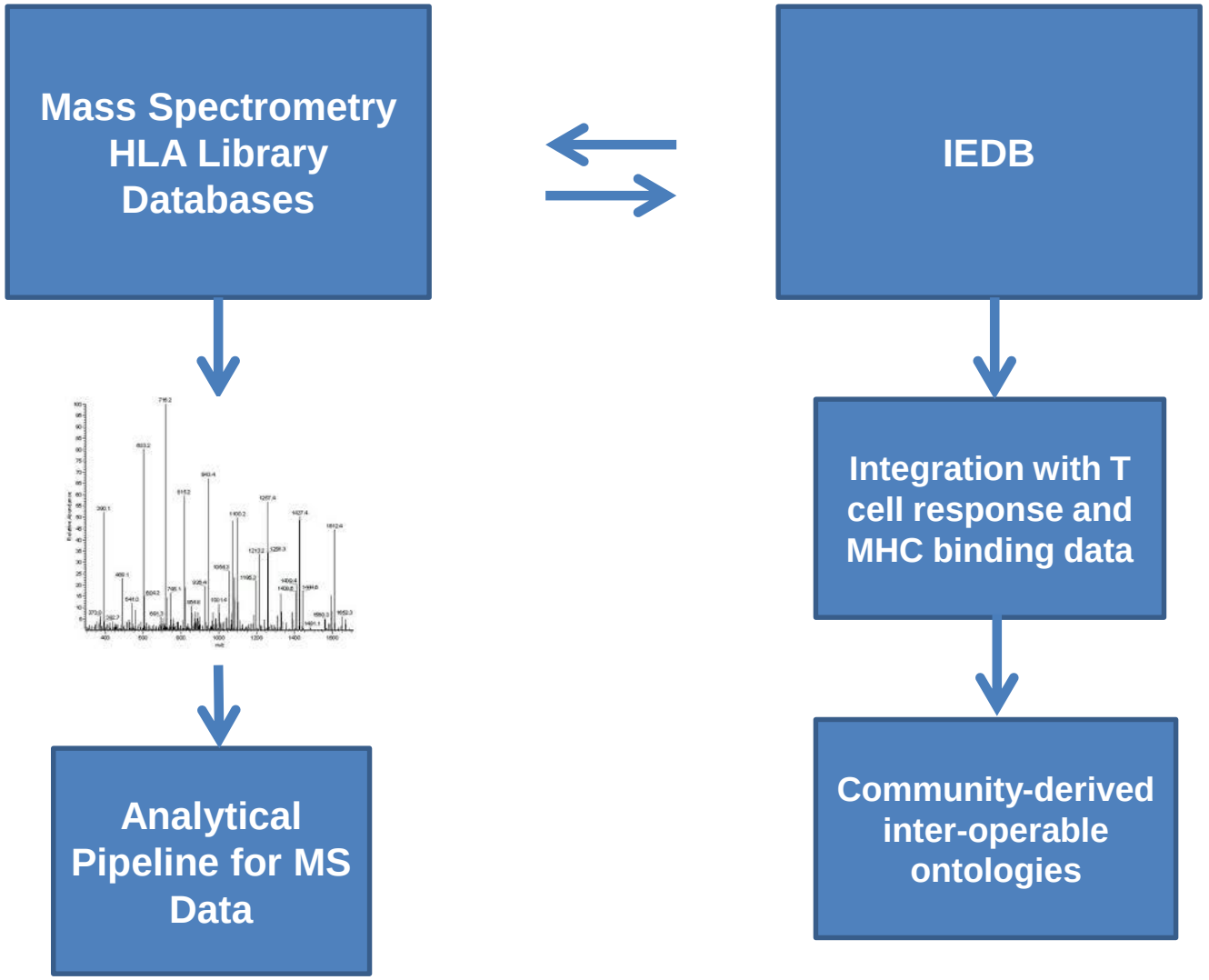
Outline

- The IEDB and naturally processed (eluted ligand) data (and associated data standards)

The IEDB contains highly granular information following rigorous data standards



Our vision for IEDB Integration with the immunopeptidomic community



Metadata and the IEDB

Proteomics. 2018 Jun; 18(12): 1800110.
Published online 2018 Jun 27. doi: 10.1002/pmic.201800110
PMCID: PMC6033177
PMID: 29791771

Minimal Information About an Immuno-Peptidomics Experiment (MIAIPE)

Jennie R. Lill, [1] Peter A. van Veelen, [2] Stefan Tenzer, [3] Arie Admon, [4] Etienne Caron, [5] Joshua E. Elias, [6] Albert J.R. Heck, [7,8] Miguel Marcilla, [9] Fabio Marino, [10,11] Markus Müller, [12] Bjoern Peters, [13] Anthony Purcell, [14] Alessandro Sette, [13,15] Theo Sturm, [7,8] Nicola Ternette, [16] Juan Antonio Vizcaíno, [17] and Michal Bassani-Sternberg [10,11]

- Ligand origin at host, cellular and protein level
  - IEDB utilizes taxonomy, cell-ontology and protein trees based on rigorous, community derived and interoperable ontologies and nomenclatures
- Information related to the biological process related to the experiment (e.g. infection, transfection, pulse with exogenous protein)
- Technical information related to the experiment, elution (MS, technique for assignment, etc.)
- HLA/MHC isolation classification and assignment based on published, rigorous and transparent MHC ontology standards
- Separate records for each literature or submission origin

T cell and MHC binding data is mostly non-self while elution data (NP) is mostly self

	Unique epitopes		
	MHC	**T cell**	**NP**
Total	**52,060**	**38,182**	**141,331**
Human	42,733	25,656	126,078
Rodent	6,416	13,052	15,679
Other hosts	628	1,380	157
Non-human primates	3,592	602	0
Self	**5,161**	**7,461**	**139,413**
Non-self	**39,801**	**28,331**	**1529**
Viruses	31,512	15,178	959
Allergen	1,264	4,458	133
Bacteria	5,219	6,332	128
Parasites	1,564	2,027	307
Fungus	242	336	2

Vaughan K, Xu X, Caron E, Peters B, Sette A. Expert Rev Proteomics. 2017 Sep;14(9):729-736.

Exemplary studies starting to fill in the gaps

- Croft, Purcell and La Gruta identified identified 21 influenza A virus (IAV)-derived peptides presented by murine H-2 b class I complexes following direct in vitro infection or cross-presentation
- Cross-presented epitope abundance and peptide-MHCI binding strength (IC 50) to be the most powerful predictors of CTL response magnitude.

Ting Wu et al, submitted

Conclusions (I)

- The IEDB, currently contains more than 320,000 elution data (140,000 elution data in 2017)
- Data resources integrating immunological metadata and NP data will become available to the user community
- Integration of NP data housed within the IEDB with SWATHAtlas [HUPO-HIPP partner]
  - ensure interoperability of data repositories to allow full access to all NP data
  - for linking peptidome data with immune reactivity data
- Elution data is mostly derived from self-antigens and fewer ligands from pathogens and allergens
  - the overlap between NPs, T cell epitopes and MHC binding data is poor
  - as a result it is difficult to address correspondence

Outline

- The IEDB and naturally processed (eluted ligand) data (and associated data standards)
- The need for global benchmarking (binding, eluted ligands, T cell recognition)

First comprehensive benchmark of MHC binding predictions with initiation of IEDB

- 48 MHC alleles, 88 datasets
- 48,828 IC_{50} values
- 50 – 3000 data points per dataset

Results:

- Training data volume >> method
- AUC values > 0.9 for IEDB methods
- NetMHC >SMM >> ARB

Allele	Peptide Length	#IC50	Allele	Peptide Length	#IC50	Allele	Peptide Length	#IC50	Allele	Peptide Length	#IC50
HLA A*0101	9	1157	HLA A*3002	9	92	HLA B*4501	9	114	Mamu A*01	8	383
HLA A*0101	10	56	HLA A*3101	9	1869	HLA B*5101	9	244	Mamu A*01	9	525
HLA A*0201	9	3089	HLA A*3101	10	1057	HLA B*5101	10	177	Mamu A*01	10	477
HLA A*0201	10	1316	HLA A*3301	9	1140	HLA B*5301	9	254	Mamu A*01	11	293
HLA A*0202	9	1447	HLA A*3301	10	1055	HLA B*5301	10	177	Mamu A*02	8	150
HLA A*0202	10	1056	HLA A*6801	9	1141	HLA B*5401	9	255	Mamu A*02	9	283
HLA A*0203	9	1443	HLA A*6801	10	1055	HLA B*5401	10	177	Mamu A*02	10	211
HLA A*0203	10	1055	HLA A*6802	9	1434	HLA B*5701	9	59	Mamu A*02	11	201
HLA A*0206	9	1437	HLA A*6802	10	1051	HLA B*5801	9	988	Mamu A*11	8	217
HLA A*0206	10	1054	HLA A*6901	9	833	H-2 Db	9	303	Mamu A*11	9	468
HLA A*0301	9	2094	HLA B*0702	9	1262	H-2 Db	10	134	Mamu A*11	10	277
HLA A*0301	10	1082	HLA B*0702	10	205	H-2 Dd	9	85	Mamu A*11	11	214
HLA A*1101	9	1985	HLA B*0801	9	708	H-2 Dd	10	75	Mamu B*01	8	155
HLA A*1101	10	1093	HLA B*1501	9	978	H-2 Kb	8	480	Mamu B*01	9	205
HLA A*2301	9	104	HLA B*1801	9	118	H-2 Kb	9	223	Mamu B*01	10	185
HLA A*2402	9	197	HLA B*2705	9	969	H-2 Kd	9	176	Mamu B*01	11	208
HLA A*2402	10	78	HLA B*3501	9	736	H-2 Kd	10	70	Mamu B*17	8	154
HLA A*2403	9	254	HLA B*3501	10	177	H-2 Kk	8	80	Mamu B*17	9	300
HLA A*2601	9	672	HLA B*4001	9	1078	H-2 Kk	9	164	Mamu B*17	10	198
HLA A*2902	9	160	HLA B*4002	9	118	H-2 Kk	10	57	Mamu B*17	11	191
HLA A*2902	10	55	HLA B*4402	9	119	H-2 Kk	11	51	Patr A*0901	11	89
HLA A*3001	9	669	HLA B*4403	9	119	H-2 Ld	9	102	Patr B*0101	9	132

Peters, PLoS Comp Biol, 2006

Bioinformatics. 2015 Jul 1; 31(13): 2174–2181.
Published online 2015 Feb 25. doi: 10.1093/bioinformatics/btv123
PMCID: PMC4481849
PMID: 25717196

Automated benchmarking of peptide-MHC class I binding predictions

Thomas Trolle,[1] Imir G. Metushi,[2] Jason A. Greenbaum,[2] Yohan Kim,[2] John Sidney,[2] Ole Lund,[1] Alessandro Sette,[2] Bjoern Peters,[2,*] and Morten Nielsen[1,3,*]

An automated benchmarking platform for MHC class II binding prediction methods

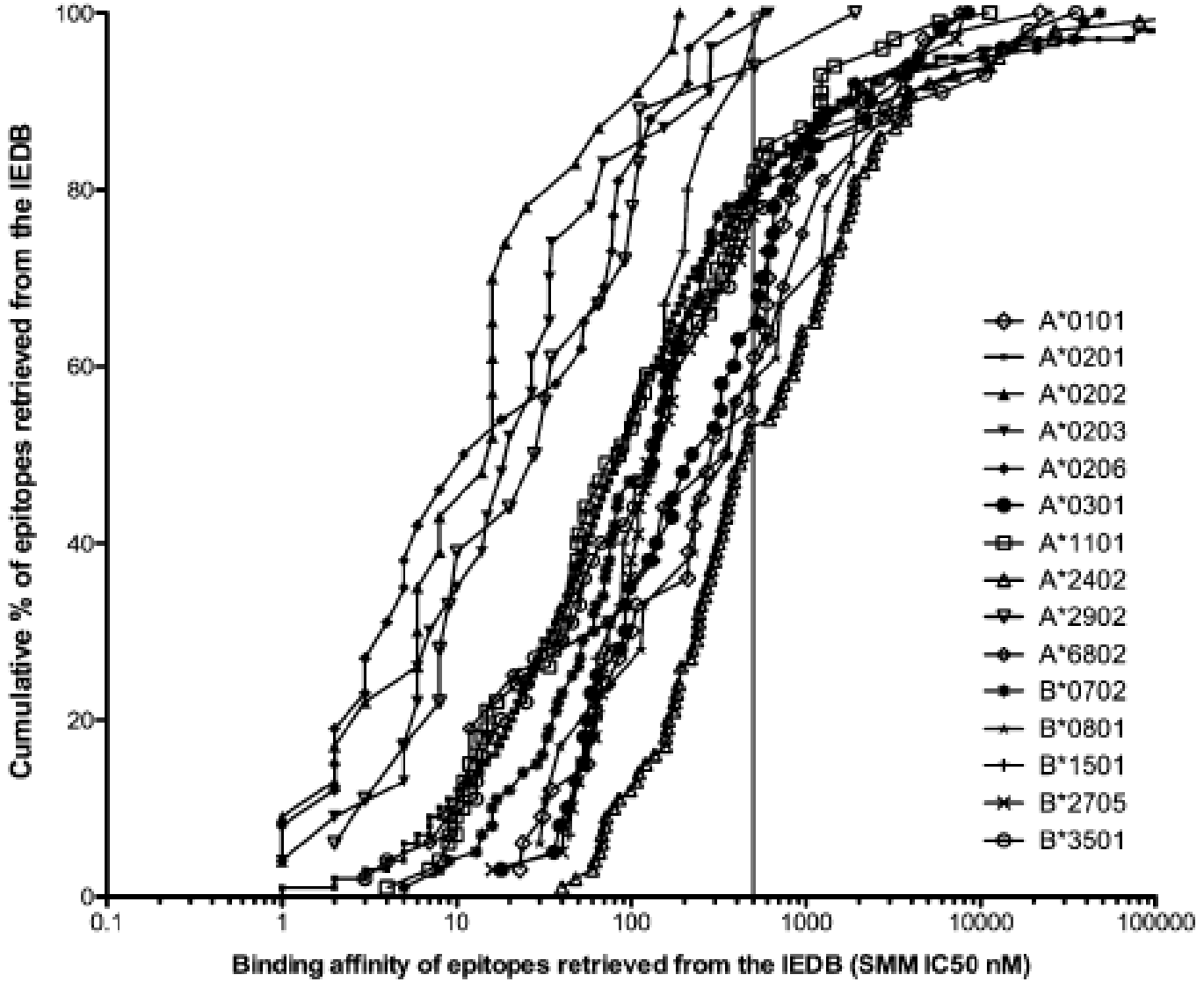
Massimo Andreatta, Thomas Trolle, Zhen Yan, Jason A Greenbaum, Bjoern Peters, Morten Nielsen ✉

Bioinformatics, Volume 34, Issue 9, 1 May 2018, Pages 1522–1528, https://doi.org/10.1093/bioinformatics/btx820

MHC binding affinity as a predictor of immunogenicity

~80% of epitopes bind <500 nM, supporting historic threshold

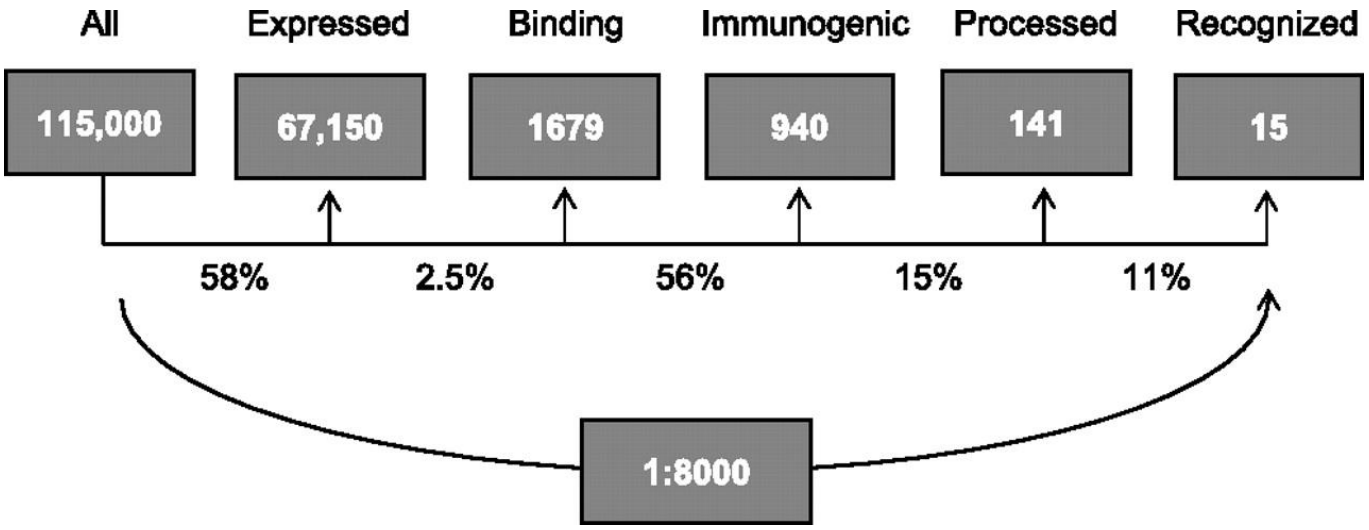
Different alleles have different affinity distribution → ranks / allele specific thresholds are preferred when combining



Paul, J Immunol, 2013

Benchmarking MHC binding to predict MHC binding vs T cell immunogenicity

- Predictive tools trained on MHC binding predict binding generally well
- Benchmarking MHC binding as a predictor of T cell immunogenicity
  - Typically all T cell epitopes are binders, but only 5- 10 % of binders are immunogenic

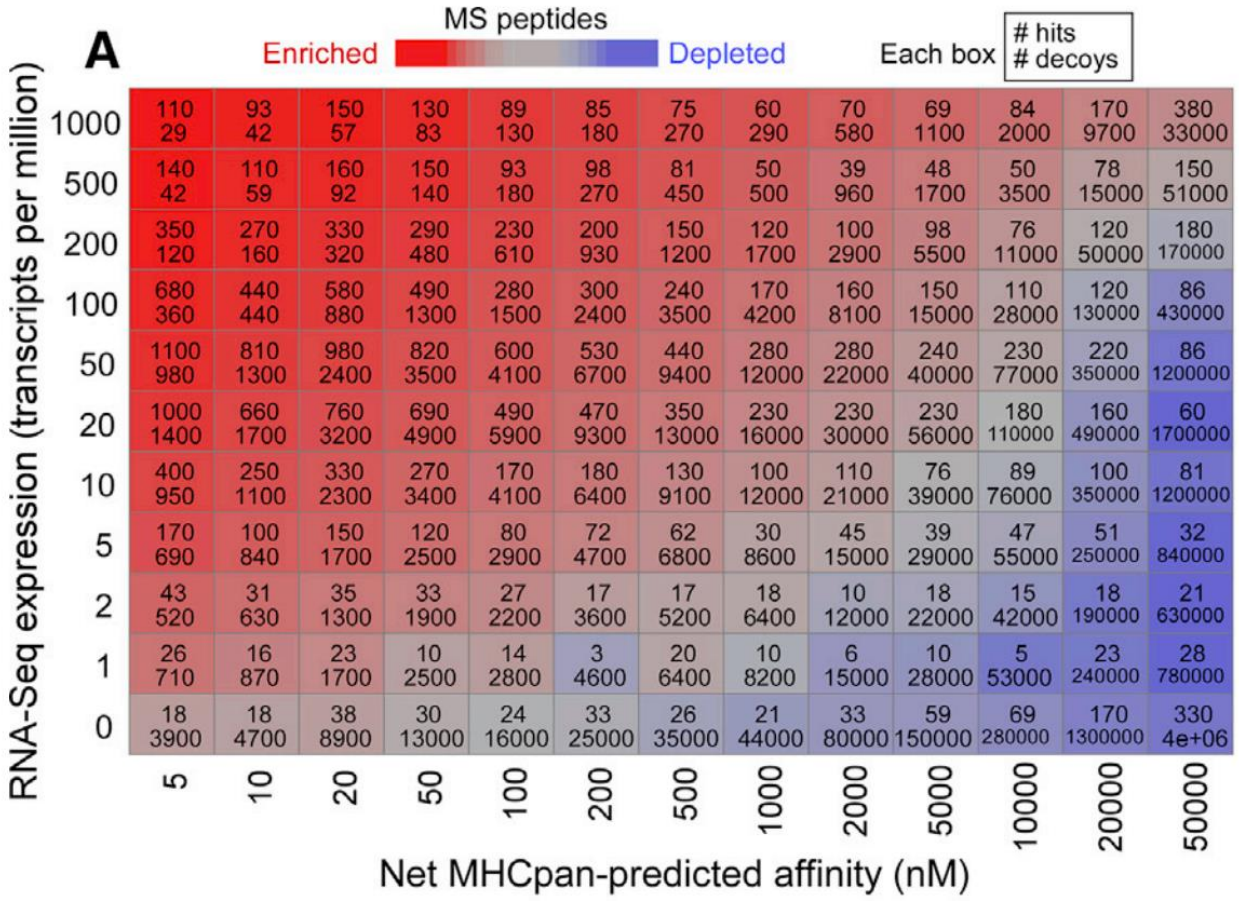


Quantitative impact of variables influencing immunodominance. Assarsson et al. J Immunol 2007;178:7890-7901

Benchmarking elution data to predict eluted peptides and T cell immunogenicity

- Elution data are increasingly used to train algorithms, that predict elution data very well
  - Combination with tools trained on binding data affords additional gains (later part of the talk)
- However, benchmarking of elution data (real data, not predicted) to predict T cell epitope is largely missing
- It is likely that all “true” epitopes are naturally processed but
  - How many of them are detected vs missed given the limits of sensitivity of the assays?
- Likewise, how many of the eluted ligands are immunogenic?

Binding affinity and Presentation

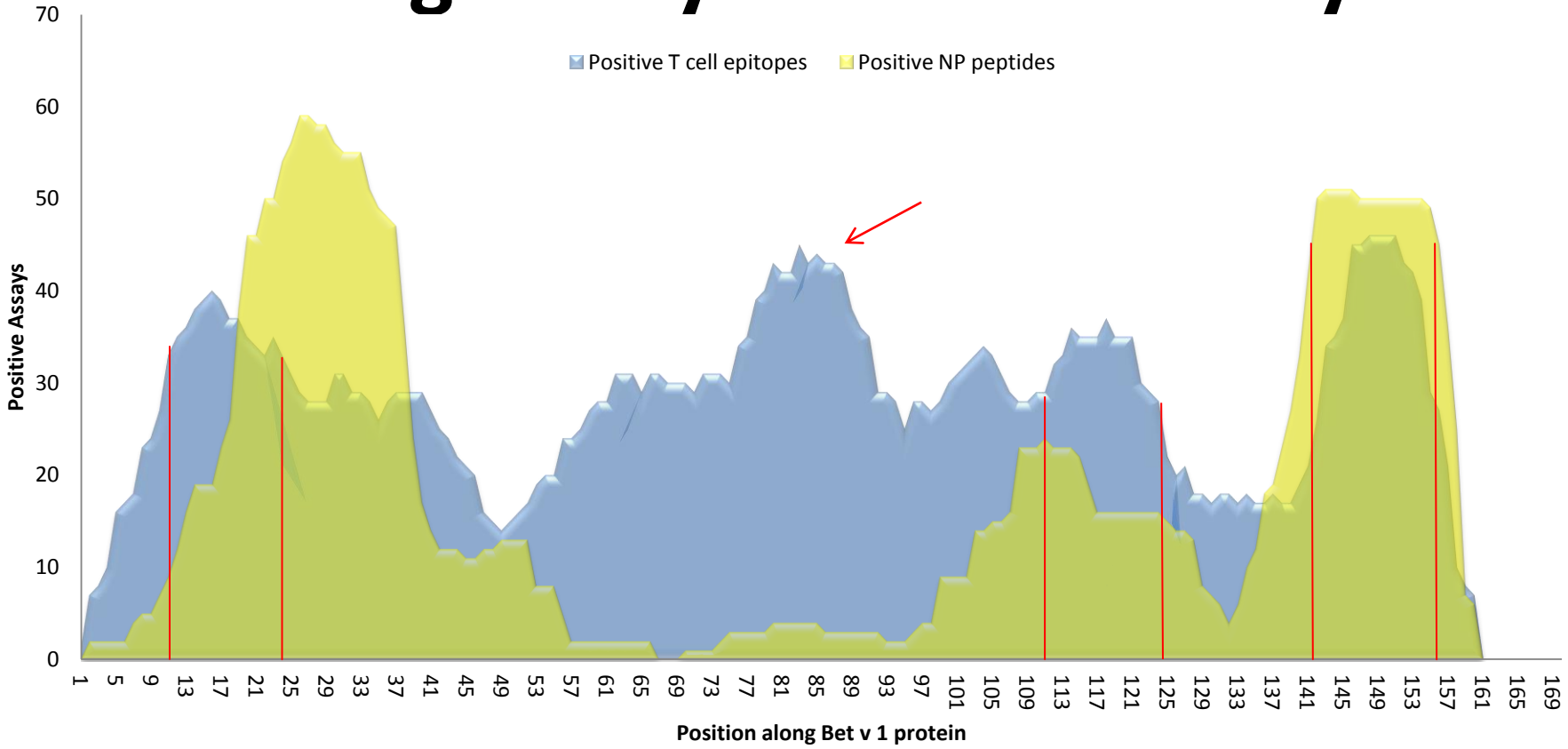


"This revealed a multiplicative relationship between expression and affinity, in which **a 10-fold increase in expression could approximately compensate for a 90% decrease in binding potential**"

Abelin et al, Immunity *46*, 315-326 (2017)

Slide courtesy of Josh Elias

Benchmarking eluted peptides and immunogenicity in the Bet v 1 system



J Allergy Clin Immunol. 2010 Mar;125(3):711-8, 718.e1-718.e2. doi: 10.1016/j.jaci.2009.10.052. Epub 2010 Feb 4.

Naturally processed T cell-activating peptides of the major birch pollen allergen.

Mutschlechner S[1], Egger M, Briza P, Wallner M, Lackner P, Karle A, Vogt AB, Fischer GF, Bohle B, Ferreira F.

Further consideration relating to binding affinity and abundance of eluted ligands

- High abundance can compensate for low MHC affinity
  - but how effective will such ligand be in terms of immunogenicity???
- Conversely, a low abundance ligand that binds with high affinity may be
  - less easily detected
  - more strongly immunogenic

Further benchmarking studies

- a recent study in collaboration with Josh Elias, reported elution data from the DENV infected Raji cell line
- the *in vitro* elution data identified only 10/58 previously described DENV T-cell epitopes, restricted by the Raji expressed HLAs
- the study identified 5 low binding epitopes, and at least two of these novel MHC ligands were recognized by T-cells from DENV-infected patients
- while HLA binding over predicts (few binders are actually epitopes); elution data may under predict (it identifies only the tip of the iceberg)
- A further issue. Raji is a B cell line, which is not a major cell type infected by DENV *in vivo*

Swaminathan et al. Submitted

Benchmarking elution data to predict immunogenicity

- How representative is *in vitro* culture with cell lines of *in vivo* processing in tissues?

Sci Data. 2018; 5: 180157.
Published online 2018 Aug 7. doi: 10.1038/sdata.2018.157
Data Descriptor

PMCID: PMC6080492
PMID: 30084848

A tissue-based draft map of the murine MHC class I immunopeptidome

Heiko Schuster,[1,2,3,*] Wenguang Shao,[4,*] Tobias Weiss,[5] Patrick G.A. Pedrioli,[4] Patrick Roth,[5] Michael Weller,[5] David S. Campbell,[6] Eric W. Deutsch,[6] Robert L. Moritz,[6] Oliver Planz,[1] Hans-Georg Rammensee,[1,3] Ruedi Aebersold,[a,4,7] and Etienne Caron[b,4]

- In this study, over 3,000 ligands were identified from 18 different tissues, and over 50% of the eluted ligands were found only in 2or 3 tissue
- This underlines the remarkable tissue specificity of protein expression and pathogen tropism

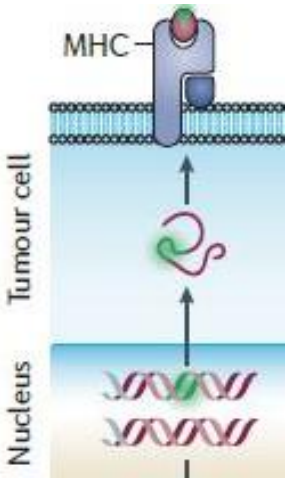
Conclusions (II)

- MHC binding predictions predict MHC binding effectively
  - Ongoing Automated benchmarking on going for both class I and class II
- MHC binding predictions have been benchmarked for predicting T cell immunogenicity
  - More in the next section
- Availability of large elution datasets continue to increases
  - These data are being utilized to derive algorithms predicting elution data
- Benchmarking studies will start to allow benchmarking elution data in terms of prediction of T cell epitopes

Outline

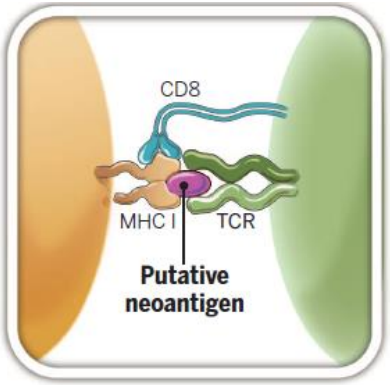
- The IEDB and naturally processed (eluted ligand) data (and associated data standards)
- The need for global benchmarking (binding, eluted ligands, T cell recognition)
- A cancer epitope prediction pipeline and benefitting from training with elution data

A cancer epitope prediction pipeline



- Cancers genomes accumulate mutations
- Mutations in coding regions are translated mutated protein sequences
- Mutated peptides can be presented as epitopes on MHC to T cells
- How well do tools perform in predicting immunogenic neoepitopes?

Neoepitopes are recognized by tumor-infiltrating lymphocytes **(TILs)**



Coulie et al, Nat Rev Cancer. 2014 Feb;14(2):135-46
Schumacher & Schreiber, Science. 2015 Apr 3;348(6230):69-74

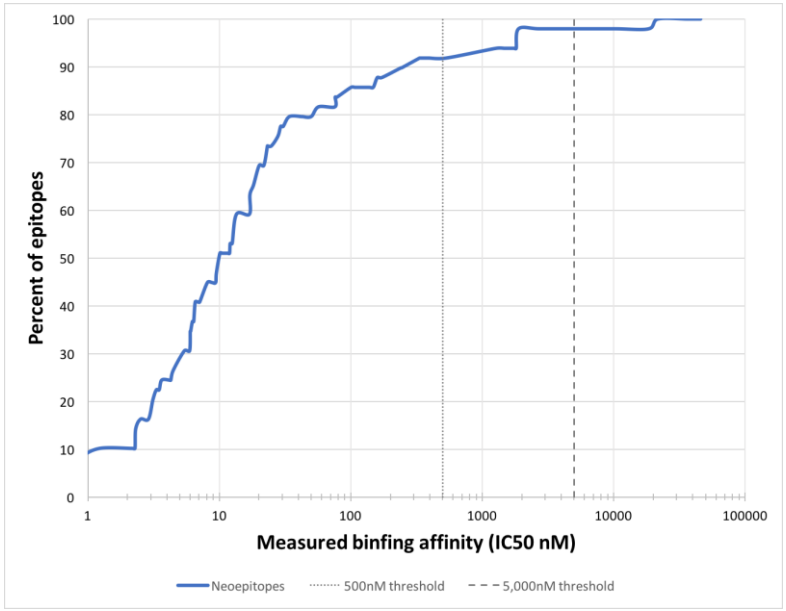
Datasets utilized in the study

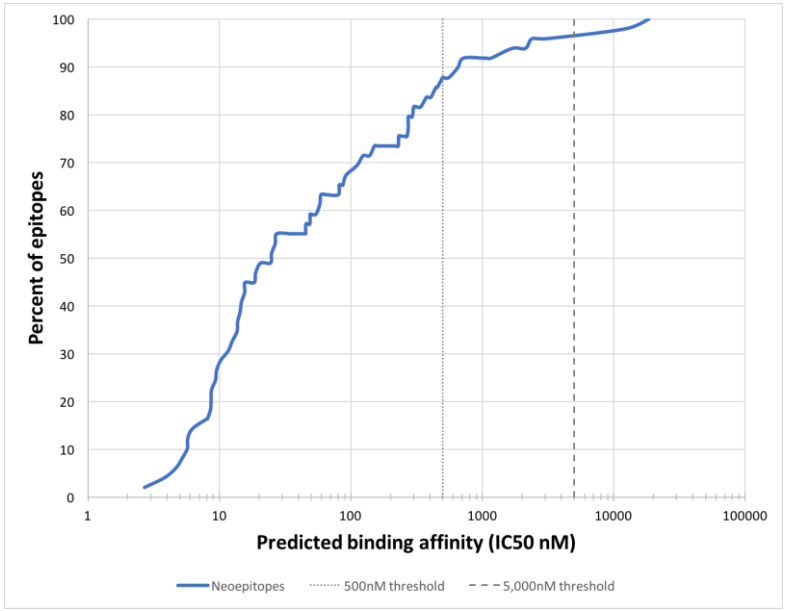
- Training set of 78 neoepitopes curated from literature
  - demonstrated T cell response (e.g. ELISPOT)
  - neoantigen specific T cell recognizes cancer cell
- Negative control control data set generated from the same antigens
- Validation dataset provided by collaborators at NCI

Epitope Set	total number of analyzed peptides
Neoepitopes round 1	49
Neoepitopes round 2	29
NCI 25mers	52 positives 2,760 negatives

Oncoimmunology, in press

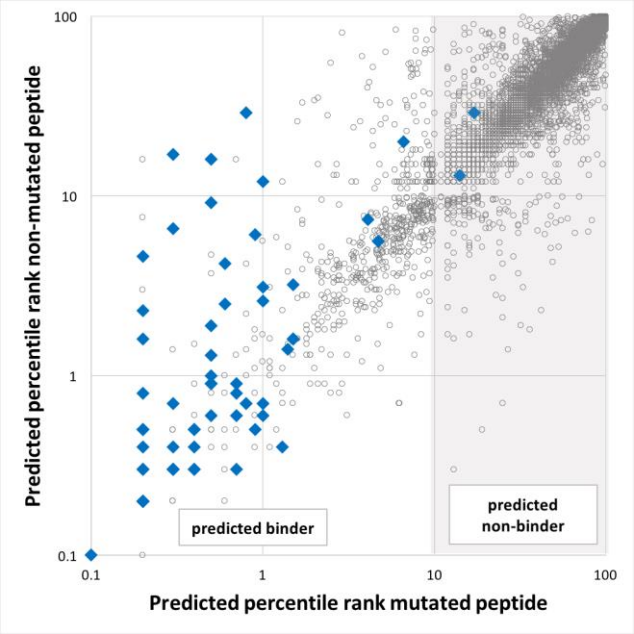
Previously established 500nM threshold identifies >90% of neoepitopes



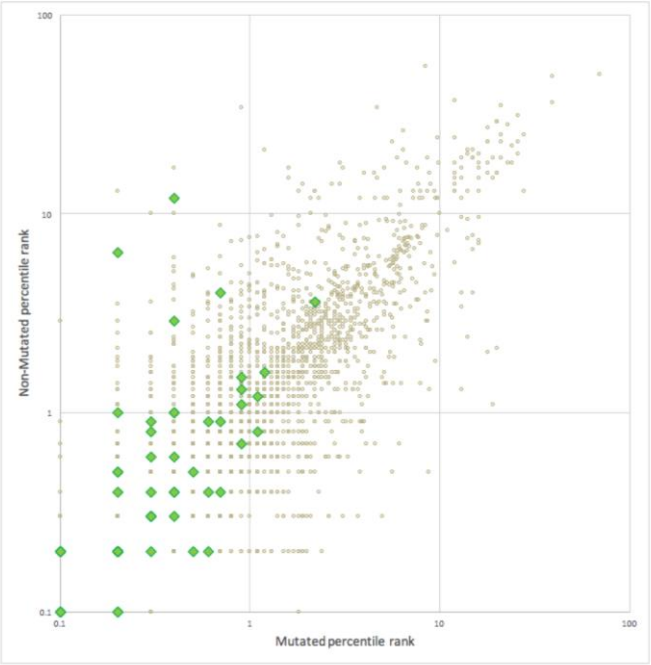


NetMHCpan 2.8 for HLA binding prediction

Neoepitopes are predicted binders



Literature data



NCI data

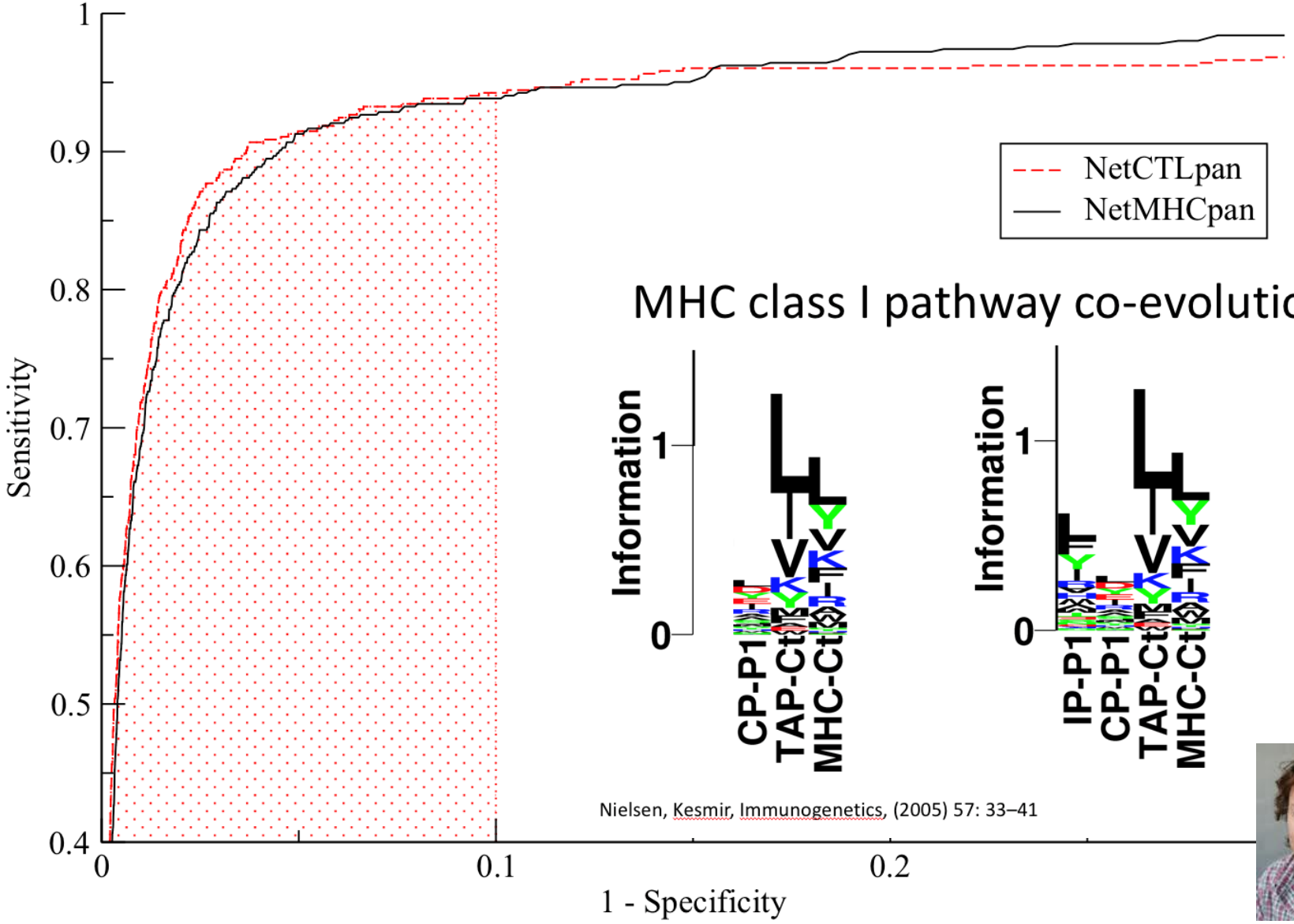
MHC Binding Predictions alone yields best performance

Tested tools:

- binding prediction
- proteasomal cleavage prediction
- TAP transport prediction
- peptide-MHC stability prediction
- similarity assessment (BLOSUM62)
- various combinations

Method	non-binders (percentile rank >= 10) removed
NetMHCpan - IC50	0.920
NetMHCpan - percentile rank	0.931
NetMHCpan - length-rescaled rank	0.952
NetMHCpan - length-rescaled rank & AP3 discarded	0.954
Proteasomal cleavage score	0.572
TAP transport score	0.584
Combined proteasomal cleavage & TAP transport score	0.589
Combined processing & binding	0.906
NetMHCstabpan score	0.860
NetMHCstabpan rank	0.890
Combined NetMHCstabpan score and binding	0.917
Combined NetMHCstabpan rank and binding	0.884
Similarity score	0.545
Immunogenicity score	0.562
Immunogenicity Threshold	0.953
Combined	0.956

Does prediction of antigen processing matter? Not really…



Nielsen, Kesmir, Immunogenetics, (2005) 57: 33–41

Size adjustment of predicted percentile ranks

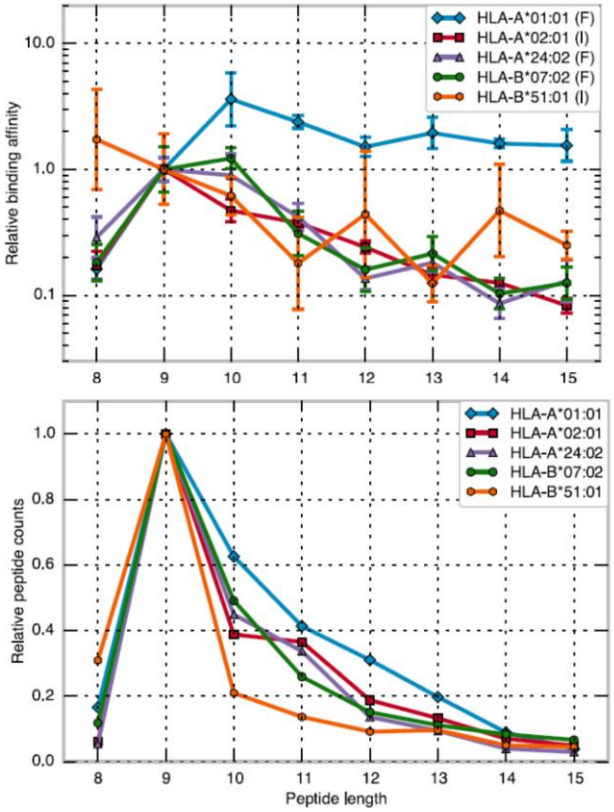
- Different MHC alleles have different length preferences
- Are these differences also seen in the length distribution of naturally processed peptides?
- Analysis of peptides eluted from secreted MHCs
- Predominant presentation of 9mers for all analyzed MHCs
- Allele-specific length preferences for eluted ligands

MHC binding length preference

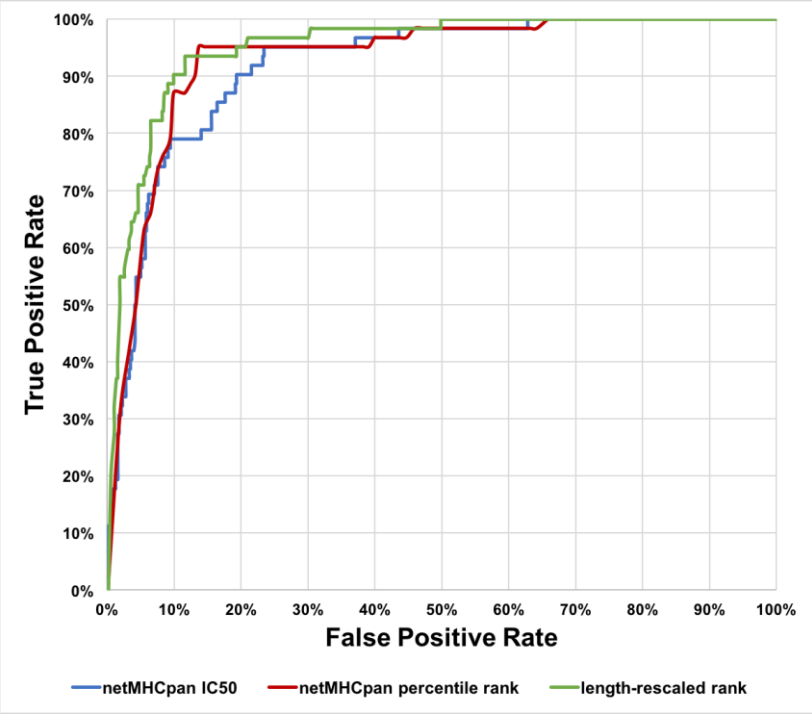
differs from

Length distribution of naturally processed peptides

Trolle, J Immunol, 2016



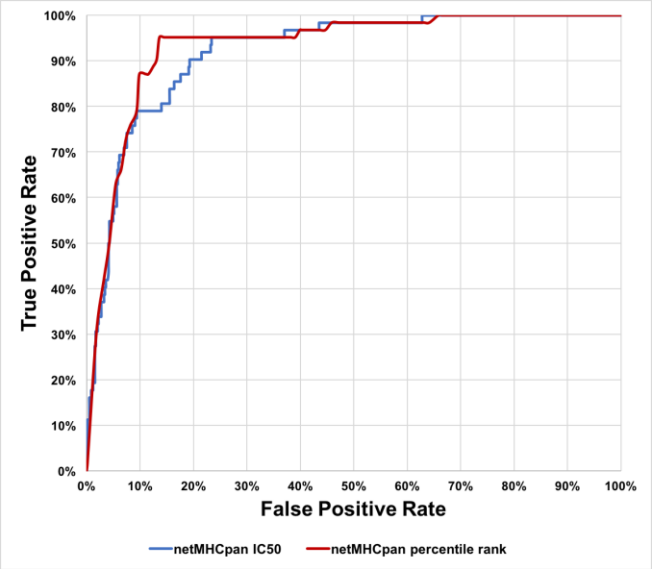
Length adjustment of prediction further boosts performance



Method	AUC
NetMHCpan - IC50	0.920
NetMHCpan - percentile rank	0.931
NetMHCpan - length-rescaled rank	0.952

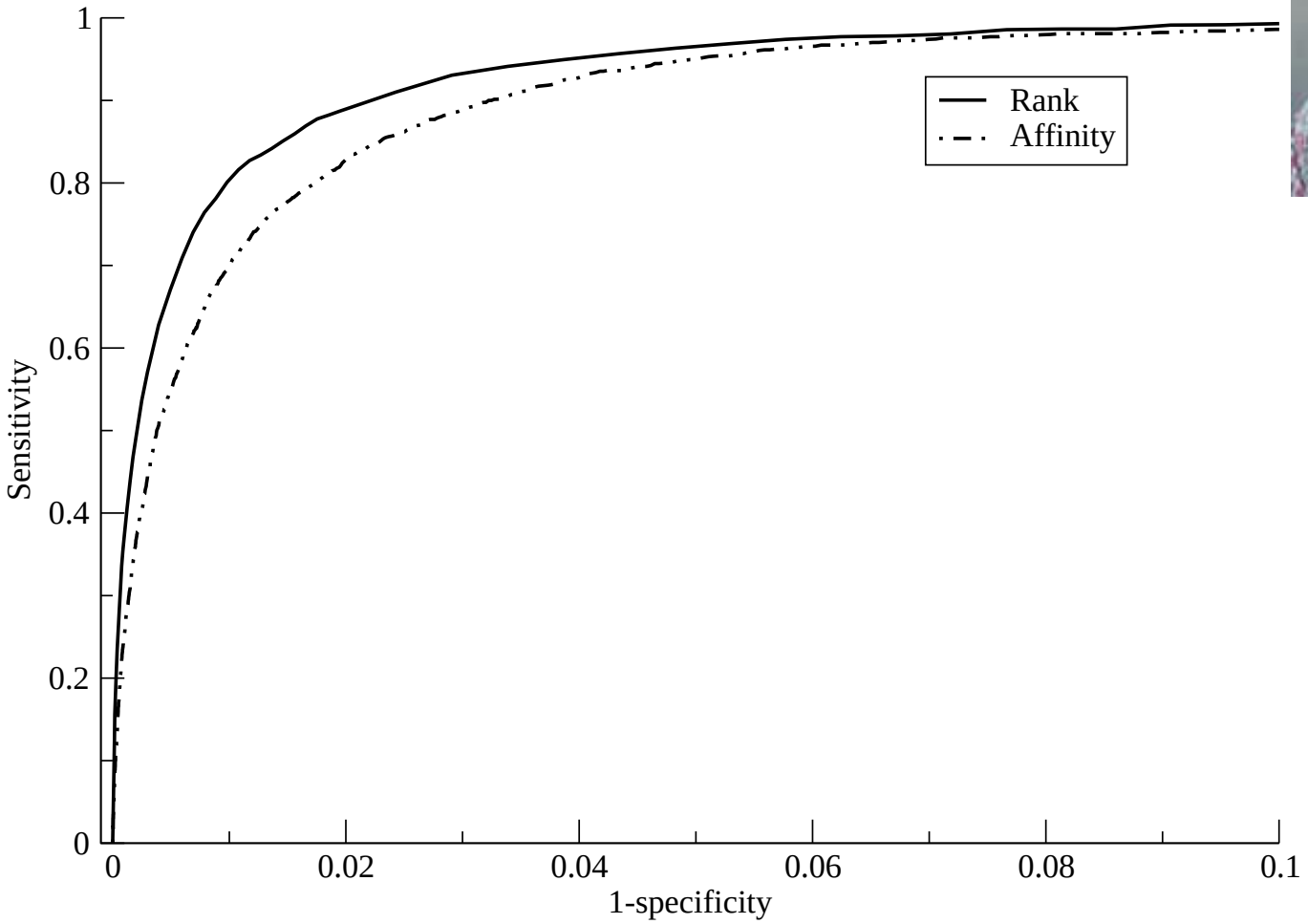
netMHCpan Predicted IC50 vs percentile rank

- Some alleles intrinsically bind more peptides than others
  - Different alleles have different affinity thresholds
- Percentile ranks normalize predictions across different alleles
  - Generated by comparing its score against the scores of 200,000 random natural peptides of the same length of the query peptide
- Prediction based on percentile ranks outperforms IC50



Method	AUC
NetMHCpan - IC50	0.920
NetMHCpan - percentile rank	0.931

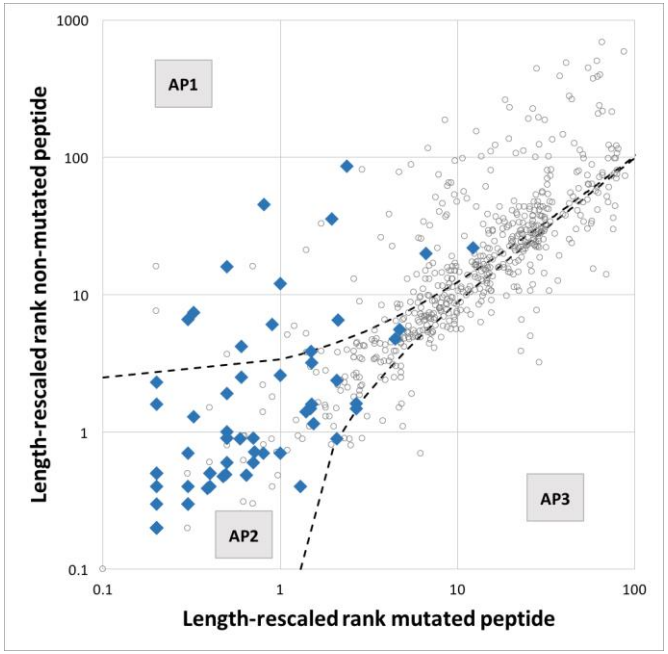
Findings consistent with previous studies from Nielsen's group



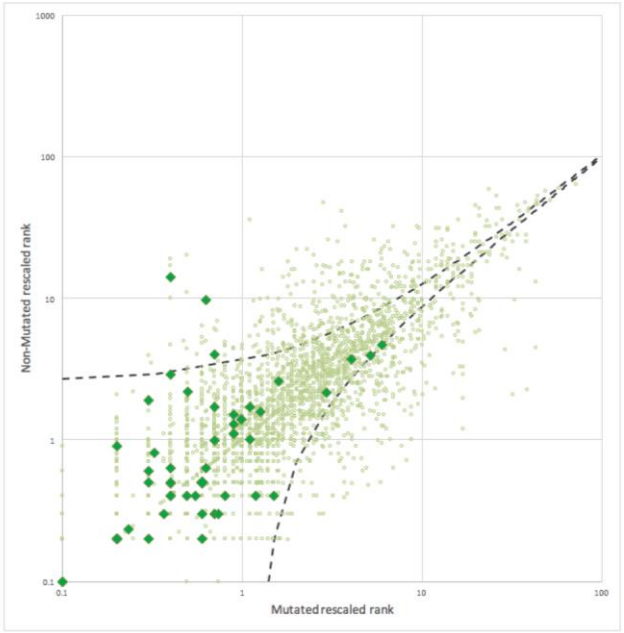
Andreatta M, Nielsen M, Bioinformatics (2016)

Nielsen M, Andreatta, M, Genome Medicine, (2016)

Binding affinity of most neoepitopes is comparable to wildtype peptides



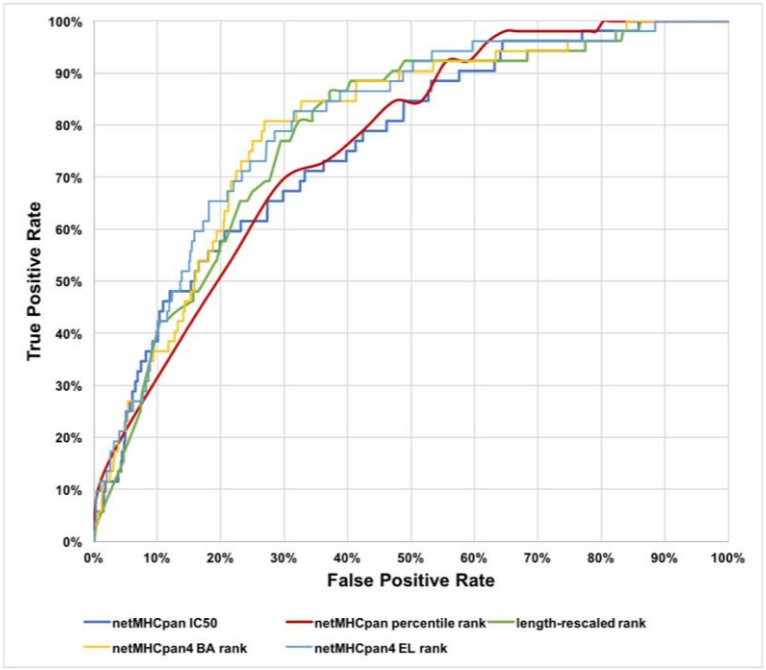
Literature data



NCI data

NetMHCpan version 4 outperforms all tools

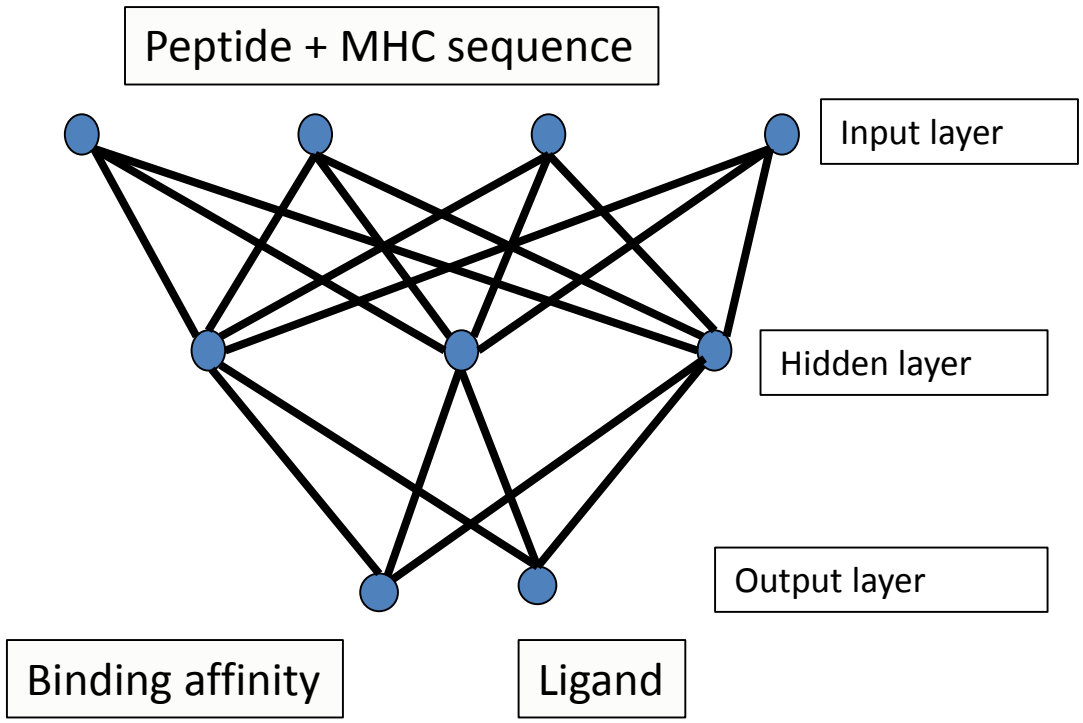
- NetMHCpan4 was trained on naturally eluted ligands as well as on binding affinity data.
- Likelihood of a peptide becoming a natural ligand (EL)
- Predicted binding affinity (BA)



NCI data

Method	All peptides	non-binders (percentile rank >= 10) removed
NetMHCpan - IC50	0.762	0.755
NetMHCpan - percentile rank	0.760	0.754
NetMHCpan - length-rescaled rank	0.783	0.777
Combined relaxed filters	0.786	0.782
NetMHCpan-4 BA percentile rank	0.798	0.792
NetMHCpan-4 EL percentile rank	**0.807**	**0.802**

How to learn from elution data?



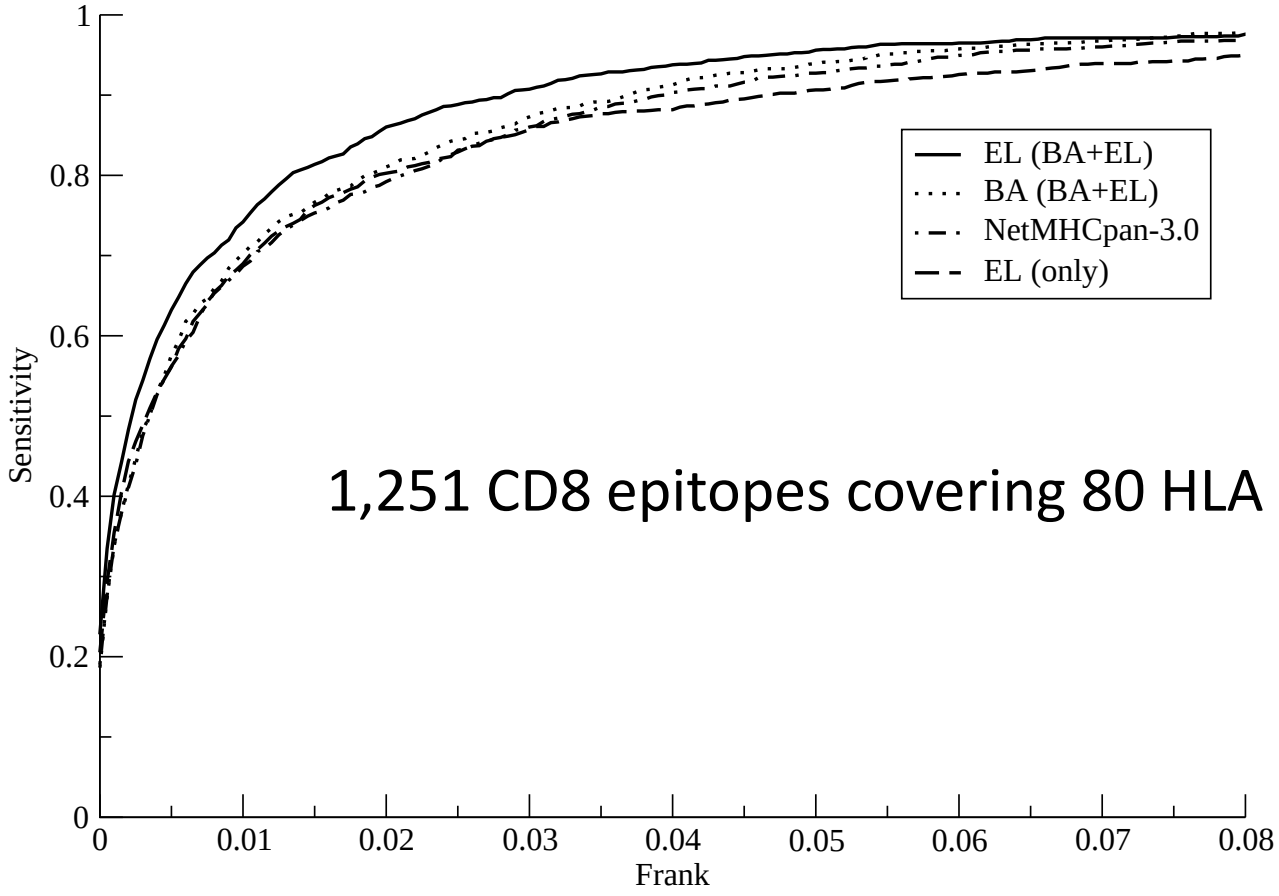
- Nielsen's group expanded the NNalign approach by adding a second output neuron
- Training is performed on both data simultaneously
- Resulting model is able to predict binding affinity value and likelihood of peptide being an eluted ligand

185,985 data points covering 153 MHC-I molecules

84,717 data points covering 55 HLA-I molecules

V. Jurtz et al. J Immunol. 2017

Validation of the approach on external data sets (II)



1,251 CD8 epitopes covering 80 HLA molecules

V. Jurtz et al. J Immunol. 2017

Conclusions (III)

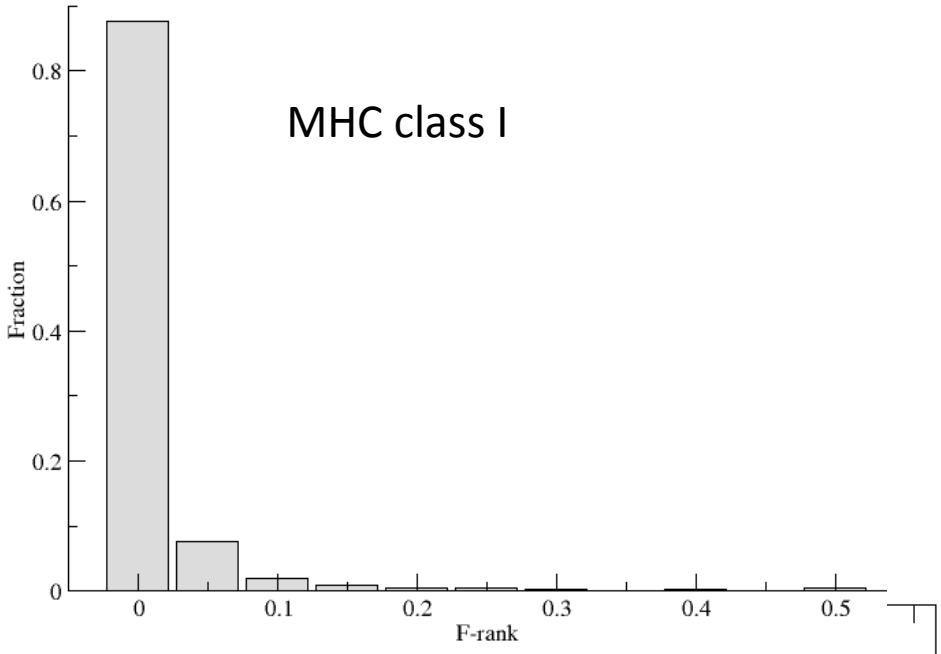
- Peptide binding can be accurately predicted using state-of-the-art prediction methods
- All CD8+ (neo)epitopes are high affinity binders to MHC
- Processing has limited impact on the prediction of CD8 T cell epitopes
- Integration of eluted ligand data into the prediction pipeline improves prediction accuracy for both MHC class ligands and T cell epitopes

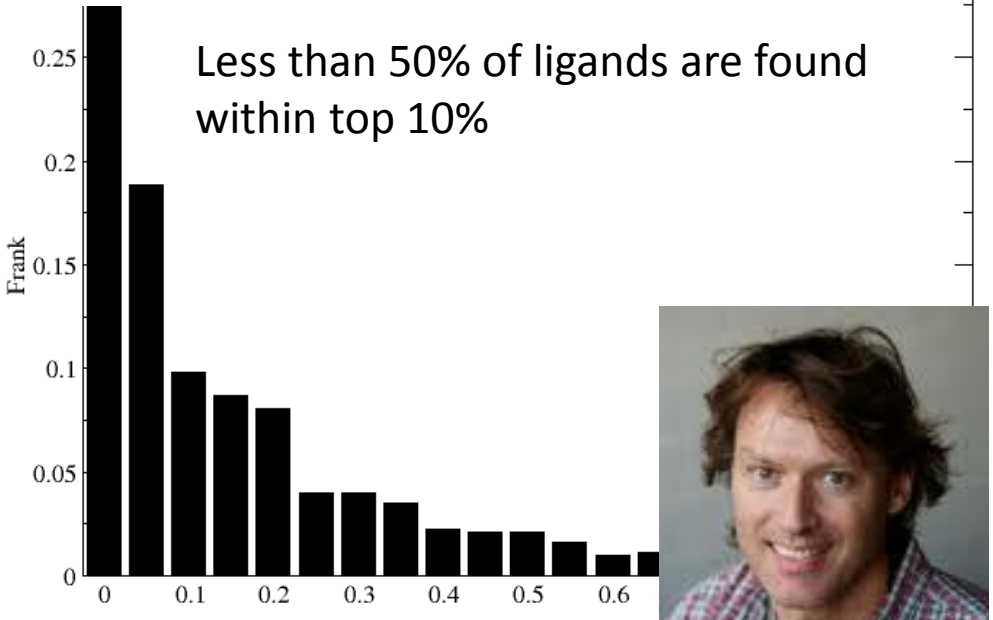
Outline

- The IEDB and naturally processed (eluted ligand) data (and associated data standards)
- The need for global benchmarking (binding, eluted ligands, T cell recognition)
- A cancer epitope prediction pipeline and benefitting from training with elution data
- Prediction of HLA class II epitopes

Predicting MHC class II ligands and T helper epitopes

- Performance remains relatively low (PCC=0.6-0.7) and many FP's when predicting T helper epitopes (MHC class II ligands)
- We have large data sets (>1000 measurements) available for most prevalent class II molecules
- And the picture does not change (much) with more data
- Same situation for other state-of-the-art methods (including measured binding affinity)

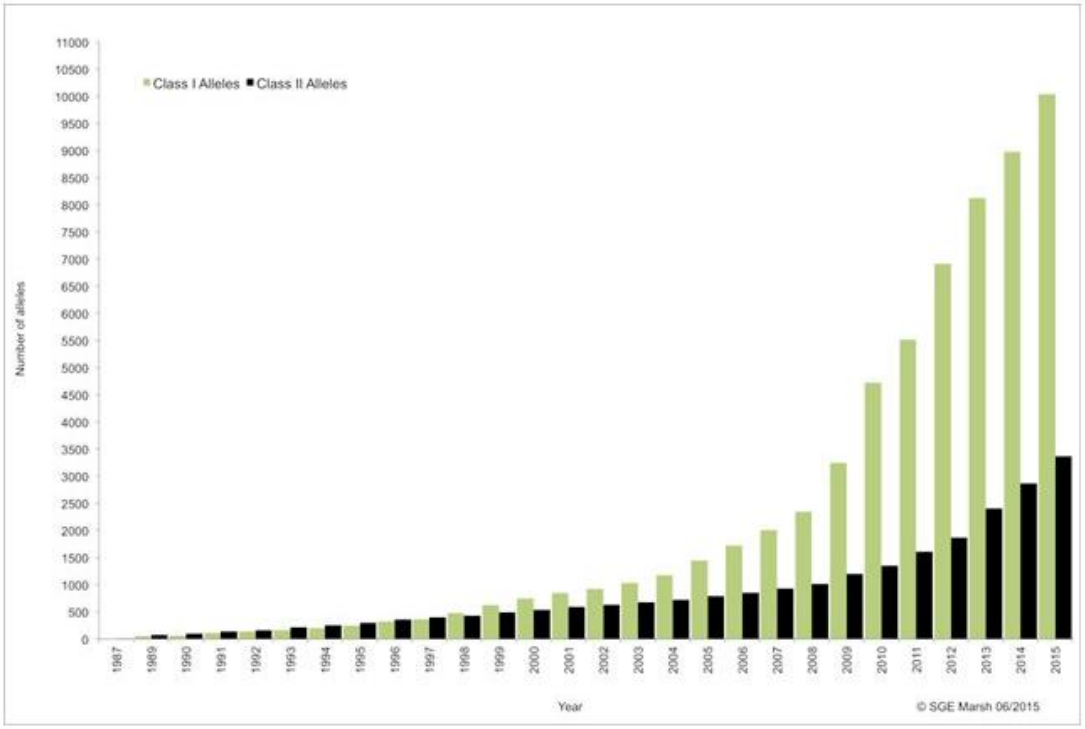




Key points for strategies using HLA class II binding to predict TCR recognition

- HLA binding is necessary but not sufficient for TCR recognition
- HLA binding predictions predict binding but not necessarily TCR recognition
- HLA binding predictions are allele specific
- However, most applications require predictions at the level of
  - individual subjects ->8 alleles
  - responding/treated population ->hundreds of alleles (usually not typed)
- What is required is an actionable strategies to target not alleles, but individuals and populations

Which HLA alleles should be considered?



IMGT/HLA database

12.672 HLA alleles

A mean of 450 new alleles/ Year

A much more limited panel of HLA class II alleles allows for global coverage

Locus	Molecule	Phenotype frequency
DRB1	DRB1*0101	5.4
	DRB1*0301	13.7
	DRB1*0401	4.6
	DRB1*0405	6.2
	DRB1*0701	13.5
	DRB1*0802	4.9
	DRB1*0901	6.2
	DRB1*1101	11.8
	DRB1*1201	3.9
	DRB1*1302	7.7
	DRB1*1501	12.2
	Combined	71.1
DRB3/4/5	DRB3*0101	26.1
	DRB3*0202	34.3
	DRB4*0101	41.8
	DRB5*0101	16.0
	Combined	87.7

Locus	Molecule	Phenotype frequency
DQA1/DQB1	DQA1*0501/DQB1*0201	11.3
	DQA1*0501/DQB1*0301	35.1
	DQA1*0301/DQB1*0302	19.0
	DQA1*0401/DQB1*0402	12.8
	DQA1*0101/DQB1*0501	14.6
	DQA1*0102/DQB1*0602	14.6
	Combined	81.6
DPA1/DPB1	DPA1*0201/DPB1*0101	16.0
	DPA1*0103/DPB1*0201	17.5
	DPA1*01/DPB1*0401	36.2
	DPA1*0301/DPB1*0402	41.6
	DPA1*0201/DPB1*0501	21.7
	DPB1*1401@	7.4
	Combined	94.5

@ No algorithm available for DPB1*1401

- Additional important considerations
  - Extensive overlap between repertoire of different allelic variants
  - Dominant epitopes tend to be “promiscuous”

Is predicting binding to a lot of HLAs necessarily best? A heuristic approach

- Peptide datasets spanning entire proteins associated with measured immune responses in exposed humans

Dataset		No. of Antigens	Total peptides	No. of donors		Reference
Der p/f (House dust mite)		4	156	20		Hinz,2015
Phl p (Timothy grass)		10	425	25		Oseroff, 2010
TB-1		4	71	18		Arlehamn, 2012
TB-2		11	499	32		Arlehamn,2015
Cockroach		6	463	19		Oseroff, 2012
Pertussis		9	785	23		Dillon, 2015
TOTAL		44	2399	137		

Paul S et al. Development and validation of a broad scheme for prediction of HLA class II restricted T cell epitopes. J Immunol Methods. 2015 PubMed PMID: 25862607

Prediction of HLA class II restricted T cell epitopes at the population level

- Optimal results obtained with a set of seven "prototypic" alleles
- These alleles are representatives of a variety of binding modes and supertypes

Paul S et al. J Immunol Methods. 2015 PubMed PMID: 25862607

frontiers in Immunology

ORIGINAL RESEARCH
published: 06 August 2018
doi: 10.3389/fimmu.2018.01795

Determination of a Predictive Cleavage Motif for Eluted Major Histocompatibility Complex Class II Ligands

Sinu Paul[1], Edita Karosiene[1], Sandeep Kumar Dhanda[1], Vanessa Jurtz[2], Lindy Edwards[1], Morten Nielsen[2,3], Alessandro Sette[1,4] and Bjoern Peters[1,4]*

- MHC II Epitope prediction algorithms are based on binding affinity measurements from allele-specific binding assays.
  - They work well for MHC binding predictions
- But MHC II "epitope prediction" performance is relatively low.
- Limitations (compared to class I):
  - MHC molecule structure
  - Longer peptides
  - Binding core & flanking residues
  - Availability of data
  - **Other factors → antigen processing?**

frontiers in Immunology

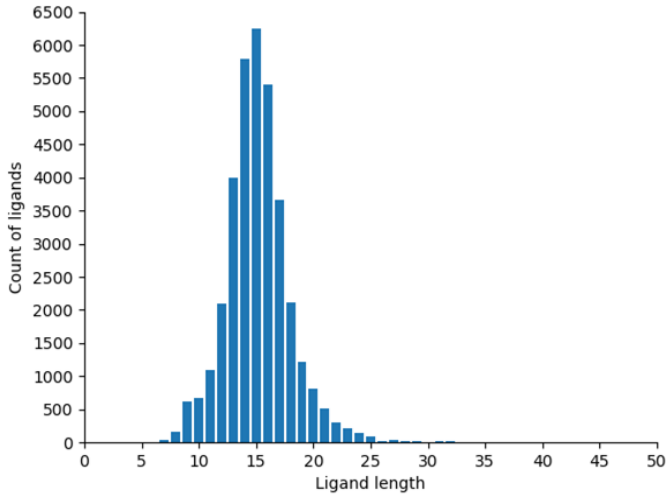
ORIGINAL RESEARCH
published: 06 August 2018
doi: 10.3389/fimmu.2018.01795

Determination of a Predictive Cleavage Motif for Eluted Major Histocompatibility Complex Class II Ligands

Sinu Paul[1], Edita Karosiene[1], Sandeep Kumar Dhanda[1], Vanessa Jurtz[2], Lindy Edwards[1], Morten Nielsen[2,3], Alessandro Sette[1,4] and Bjoern Peters[1,4]*

- MHC II ligand elution data collected from IEDB
- Initial data
  - ~35k peptides
  - lengths 3-46

Distribution of ligands based on length



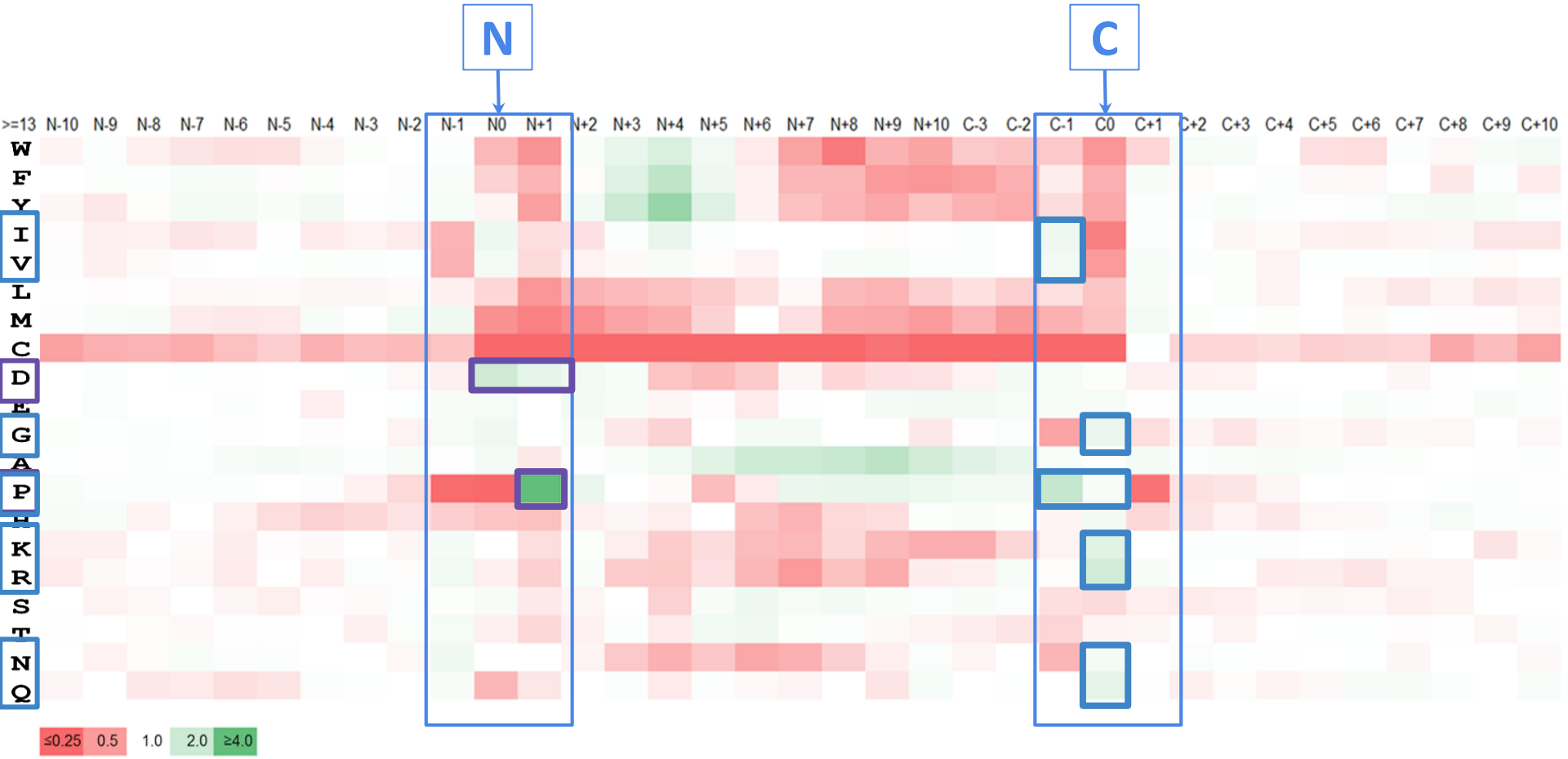
Analyses to filter and generate high quality data

Filtering step	No. of ligand entries	
Initial data	35,367	
Ligands with lengths that are ≥0.5% of total data (lengths 9-23)	34,737	
Ligands with alleles listed unambiguously and having ≥50 entries per allele (17 alleles)	28,007	
Ligands with "optimal lengths" (lengths 13-23)	24,099	*
Ligands with proteins listed unambiguously and excluding "potential false antigens"	18,286	*
Selecting unique peptides & that present in 100% identity in "parent sequences"	14,051	

* Based on probability analysis using predicted binders among ligands vs. random peptides generated from ligands

Cleavage motif

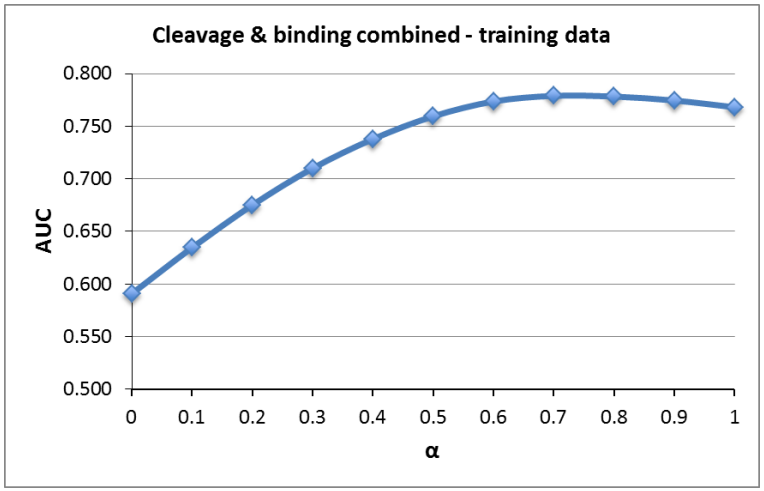
- Enrichment and depletion of amino acids within and adjacent to MHC ligands and predicted binders
- Heatmap – log transformed relative AA frequency with respect to the overall amino acid frequency of the source proteins



Combination of cleavage and binding predictions improved ligand predictions

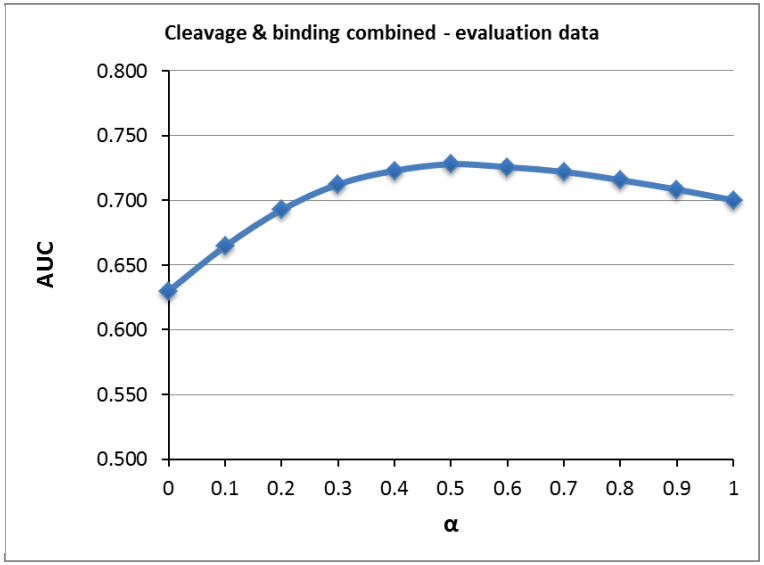
Training ligand data

α	AUC
0	0.591
0.1	0.635
0.2	0.675
0.3	0.710
0.4	0.738
0.5	0.759
0.6	0.774
0.7	0.779
0.8	0.778
0.9	0.774
1	0.768



Evaluation ligand data

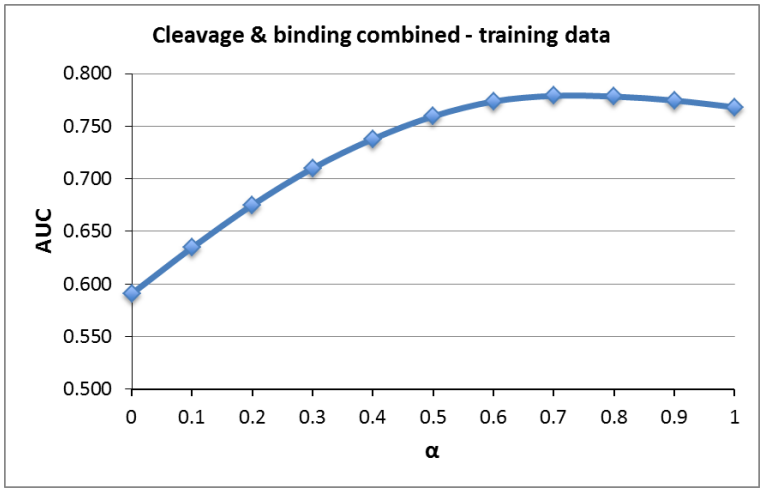
α	AUC
0	0.630
0.1	0.665
0.2	0.693
0.3	0.712
0.4	0.723
0.5	0.728
0.6	0.726
0.7	0.722
0.8	0.716
0.9	0.708
1	0.700



Combination of cleavage and binding does not improve epitope predictions

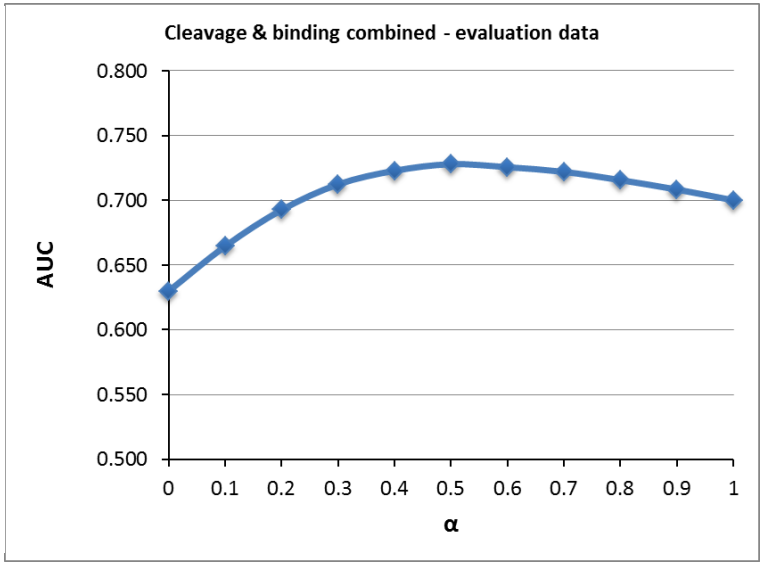
Training ligand data

α	AUC
0	0.591
0.1	0.635
0.2	0.675
0.3	0.710
0.4	0.738
0.5	0.759
0.6	0.774
0.7	0.779
0.8	0.778
0.9	0.774
1	0.768



Evaluation ligand data

α	AUC
0	0.630
0.1	0.665
0.2	0.693
0.3	0.712
0.4	0.723
0.5	0.728
0.6	0.726
0.7	0.722
0.8	0.716
0.9	0.708
1	0.700



- Used 3 different sets of epitopes
  - Epitopes identified from in-house experiments
  - Epitopes identified by tetramer mapping studies (collected from IEDB)
  - Epitopes from five other studies curated by IEDB that contained 15-mer peptides spanning six proteins

Predicting HLA class II T cell epitopes

- Like in the case of class I, processing/cleavage predictions do not improve epitope predictions
- For this reason we considered an agnostic approach, were we used T cell epitope data to directly train predictive algorithms
- Used in-house data and IEDB-derived tetramer as training set

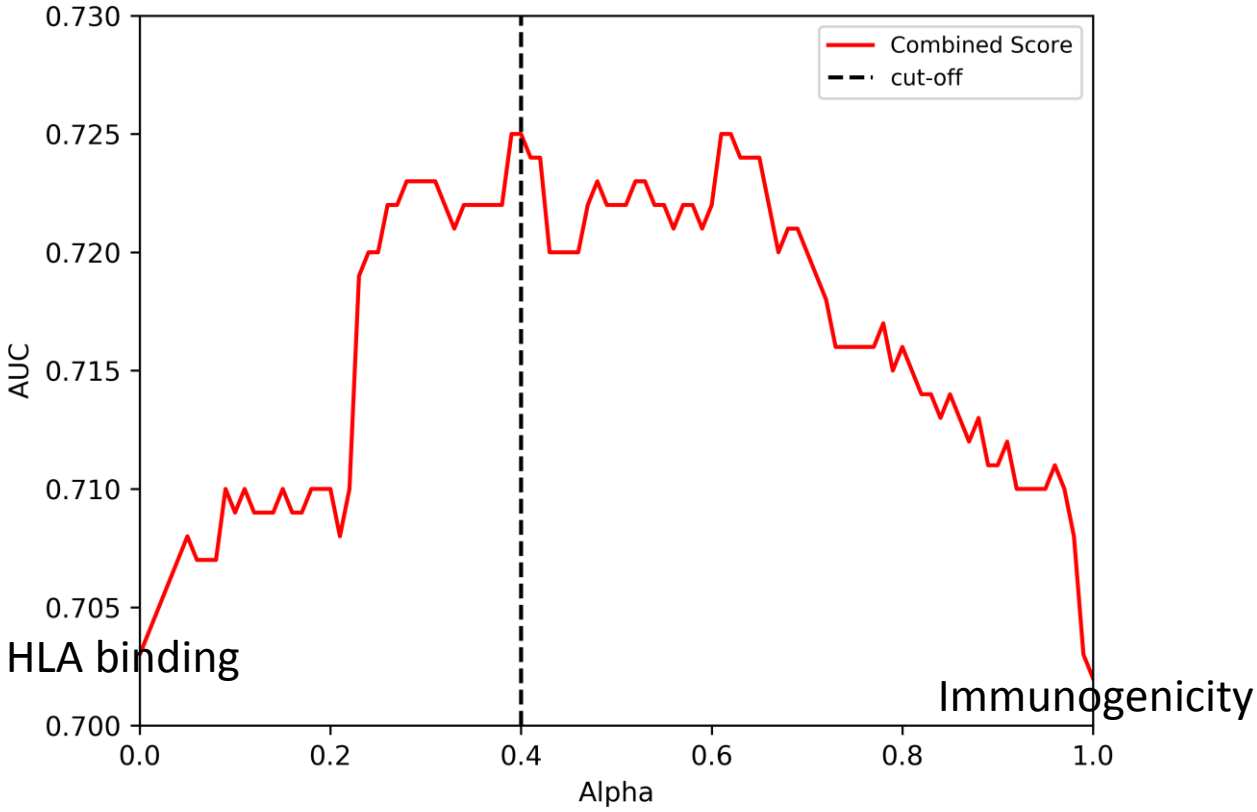
In house training dataset

Antigen (s)	Selection method	# of donors	Reference	# epitopes	# of control peptides
Mycobacterium tuberculosis	Overlapping	18	(Arlehamn et al., 2012)	65	53
	Predicted	28	(Lindestam Arlehamn et al., 2013)		1043
	Overlapping	61	(Lindestam Arlehamn et al., 2016)		362
	Confirmed epitopes	61	(Lindestam Arlehamn et al., 2016)		137
Timothy Grass	Overlapping	25	(Oseroff et al., 2010)	60	360
	Predicted	35	(Schulten et al., 2013)		360
	Overlapping	21	(Westernberg et al., 2016)		6
	Overlapping	37	(Hinz et al., 2015)		0
House Dust Mite	Overlapping	20	(Hinz et al., 2015)	52	6
Cockroach	Overlapping	19	(Dillon et al., 2015)	71	521
Dengue Antigens	Predicted	150	(Weiskopf et al., 2015a)	325	140
Erythropoietin	Overlapping	5	(Tangri et al., 2005)	9	11
CRJ1 and CRJ2	Overlapping	54	(Oseroff et al., 2016)	30	18
Mouse allergens	Predicted	22	(Schulten et al, submitted)	82	885
Novel house dust mite	Predicted	20	(Oseroff et al., 2017)	105	186
Pertussis Vaccine	Overlapping	53	(Bancroft et al., 2016)	100	202
Ragweed allergens	Overlapping	25	(Pham et al., 2016)	15	183
Tetanus		20	(Antunes et al., 2017)	28	98
ZIKV polyprotein	Overlapping	18	(Grifoni et al., unpublished)	48	529
Yellow fever virus	Overlapping	42	(Weiskopf et al, unpublished)	42	639
Overall				**1032**	**5739**

Validation dataset

- Reported in literature from other labs
- Studies measuring T cell reactivity using complete sets of overlapping peptides spanning antigens of interest and exposed patient populations
- After excluding antigens included in the in-house datasets, 57 papers were selected
- Final set contained 530 dominant epitopes and 1758 non-epitopes

Gain in performance by combining binding and immunogenicity predictions



Conclusions (IV)

- HLA class II predictions are less accurate than class I
- However, the extensive repertoire overlap and the phenomenon of epitope promiscuity are also a prominent factor
- As in the case of class I, processing predictions do not improve epitope predictions
- Population approaches and training with T cell epitope data most promising approaches

Acknowledgements

LJI Group

Bjoern Peters

Kerrie Vaughan

Sinu Paul

Zeynep Kosaloglu-Yalcin

Manasa Lanka

John Sidney

Jason Greenbaum

Manasa Lanka

Ashmitaa Premlal

Luise Sternberg

Angela Frentzen

Sandeep Dhanda

NCI

Paul Robbins

Steven Rosenberg

Jared Gartner

DTU-Bioinformatics

Morten Nielsen

Carolina Barra

Esteban Lanzaroti

Massimo Andreatta

Vanessa Jurtz

Kamilla Kjærgård Jensen

Other Collaborators

Søren Buus, IMMI, University of Copenhagen: MHC binding/Tetramers

Nicola Ternette, Nuffield Department of Medicine, Oxford

Sine Reker Hadrup, DTU Nano

Etienne Caron, ETH Zurich

Josh Elias, Stanford