

Supplementary Information

Deciphering HLA-I motifs across HLA peptidomes improves neo-antigen predictions and identifies allostery regulating HLA specificity

Michal Bassani-Sternberg^{1,2,*}, Chloé Chong^{1,2}, Philippe Guillaume^{1,2}, Marthe Solleder^{1,3}, HuiSong Pak^{1,2}, Philippe O Gannon², Lana E Kandalaf^{1,2}, George Coukos^{1,2}, David Gfeller^{1,2,3,*}

¹Ludwig Centre for Cancer Research, University of Lausanne, 1066 Epalinges, Switzerland.

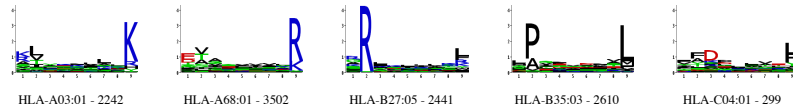
²Department of Fundamental Oncology, University Hospital of Lausanne, Lausanne, Switzerland.

³Swiss Institute of Bioinformatics (SIB), 1015 Lausanne, Switzerland.

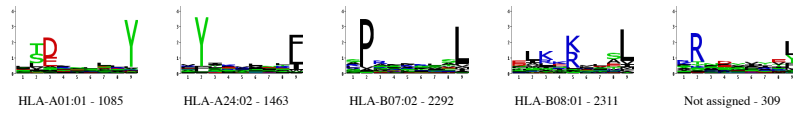
*To whom correspondence should be sent: David.Gfeller@unil.ch and Michal.Bassani@chuv.ch.

Supplementary Figures

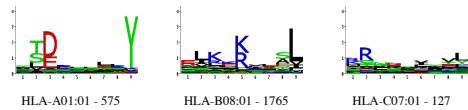
Mel_15



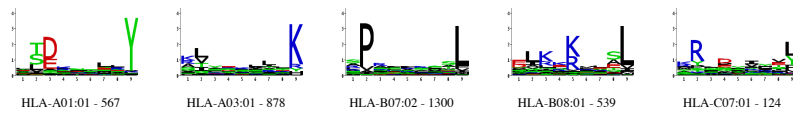
Mel_16



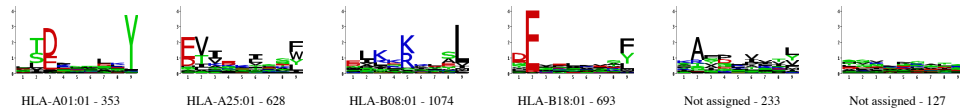
Mel_12



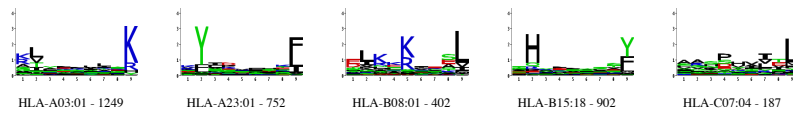
Mel_8



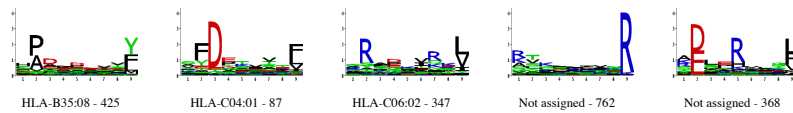
Mel_5



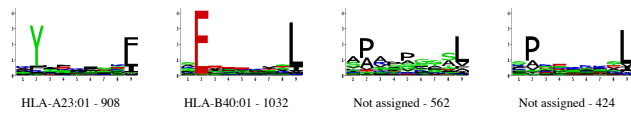
Fibroblast



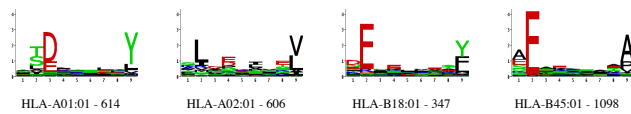
HCC1143



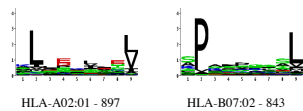
HCC1937



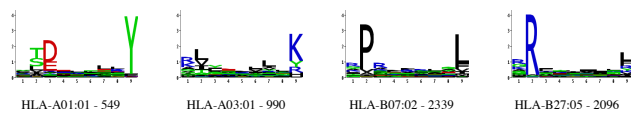
HCT116



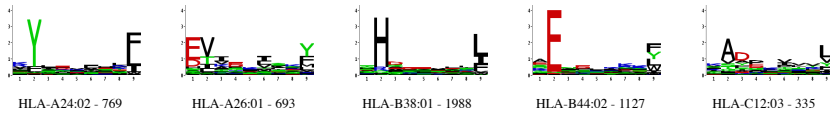
JY



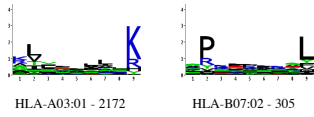
Bcell



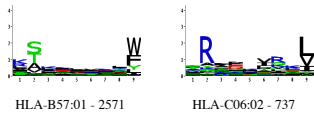
MAVER_1



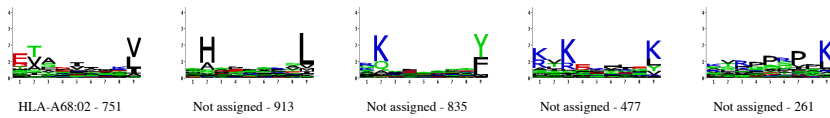
HEK293



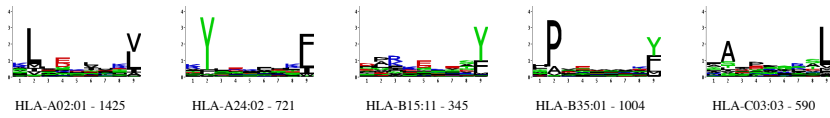
HL_60



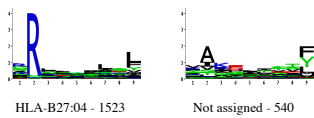
RPMI8226



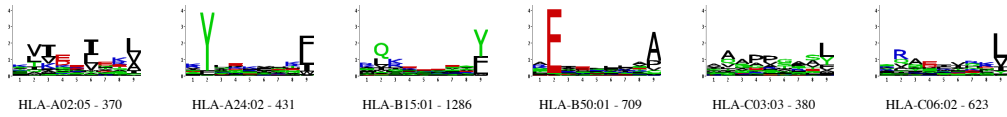
THP_1



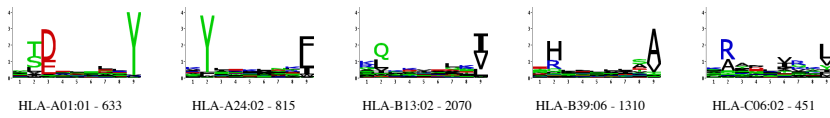
CA46



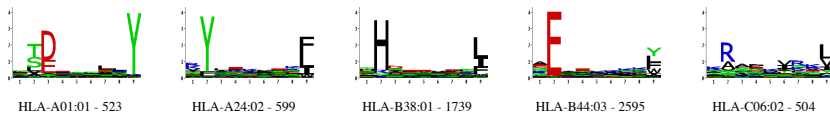
CD165



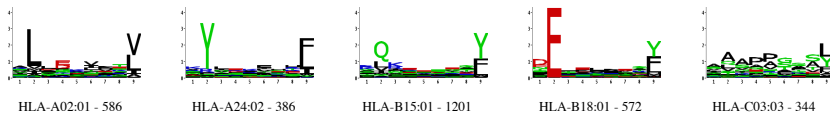
CM467



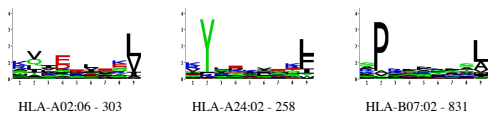
GD149



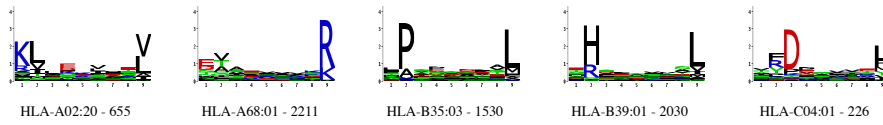
MD155



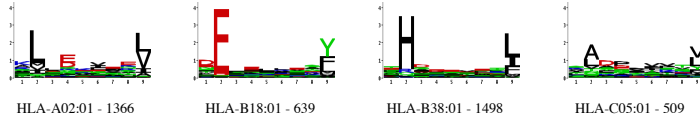
PD42



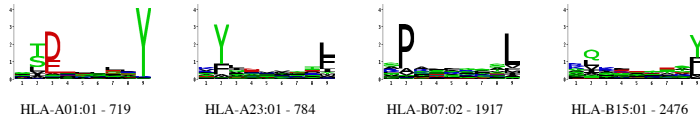
RA957



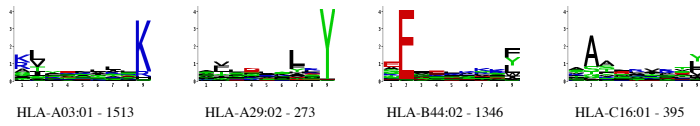
TIL1



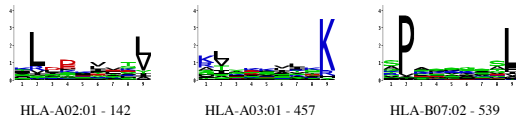
TIL3



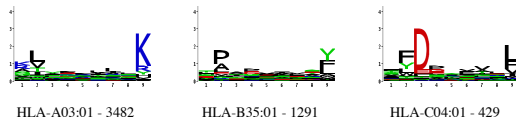
Apher1



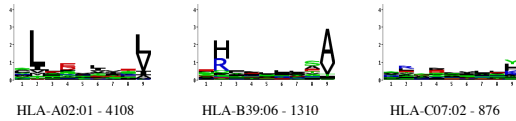
Apher6



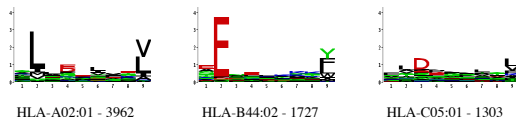
LNT-229-all



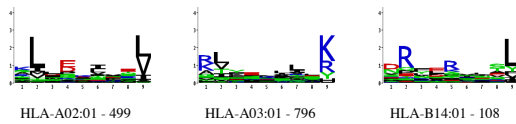
T98G-all



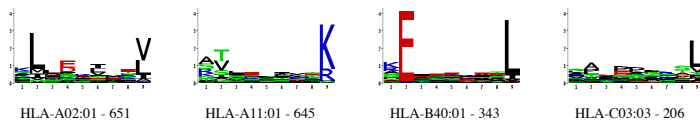
U-87-all



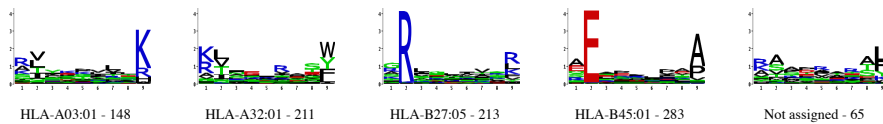
Mel-624



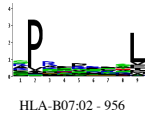
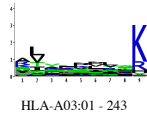
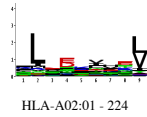
SK-Mel-5



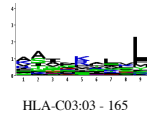
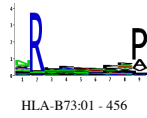
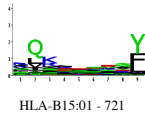
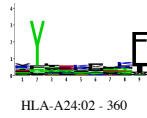
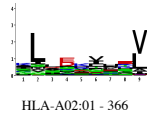
pat_AC2



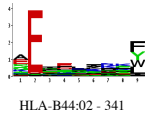
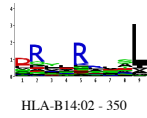
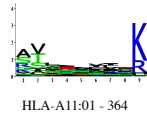
pat_C



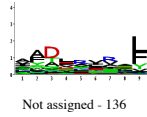
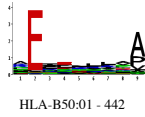
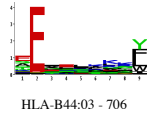
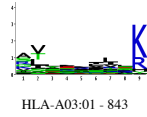
pat_CELG



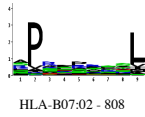
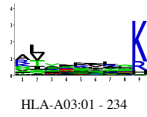
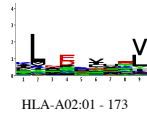
pat_CP2



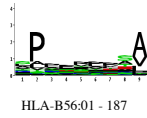
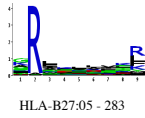
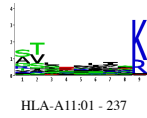
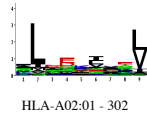
pat_FL



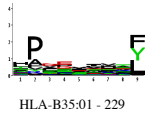
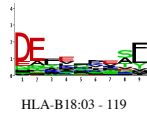
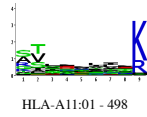
pat_J



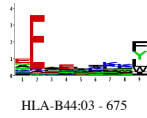
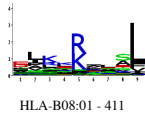
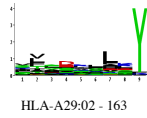
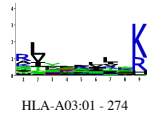
pat_JPB3



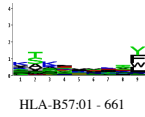
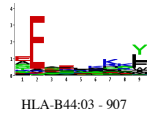
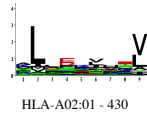
pat_JT2



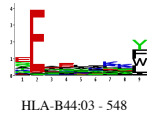
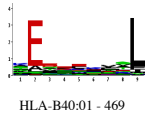
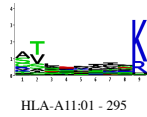
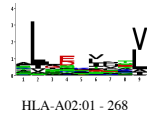
pat_M



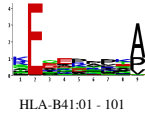
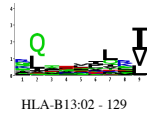
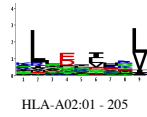
pat_MA



pat_ML



pat_NS2



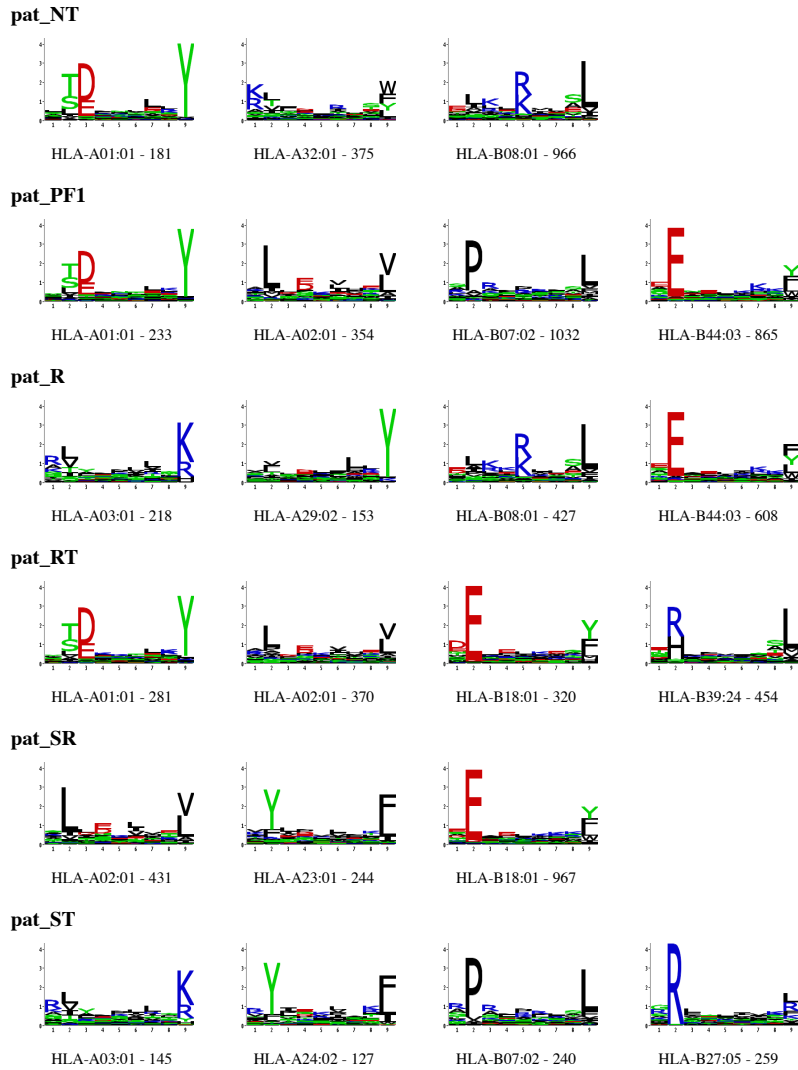


Fig A: Motifs identified across all HLA peptidomics datasets analyzed in this work. The allele to which the logos have been mapped by the approach presented in this work and the number of peptides assigned to this allele is indicated below each motif. The few motifs that could not be assigned to an allele are annotated as “Not assigned”.

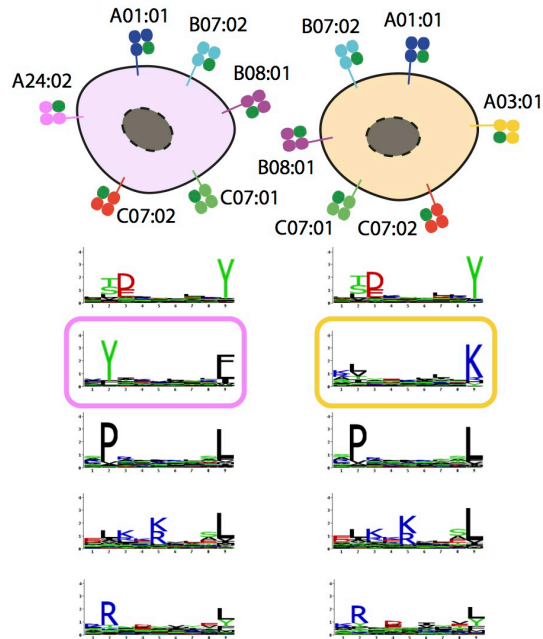


Fig B: Strategy to annotate motifs in cases where two samples share all but one motif. The left sample (Mel_16) and the right sample (Mel_8) come from two different melanoma patients from [1].

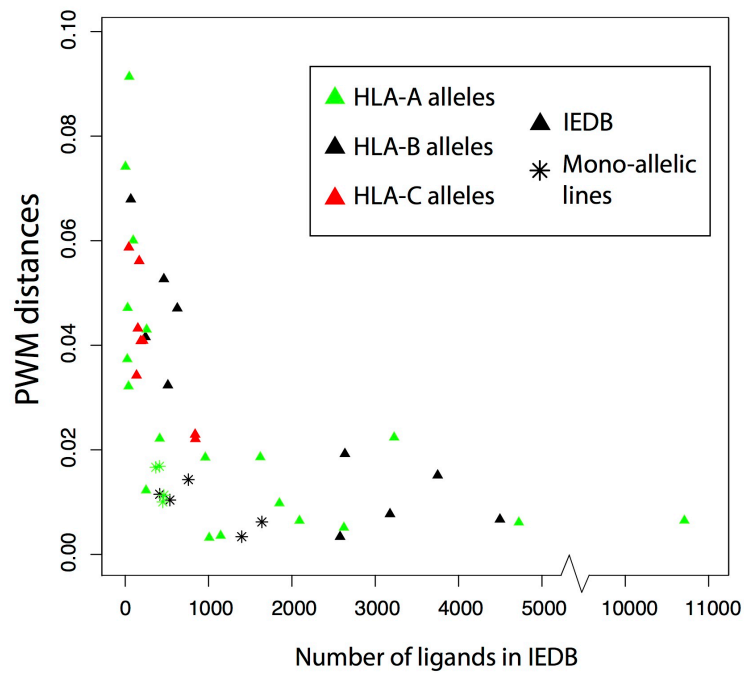


Fig C: PWM distances between the motif predicted in this work and motifs derived from IEDB, as a function of the number of ligands in IEDB.

HLA alleles	predicted motifs	MS IEDB-derived motifs	non-MS IEDB-derived motifs
A01:01*	2076 peptides	1591 peptides	845 peptides
A02:01	6886 peptides	5018 peptides	4808 peptides
A02:05	370 peptides	209 peptides	28 peptides
A02:06	303 peptides	no ligand	1388 peptides
A02:20	655 peptides	no ligand	no ligand
A03:01*	6581 peptides	955 peptides	2098 peptides
A11:01*	1072 peptides	335 peptides	2187 peptides
A23:01	1800 peptides	TFLPFIHTI	510 peptides

HLA alleles	predicted motifs	MS IEDB-derived motifs	non-MS IEDB-derived motifs
A24:02*	2344 peptides	1939 peptides	1606 peptides
A25:01	628 peptides	no ligand	65 peptides
A26:01	693 peptides	no ligand	606 peptides
A29:02*	396 peptides	3888 peptides	572 peptides
A32:01	441 peptides	2 peptides	462 peptides
A68:01	4293 peptides	12 peptides	1110 peptides
A68:02	751 peptides	24 peptides	1199 peptides

HLA alleles	predicted motifs	MS IEDB-derived motifs	non-MS IEDB-derived motifs
B07:02*	5808 peptides	3110 peptides	1403 peptides
B08:01	3460 peptides	633 peptides	975 peptides
B13:02	2106 peptides	no ligand	no ligand
B14:01	108 peptides	no ligand	no ligand
B14:02	350 peptides	17 peptides	70 peptides
B15:01*	3642 peptides	1012 peptides	2105 peptides
B15:11	345 peptides	no ligand	no ligand
B15:18	902 peptides	no ligand	no ligand
B18:01	2048 peptides	88 peptides	317 peptides
B18:03	119 peptides	no ligand	no ligand
B27:04	1523 peptides	734 peptides	23 peptides
B27:05*	3578 peptides	2083 peptides	520 peptides
B35:01*	2126 peptides	445 peptides	1220 peptides
B35:03	3106 peptides	102 peptides	HPNIEEVAL

HLA alleles	predicted motifs	MS IEDB-derived motifs	non-MS IEDB-derived motifs
B35:08	425 peptides	101 peptides	IPSIDVHHY
B38:01	2987 peptides	4 peptides	21 peptides
B39:01	2030 peptides	3 peptides	327 peptides
B39:06	2080 peptides	21 peptides	3 peptides
B39:24	454 peptides	no ligand	no ligand
B40:01*	1441 peptides	277 peptides	638 peptides
B41:01	101 peptides	40 peptides	no ligand
B44:02*	3066 peptides	1756 peptides	319 peptides
B44:03*	3367 peptides	885 peptides	253 peptides
B45:01	1250 peptides	94 peptides	123 peptides
B50:01	918 peptides	250 peptides	no ligand
B56:01	187 peptides	3 peptides	no ligand
B57:01*	2947 peptides	575 peptides	427 peptides
B73:01	456 peptides	no ligand	76 peptides

HLA alleles	predicted motifs	MS IEDB-derived motifs	non-MS IEDB-derived motifs	HLA alleles	predicted motifs	MS IEDB-derived motifs	non-MS IEDB-derived motifs
C03:03	1333 peptides 	3 peptides VYSPYEF ATAKALAI	125 peptides 	C07:02	876 peptides 	46 peptides 	120 peptides
C04:01*	805 peptides 	472 peptides ED	325 peptides 	C07:04	187 peptides 	no ligand	no ligand
C05:01	1722 peptides 	729 peptides 	84 peptides 	C12:03	335 peptides 	6 peptides 	143 peptides
C06:02	1768 peptides 	8 peptides RCEVRF	124 peptides 	C16:01	392 peptides 	214 peptides 	no ligand
C07:01	420 peptides 	10 peptides SXVTH	34 peptides 				

Fig D: Comparison between the HLA-I motifs identified by integrating data across all HLA peptidomics studies and the motifs from IEDB (either MS and non-MS data). When only one peptide was present the sequence of the peptide is indicated. Stars indicate motifs considered in the analyses of Fig J.

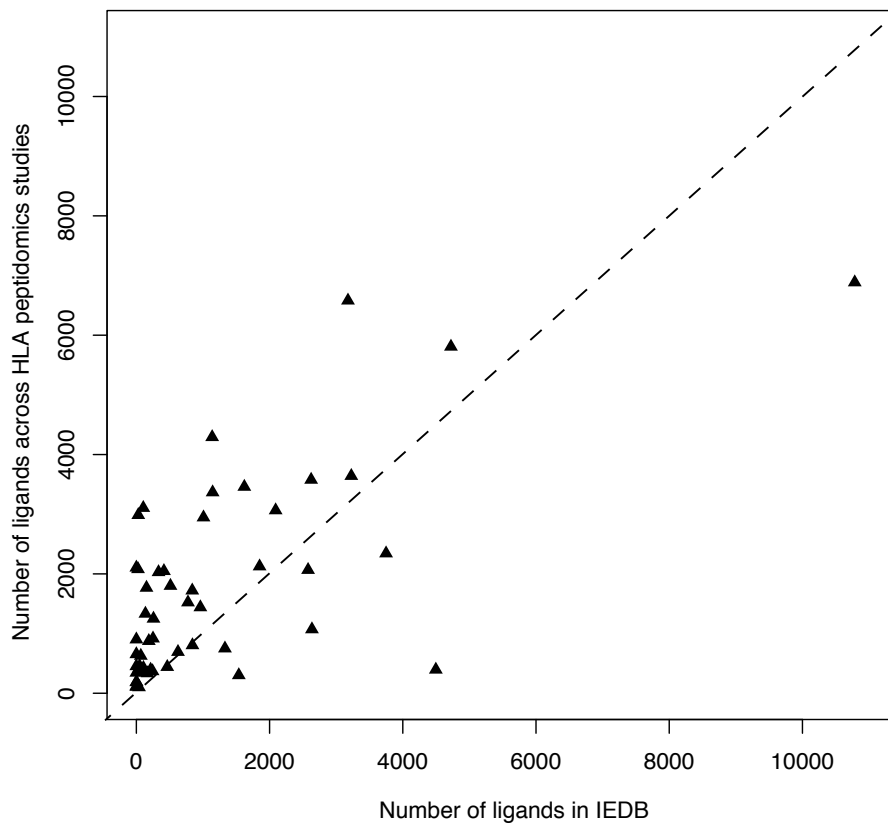


Fig E: Number of peptides assigned to each allele across the HLA peptidomics studies considered in this work, compared to IEDB data (see all raw numbers in Fig D).

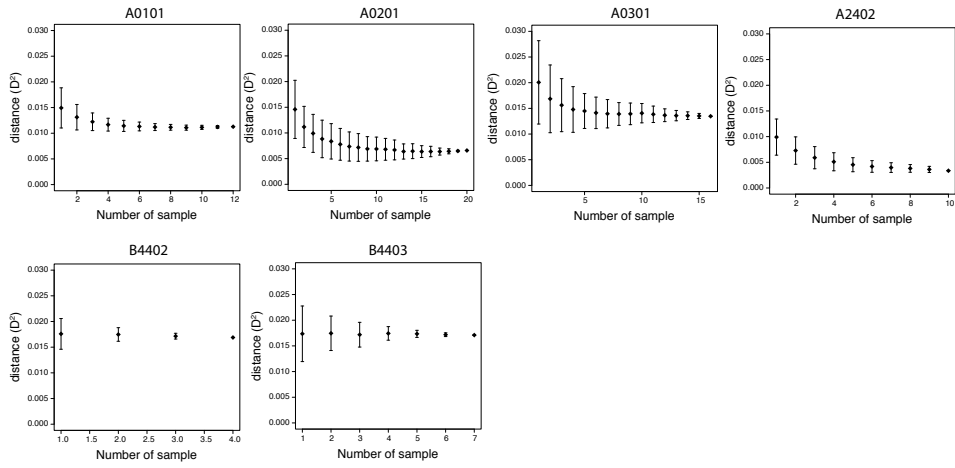
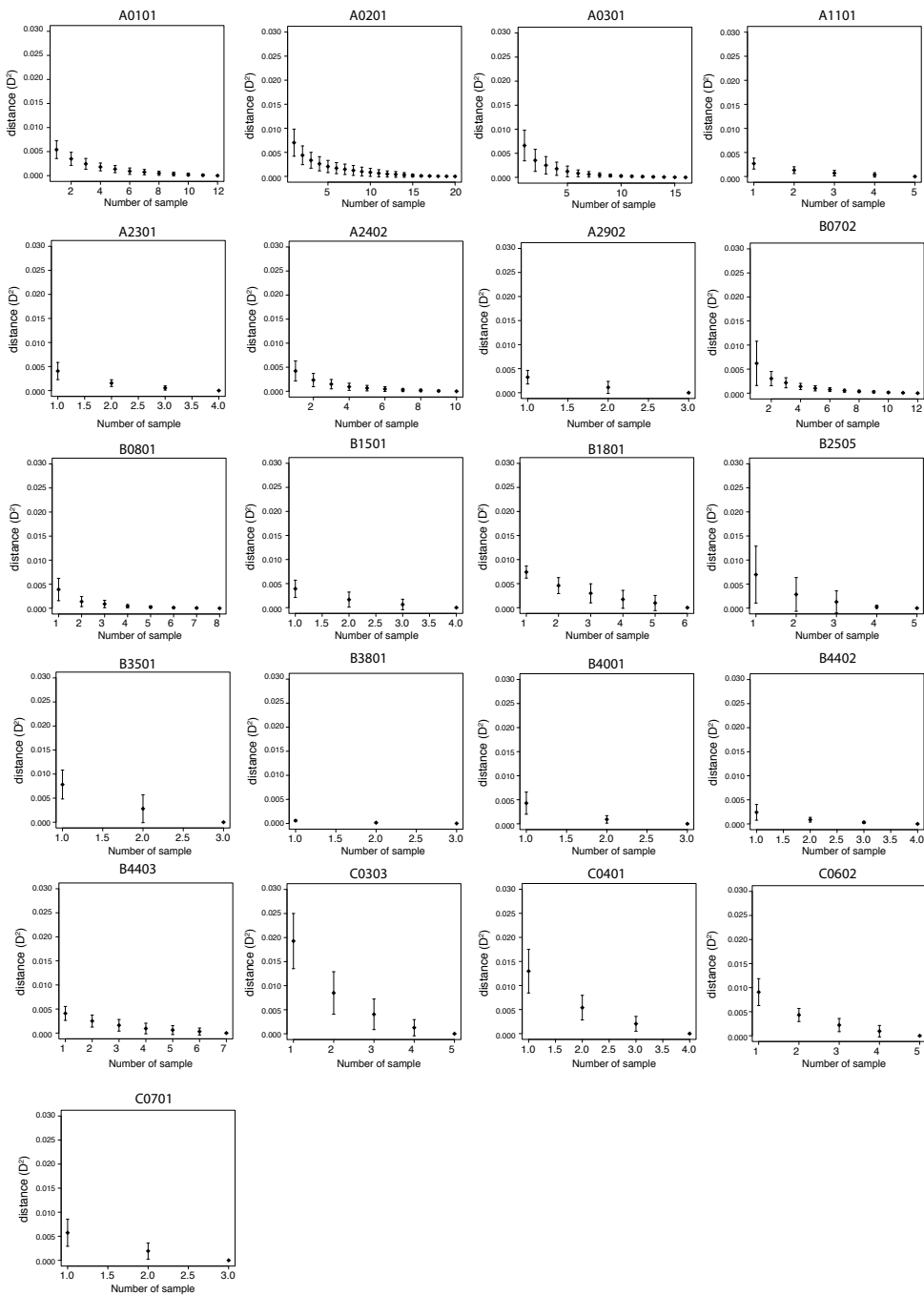
A**B**

Fig F: (a) Distance between the motifs predicted in our work and those derived from mono-allelic cell lines [2], as a function of the number of samples that are considered (error bars correspond to 100 random choices). The six alleles were selected as those with motifs detected in more than 3 samples in our dataset of pooled HLA peptidomics data and present in the mono-allelic cell lines of [2]. Of note, some small contaminations or batch effects are expected even in mono-allelic cell lines and we did not expect distances to fall exactly to zero. (b) Same as in A, but comparing with the motifs derived from all samples, for all alleles present in at least three samples. By construction all distances fall to 0. The most important message of this figure is that all distances (even when considering only one sample) are very small, and much smaller than the 0.078 threshold used to determine two motifs as ‘similar’.

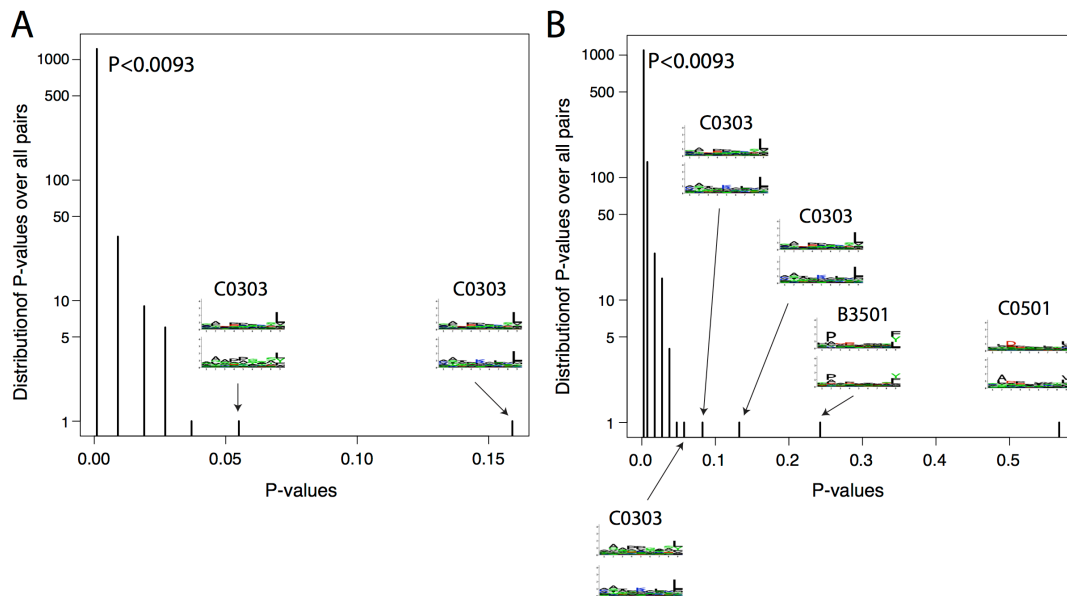


Fig G: Distribution of P-values of similarities over all pairs of motifs annotated to the same alleles. Empirical P-values were computed by comparing for each pair of motif (m_i, m_j) the similarity between m_i and the set of known HLA-I motifs, excluding the one to which m_i had been annotated (107 motifs in total, see Methods). In (a), the Euclidean distance was used and in (b) the BLiC score of ref. [3]. The first bar of the graphs shows the number of cases where the similarity was higher than to any of the known HLA-I motifs used for estimating these empirical P-values ($P < 1/107 = 0.0093$). The very few pairs of motifs that do not pass the statistical threshold are shown explicitly, as well as the allele to which they had been assigned.

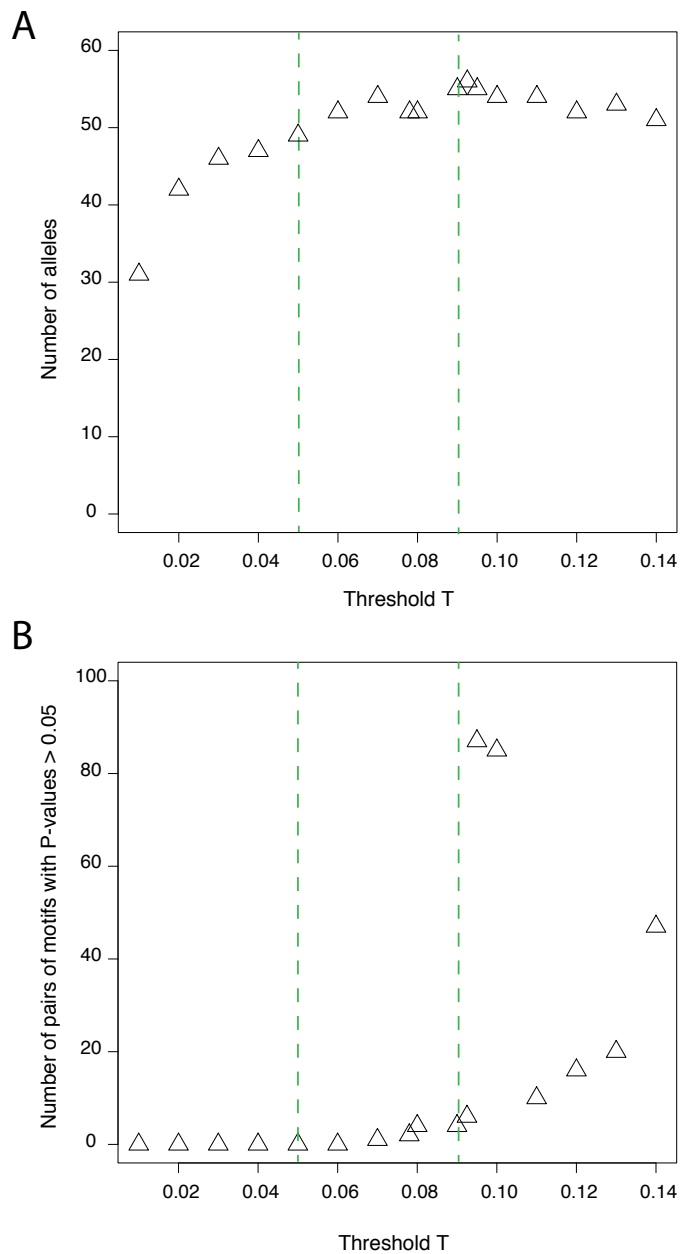


Fig H: Effect of different thresholds T for the Euclidean distance. (a) Number of HLA-I alleles with at least one motif annotated to them as a function of T. (b) Number of pairs of motifs annotated to the same allele with similarity P-values (based on Euclidean distance) larger than 0.05, as a function of T. The green bars indicate a possible range of values of T that leads to similar number of alleles being detected and few pairs of motifs with P-values > 0.05. The non-linearity can be expected since our algorithm is an optimization procedure.

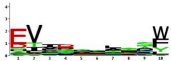
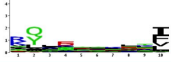
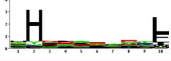
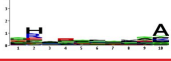
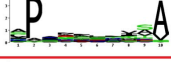
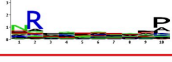
HLA alleles	Predicted logo	Known ligands
HLA-A25:01		<i>no ligands</i>
HLA-B13:02		GQFKDIITKV YILGADPLRV
HLA-B38:01		<i>no ligands</i>
HLA-B39:06		<i>no ligands</i>
HLA-B56:01		<i>no ligands</i>
HLA-B73:01		<i>no ligands</i>

Fig I: New HLA-I binding motifs identified for 10-mers across all HLA peptidomics datasets.

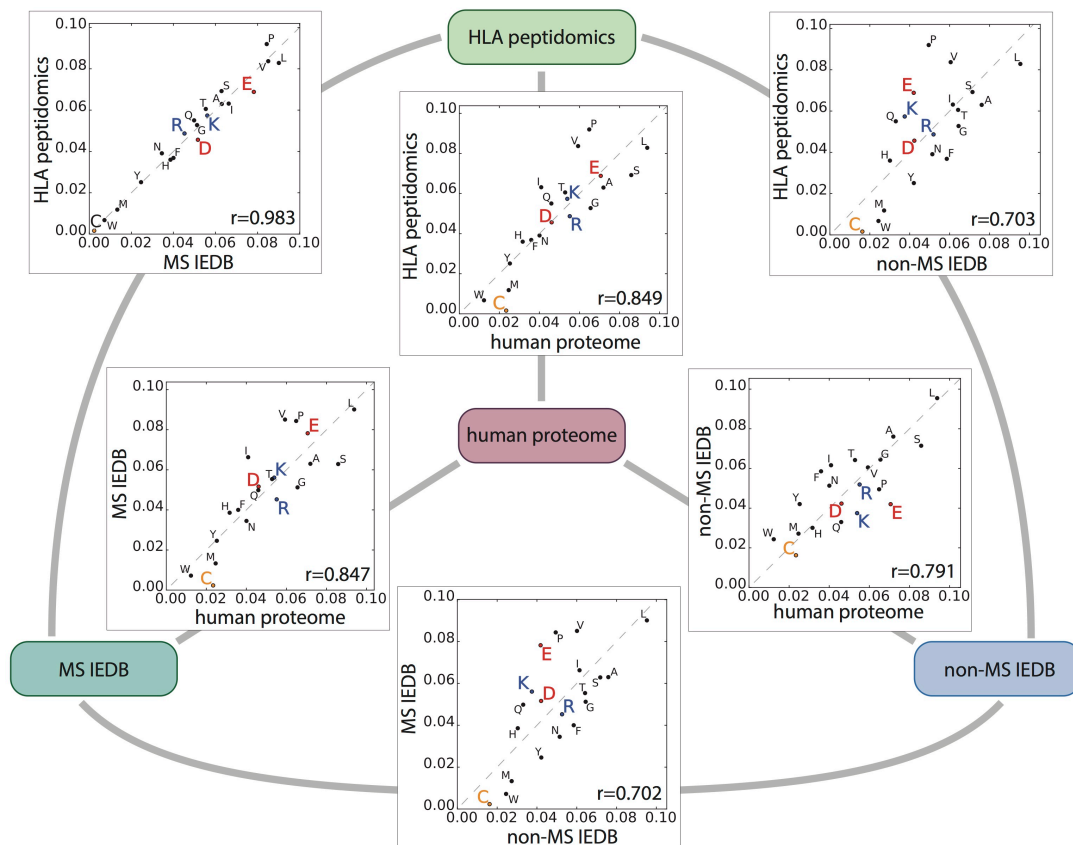


Fig J: Comparison between amino acid frequencies at positions P4 to P7 in HLA peptidomics data from unmodified cell lines or tissue samples, MS IEDB data, non-MS IEDB data, and the human proteome. Only alleles with at least 100 ligands in our dataset, IEDB non-MS and IEDB MS were used (marked with * in Fig D).

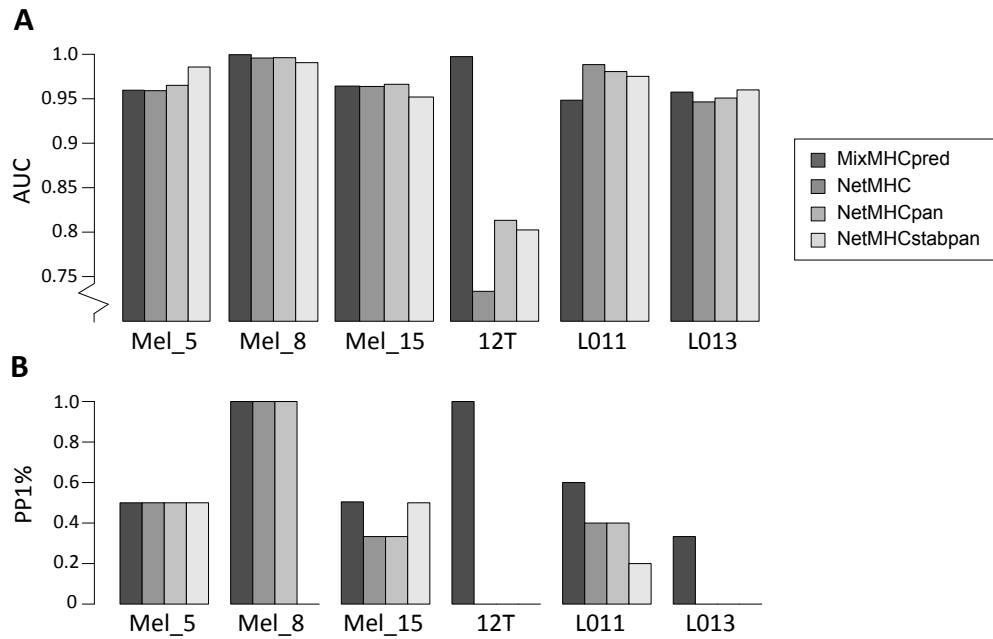


Fig K: Comparison between our predictor (MixMHCpred) and existing tools when excluding endogenous HLA-I ligands from the sample in which mutated neo-antigens had been identified (same comparison as in Fig 4B-C, other benchmarkings remain the same).

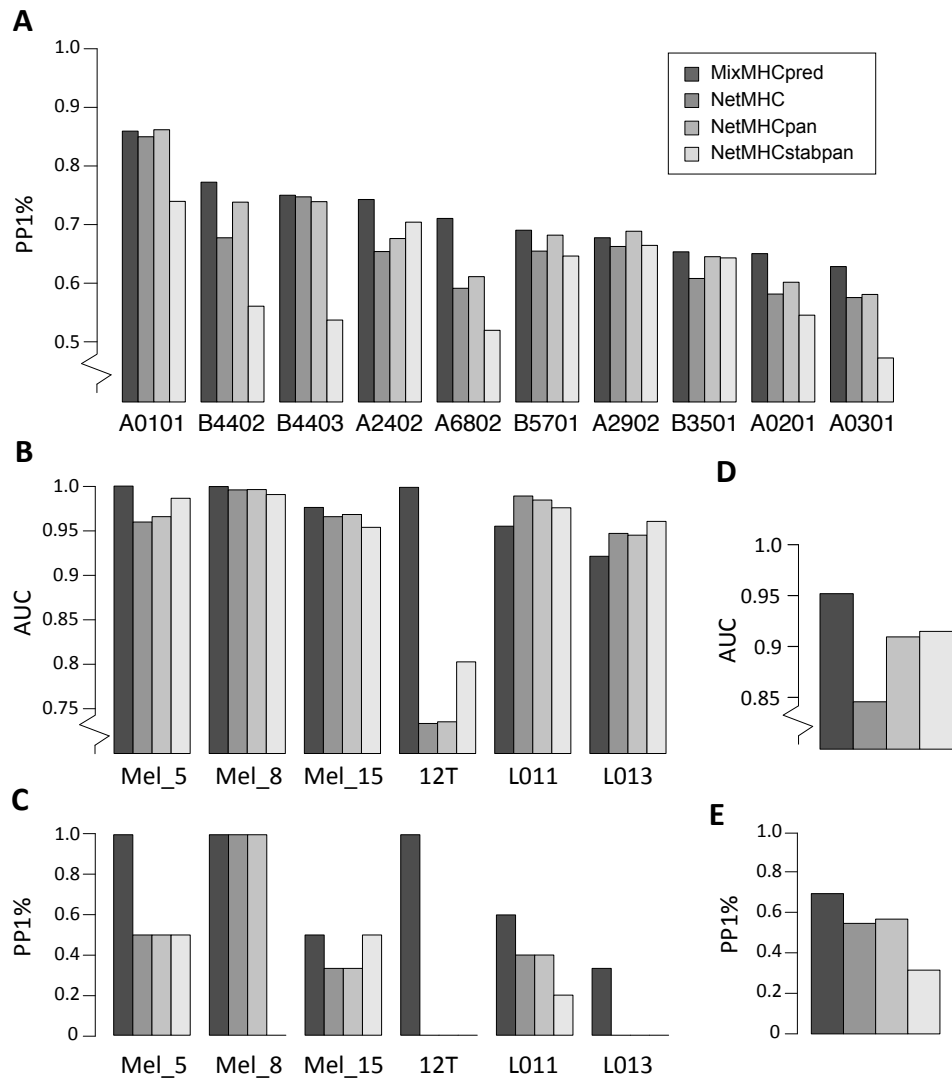


Fig L: Comparison between our predictor (MixMHCpred) and existing tools when adding 5% of noise in the pooled HLA peptidomics data used to train our predictor.

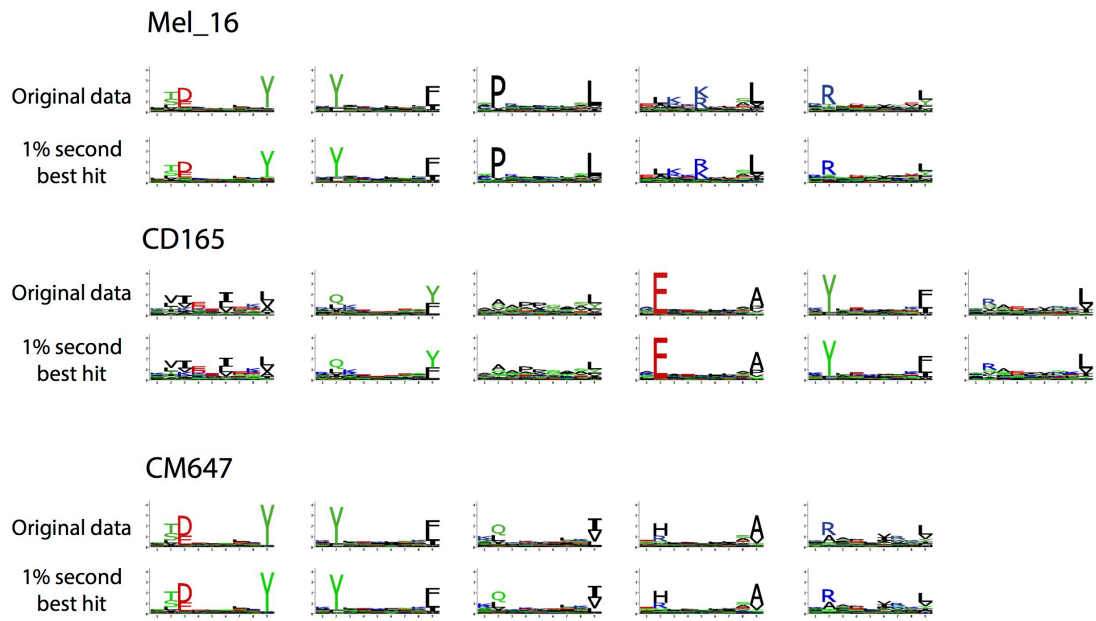


Fig M: Comparison of the deconvoluted motifs for the three samples shown in Fig 1, when using the second best hit in spectral searches for 1% of the peptides.

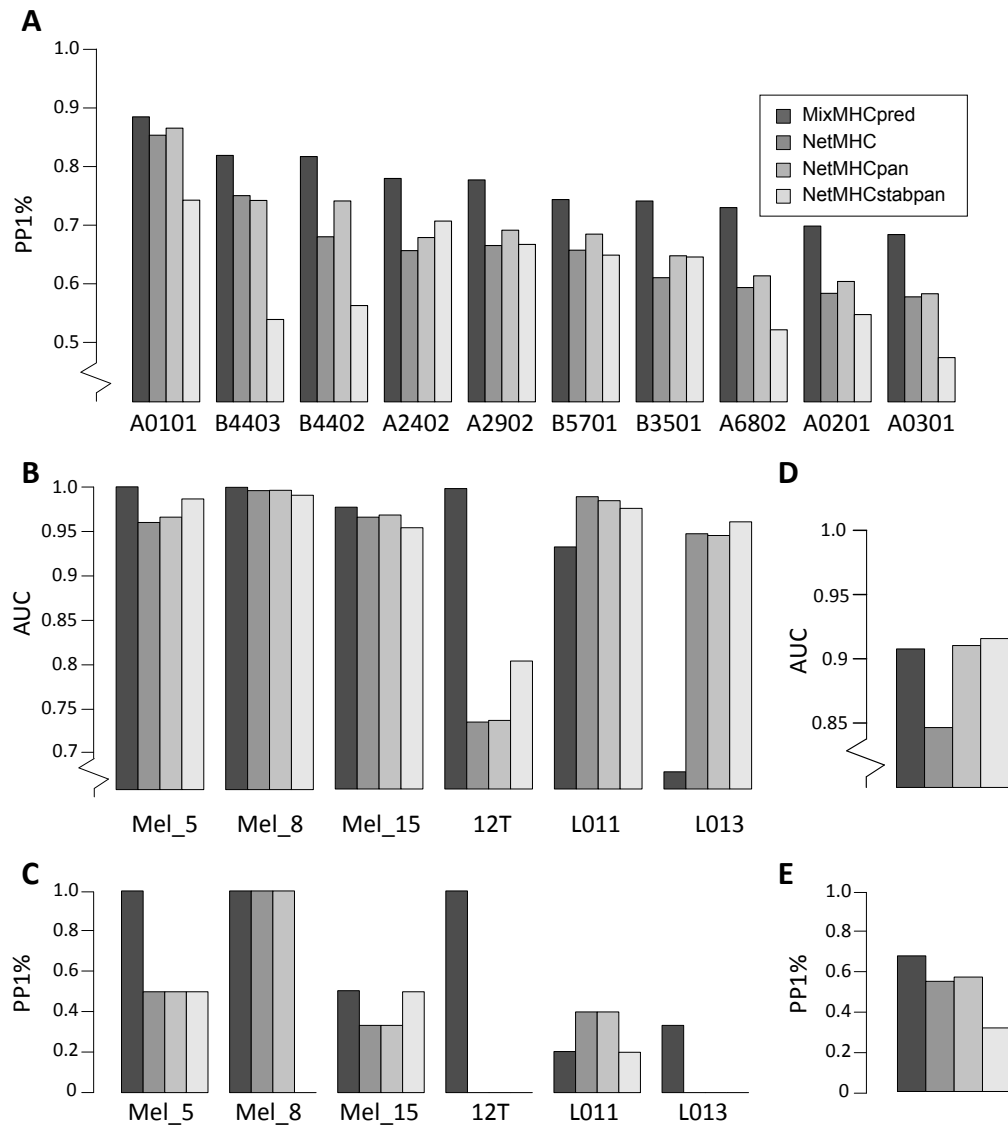
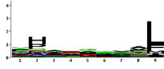
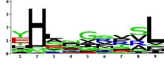
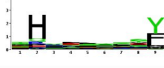
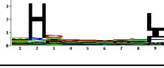
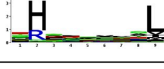

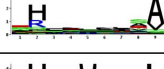



Fig N: Performance of predictors when using amino acid frequencies from the human proteome (instead of those found at non-anchor residues) in MixMHCpred (same data as in Fig 4). In this case, cysteine containing peptides are poorly predicted, which increases the performance of MixMHCpred on elution data (mainly panel A), but decreases the performance on other datasets (especially L011 and L013).

HLA alleles	Binding motif	Amino acid at position 90	HLA alleles	Binding motif	Amino acid at position 90
HLA-A02:20		asparagine	HLA-B27:10		isoleucine
HLA-A03:01		asparagine	HLA-B27:20		isoleucine
HLA-A03:19		asparagine	HLA-B40:13		isoleucine
HLA-A30:01		asparagine	HLA-B41:01		isoleucine
HLA-A30:04		asparagine	HLA-B41:02		isoleucine
HLA-A31:01		asparagine	HLA-B42:01		isoleucine
HLA-A32:01		asparagine	HLA-B47:01		isoleucine
HLA-A32:07		asparagine	HLA-B48:01		isoleucine
HLA-B13:01		isoleucine	HLA-B57:01		asparagine
HLA-B13:02		isoleucine	HLA-B57:02		asparagine
HLA-B15:01		isoleucine	HLA-B57:03	X-ray structure with KAFSPEVI	asparagine
HLA-B15:42		isoleucine	HLA-C15:02		asparagine

Fig O: Binding motifs and amino acid at position 66 for HLA-I alleles showing some preference for charged amino acids (R or K) at P1.

A

HLA alleles	Binding motif
HLA-B15:09	
HLA-B15:10	
HLA-B15:18	
HLA-B38:01	
HLA-B39:01	
HLA-B39:05	
HLA-B39:06	
HLA-B39:09	

B

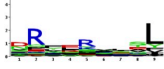
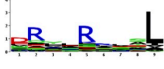
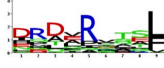
HLA alleles	Binding motif
HLA-B14:01	
HLA-B14:02	
HLA-B14:03	

Fig P: (a) Binding motifs for HLA-I alleles showing preference for histidine at P2. (b) Binding motifs for HLA-B14 alleles displaying the same P2 binding site as HLA-B15:18.

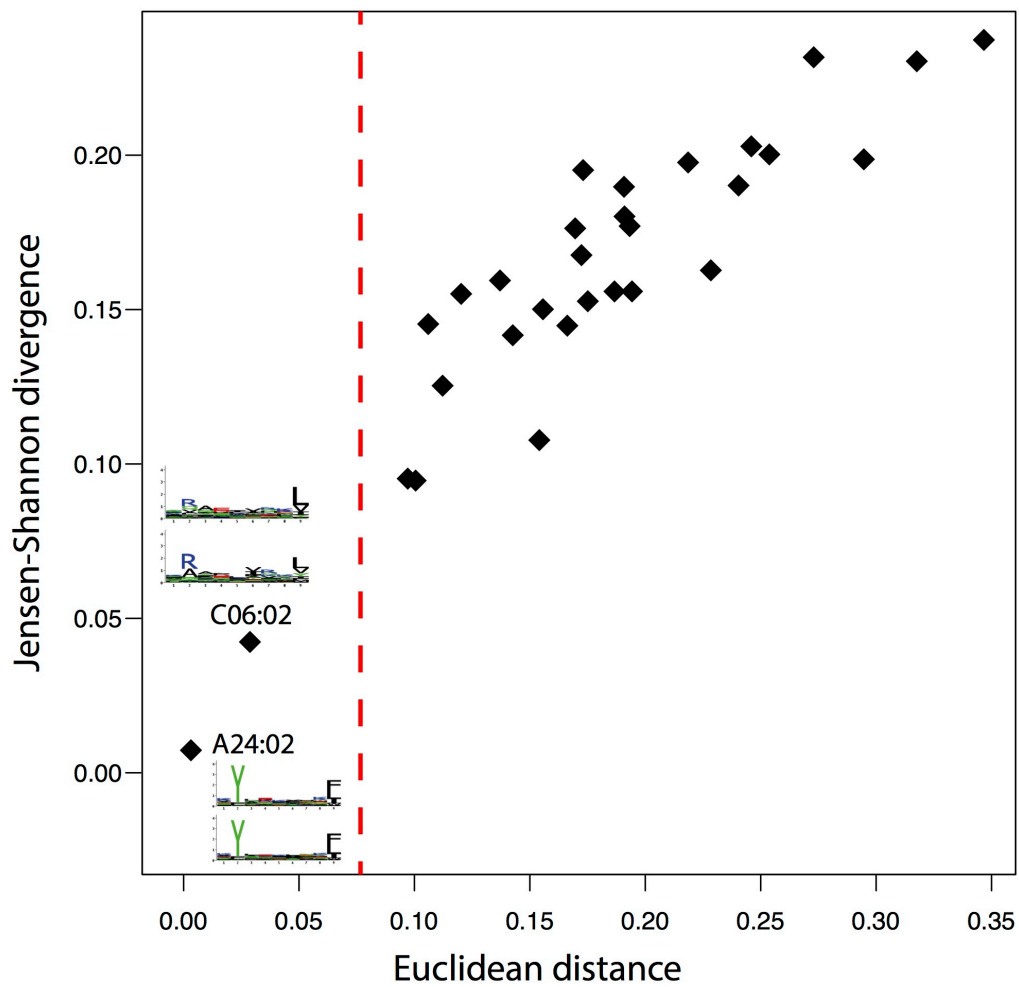


Fig Q: Comparison of the motif similarity values using Euclidean distance and Jensen-Shannon divergence (average of Jensen-Shannon divergence over all 9 positions) for all 5x6 pairs of motifs between two last samples (CD165 and CM467) shown in Fig 1. The red line shows the cut-off on Euclidean distance used in this work.

Table A: HLA peptidomics datasets analyzed in this work. Columns 2 and 3 show the number of unique peptides in each dataset. The first ten samples have been generated for this work.

Samples	9-mers	10-mers	HLA-A	HLA-A	HLA-B	HLA-B	HLA-C	HLA-C
CD165	3799	847	A02:05	A24:02	B15:01	B50:01	C03:03	C06:02
CM467	5279	921	A01:01	A24:02	B13:02	B39:06	C06:02	C12:03
GD149	5960	2063	A01:01	A24:02	B38:01	B44:03	C06:02	C12:03
MD155	3089	493	A02:01	A24:02	B15:01	B18:01	C03:03	C07:01
PD42	1392	450	A02:06	A24:02	B07:02	B55:01	C01:02	C07:02
RA957	6652	1944	A02:20	A68:01	B35:03	B39:01	C04:01	C07:02
TIL1	4012	711	A02:01	B18:01	B38:01		C05:01	
TIL3	5896	1290	A01:01	A23:01	B07:02	B15:01	C12:03	C14:02
Apher1	3527	1273	A03:01	A29:02	B44:02	B44:03	C12:03	C16:01
Apher6	1138	376	A02:01	A03:01	B07:02		C07:02	
Mel15 [1]	11094	4865	A03:01	A68:01	B27:05	B35:03	C02:02	C04:01
Mel16 [1]	7460	1769	A01:01	A24:02	B07:02	B08:01	C07:01	C07:02
Mel12 [1]	2467	417	A01:01		B08:01		C07:01	
Mel8 [1]	3408	1283	A01:01	A03:01	B07:02	B08:01	C07:01	C07:02
Mel5 [1]	3108	750	A01:01	A25:01	B08:01	B18:01		
Fibroblast [4]	3492	859	A03:01	A23:01	B08:01	B15:18	C07:02	C07:04
HCC1143 [4]	1989	284	A31:01	B35:08	B37:01	C04:01	C06:02	
HCC1937 [4]	2926	1065	A23:01	A24:02	B07:02	B40:01	C03:04	C07:02
HCT116 [4]	2665	737	A01:01	A02:01	B45:01	B18:01	C05:01	C07:01
JY [4]	1740	483	A02:01		B07:02		C07:02	
Bcell [5]	5974	2521	A01:01	A03:01	B07:02	B27:05	C02:02	C07:01
Mel-624 [6]	1403	425	A02:01	A03:01	B07:02	B14:01	C07:02	C08:01
SK-Mel-5 [6]	1845	618	A02:01	A11:01	B40:01		C03:03	
HEK293 [7]	2477	1221	A03:01		B07:02		C07:02	
MAVER_1 [7]	4912	1423	A24:02	A26:01	B38:01	B44:02	C05:01	C12:03
HL_60 [7]	3308	1437	A01:01		B57:01		C06:02	
RPMI8226 [7]	3237	647	A30:01	A68:02	B15:03	B15:10	C02:10	C03:04
THP_1 [7]	4085	656	A02:01	A24:02	B15:11	B35:01	C03:03	
CA46 [7]	2063	414	A26:03		B27:04		C12:02	
LNT-229 [8]	5202	2191	A03:01		B35:01		C04:01	
T98G [8]	6294	1413	A02:01		B39:06		C07:02	
U-87 [8]	6992	2126	A02:01		B44:02		C05:01	
pat_AC2 [9]	920	267	A03:01	A32:01	B27:05	B45:01		
pat_C [9]	1423	641	A02:01	A03:01	B07:02		C07:02	

pat_CELG [9]	2068	664	A02:01	A24:02	B15:01	B73:01	C03:03	C15:05
pat_CP2 [9]	1055	376	A11:01		B14:02	B44:02		
pat_FL [9]	2127	873	A03:01	A11:01	B44:03	B50:01		
pat_J [9]	1215	551	A02:01	A03:01	B07:02		C07:02	
pat_JPB3 [9]	1009	487	A02:01	A11:01	B27:05	B56:01		
pat_JT2 [9]	846	287	A11:01		B18:03	B35:01		
pat_M [9]	1523	574	A03:01	A29:02	B08:01	B44:03	C07:01	C16:01
pat_MA [9]	1998	808	A02:01	A29:02	B44:03	B57:01	C07:01	C16:01
pat_ML [9]	1580	811	A02:01	A11:01	B40:01	B44:03		
pat_NS2 [9]	435	91	A02:01		B13:02	B41:01		
pat_NT [9]	1522	252	A01:01	A32:01				
pat_PF1 [9]	2484	1049	A01:01	A02:01	B07:02	B44:03	C07:02	C16:01
pat_R [9]	1406	549	A03:01	A29:02	B08:01	B44:03	C07:01	C16:01
pat_RT [9]	1425	333	A01:01	A02:01	B18:01	B39:24	C05:01	C07:01
pat_SR [9]	1642	505	A02:01	A23:01	B18:01	B44:03		
pat_ST [9]	771	283	A03:01	A24:02	B07:02	B27:05		

Table B: Amino acid frequencies at non-anchor positions (P4 to P7) in HLA-I ligands from pooled HLA peptidomics data.

A	0.060782
C	0.001453
D	0.045653
E	0.073598
F	0.035112
G	0.049635
H	0.037461
I	0.06487
K	0.059036
L	0.083318
M	0.011609
N	0.039796
P	0.090261
Q	0.056825
R	0.050559
S	0.066348
T	0.059356
V	0.083667
W	0.006434
Y	0.024228

Table C: Ranking of the neo-antigens identified in two lung cancer samples [10]. Column 4 shows the ranking based on our predictor. Column 6-8 show the ranking based on NetMHC[11], NetMHCpan[12] and NetMHCstabpan[13].

Sample	Sequence	Protein	Mutation	Rank	Net-MHC	Net-MHC-pan	Net-MHC-stabpan	# Candidates
L011	SVTNEFC <u>L</u> K	DOCK10	R1149L	13	26	21	9	7414
L011	GTSAPRKKK	XIRP2	Q2514K	58	213	281	110	7414
L011	FAFQ <u>E</u> YDSF	MTFR2	D326Y	65	6	10	296	7414
L011	RSMRTVYGL <u>F</u>	OR2A2	C317F	712	83	136	221	7414
L011	GPEELGL <u>P</u> M	SOHLH2	L159M	1070	109	158	292	7414
L013	ALQSR <u>L</u> QAL	CACNA1I	R464L	3	151	178	118	2700
L013	KVCCCQILL	OR4P4	W300C	168	224	153	153	2700
L013	YSNYY <u>C</u> GLRY	KRTAP20-2	G9C	177	63	124	58	2700

Table D: List of cancer testis antigens with epitopes and HLA restriction information from the CTDatabase [14].

Protein	Epitope	Allele	Protein	Epitope	Allele
BAGE	AARAVFLAL	C16:01	MAGEA12	EGDCAPEEK	C07:01
CCDC110	NYNNFYRFL	A24:02	MAGEA2	YLQLVFGIEV	A02:01
CCDC110	EYSKECLKEF	A24:02	MAGEA2	EYLQLVFGI	A24:02
CCDC110	EYLSLSDKI	A24:02	MAGEA2	EGDCAPEEK	C07:01
CTAG1A	SLLMWITQC	A02:01	MAGEA3	EVDPIGHLY	A01:01
CTAG1A	ASGPGGGAPR	A02:01	MAGEA3	FLWGPRALV	A02:01
CTAG1A	MPFATPMEA	B35:01	MAGEA3	KVAELVHFL	A02:01
CTAG1A	MPFATPMEA	B51:01	MAGEA3	TFPDLESEF	A24:02
CTAG1A	ARGPESRLL	C06:02	MAGEA3	VAELVHFL	A24:02
CTAG1A	LAMPFATPM	C03:03	MAGEA3	MEVDPIGHLY	B18:01
CTAG2	ELVRRILSR	A68:01	MAGEA3	EVDPIGHLY	B35:01
CT83	RQKRILVNL	B15:01	MAGEA3	AELVHFL	B40:01
MAGEA1	EADPTGHSY	A01:01	MAGEA3	MEVDPIGHLY	B44:02
MAGEA1	KVLEYVIKV	A02:01	MAGEA3	EGDCAPEEK	C07:01
MAGEA1	SLFRAVITK	A03:01	MAGEA4	EVDPASNTY	A01:01
MAGEA1	EVYDGREHSA	A68:02	MAGEA4	GVYDGREHTV	A02:01
MAGEA1	RVRFFPSL	B07:02	MAGEA4	NYKRCPVI	A24:02
MAGEA1	EADPTGHSY	B35:01	MAGEA6	EVDPIGHVY	B35:01
MAGEA1	SAYGEPRKL	C03:03	MAGEA6	EGDCAPEEK	C07:01
MAGEA1	SAYGEPRKL	C16:01	MAGEA6	ISGGPRISY	C16:01
MAGEA10	GLYDGMEHL	A02:01	MAGEA9	ALSVMGVYV	A02:01
MAGEA12	FLWGPRALV	A02:01	SAGE1	LYATVIHDI	A24:02
MAGEA12	FLWGPRALV	C07:01	SSX2	KASEKIFYV	A02:01

References

1. Bassani-Sternberg M, Bräunlein E, Klar R, Engleitner T, Sinitcyn P, Audehm S, et al. Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. *Nat Commun. Nature Publishing Group*; 2016;7: 13404. doi:10.1038/ncomms13404
2. Abelin JG, Keskin DB, Sarkizova S, Hartigan CR, Zhang W, Sidney J, et al. Mass Spectrometry Profiling of HLA-Associated Peptidomes in Mono-allelic Cells Enables More Accurate Epitope Prediction. *Immunity*. 2017;46: 315–326. doi:10.1016/j.immuni.2017.02.007
3. Habib N, Kaplan T, Margalit H, Friedman N. A novel Bayesian DNA motif comparison method for clustering and retrieval. Fraenkel E, editor. *PLoS Comput Biol*. 2008;4: e1000010. doi:10.1371/journal.pcbi.1000010
4. Bassani-Sternberg M, Pletscher-Frankild S, Jensen LJ, Mann M. Mass spectrometry of human leukocyte antigen class I peptidomes reveals strong effects of protein abundance and turnover on antigen presentation. *Mol Cell Proteomics*. 2015;14: 658–673. doi:10.1074/mcp.M114.042812
5. Mommen GPM, Frese CK, Meiring HD, van Gaans-van den Brink J, de Jong APJM, van Els CACM, et al. Expanding the detectable HLA peptide repertoire using electron-transfer/higher-energy collision dissociation (ETHcD). *Proc Natl Acad Sci USA. National Acad Sciences*; 2014;111: 4507–4512. doi:10.1073/pnas.1321458111
6. Gloger A, Ritz D, Fugmann T, Neri D. Mass spectrometric analysis of the HLA class I peptidome of melanoma cell lines as a promising tool for the identification of putative tumor-associated HLA epitopes. *Cancer Immunol Immunother*. 2016;65: 1377–1393. doi:10.1007/s00262-016-1897-3
7. Ritz D, Gloger A, Weide B, Garbe C, Neri D, Fugmann T. High-sensitivity HLA class I peptidome analysis enables a precise definition of peptide motifs and the identification of peptides from cell lines and patients' sera. *Proteomics*. 2016;; n/a–n/a. doi:10.1002/pmic.201500445
8. Shraibman B, Kadosh DM, Barnea E, Admon A. Human Leukocyte Antigen (HLA) Peptides Derived from Tumor Antigens Induced by Inhibition of DNA Methylation for Development of Drug-facilitated Immunotherapy. *Mol Cell Proteomics. American Society for Biochemistry and Molecular Biology*; 2016;15: 3058–3070. doi:10.1074/mcp.M116.060350
9. Pearson H, Daouda T, Granados DP, Durette C, Bonneil E, Courcelles M, et al. MHC class I-associated peptides derive from selective regions of the human genome. *J Clin Invest. American Society for Clinical Investigation*; 2016;126. doi:10.1172/JCI88590
10. Bentzen AK, Marquard AM, Lyngaa R, Saini SK, Ramskov S, Donia M, et al. Large-scale detection of antigen-specific T cells using peptide-MHC-I multimers labeled with DNA barcodes. *Nat Biotechnol. Nature Research*;

2016;34: 1037–1045. doi:10.1038/nbt.3662

11. Andreatta M, Nielsen M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinformatics*. 2015. doi:10.1093/bioinformatics/btv639
12. Nielsen M, Andreatta M. NetMHCpan-3.0; improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length datasets. *Genome Med. BioMed Central*; 2016;8: 33. doi:10.1186/s13073-016-0288-x
13. Rasmussen M, Fenoy E, Harndahl M, Kristensen AB, Nielsen IK, Nielsen M, et al. Pan-Specific Prediction of Peptide-MHC Class I Complex Stability, a Correlate of T Cell Immunogenicity. *J Immunol. American Association of Immunologists*; 2016;197: 1517–1524. doi:10.4049/jimmunol.1600582
14. Almeida LG, Sakabe NJ, deOliveira AR, Silva MCC, Mundstein AS, Cohen T, et al. CTdatabase: a knowledge-base of high-throughput and curated data on cancer-testis antigens. *Nucleic Acids Res*. 2009;37: D816–9. doi:10.1093/nar/gkn673