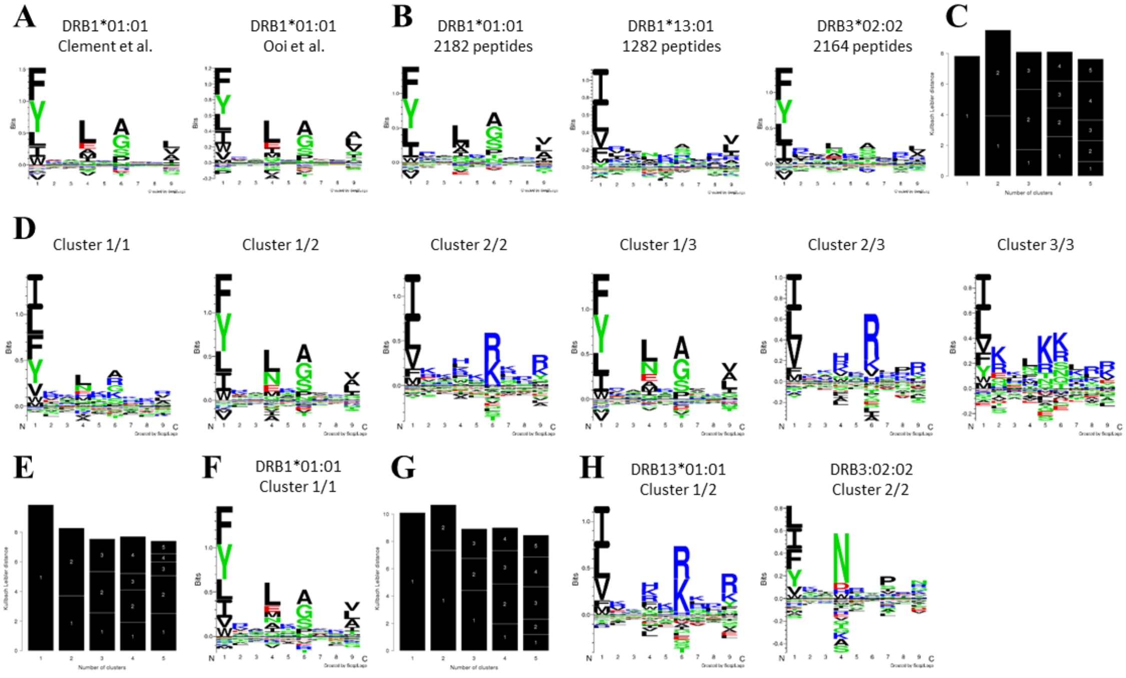
Supporting Information



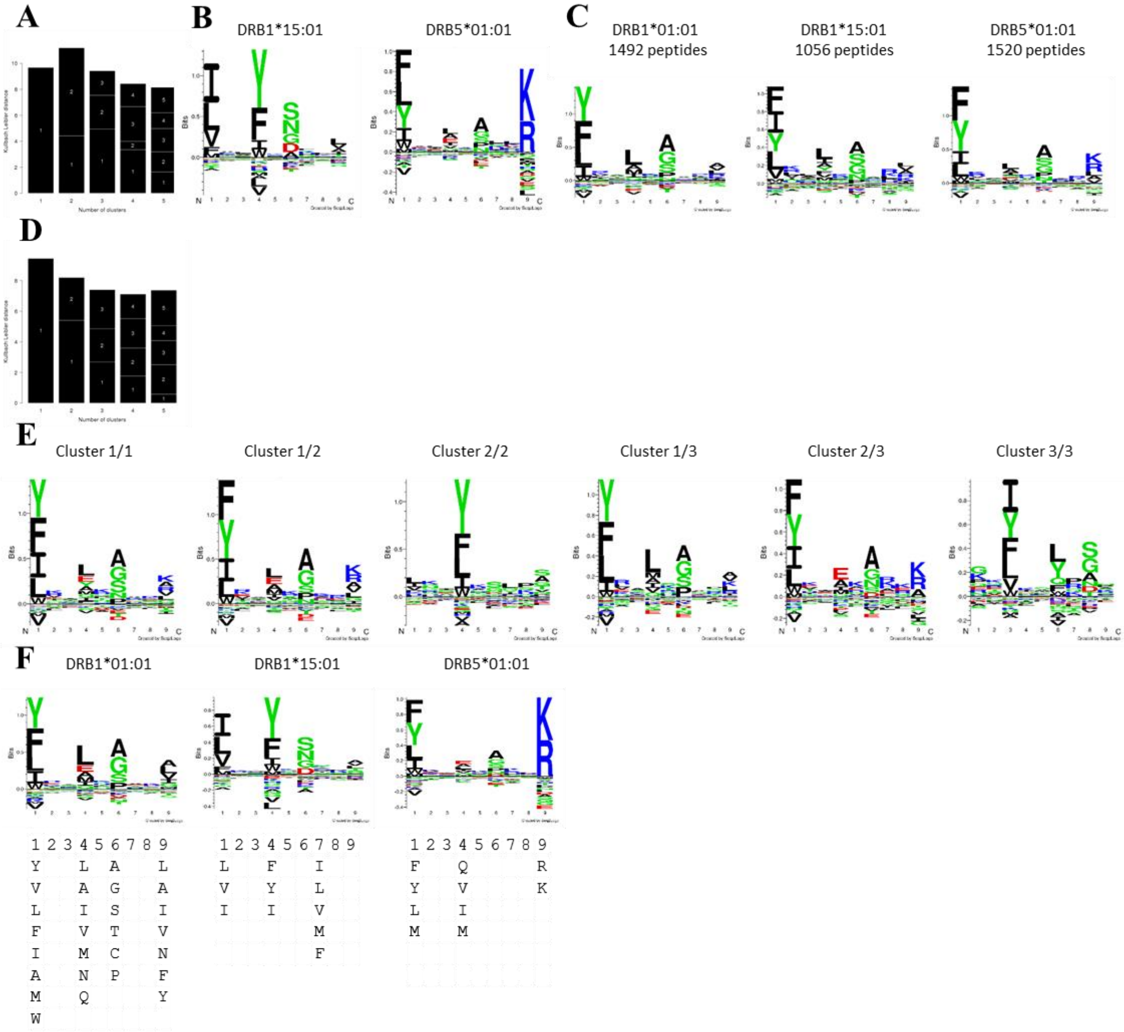
1	2	3	4	5	6	7	8	9
Y			L		A			L
V			A		G			A
L			I		S			I
F			V		T			V
I			M		C			N
A			N		P			F
M			Q					Y
W								

1	2	3	4	5	6	7	8	9
I			L		R			Y
L			V		K			F
V			M					A
			A					S
			W					T
			Y					

1	2	3	4	5	6	7	8	9
Y			N		A			L
F					S			V
I					P			I
L					D			S
					E			G

Supporting Information Figure 1: Assignment of peptides identified from Maver-1 cells to respective HLA class II alleles. (A) Definition of the DRB1*01:01 motif from the literature. Peptide sequences identified from DRB1*01:01 homozygous cell lines reported by Clement et al. and Ooi et al. were subjected to Gibbs Cluster-2.0 analysis. Left panel: Sequence logo for the peptide sequences reported by Clement et al. [Clement, C. C., et al. JBC 2016, 291, 11, 5576-5595]. Right panel: Sequence logo for the peptide sequences reported by Ooi et al. [Ooi, J. D., et al. Nature 2017, 545, 243-250]. **(B)** NetMHCIIpan 3.1 analysis of peptide sequences identified from Maver-1 cells. All peptide sequences from Maver-1 cells were submitted to NetMHCIIpan 3.1. HLA-specific motifs were plotted from core sequences of peptides with %Rank below 10. **(C-D)** Gibbs Cluster-2.0 analysis of the HLA DR peptide sequences identified from Maver-1 cells. **(C)**

Information content of the results between one and five clusters per solution. The solution with two clusters (i.e. with the largest Kullbach Leibler distance) was proposed as the optimal result. **(D)** Sequence logos reported by Gibbs Cluster-2.0 for the solutions (from left to right) with one, two, and three clusters, respectively. **(E-F)** To identify DRB3*02:02-eluted peptides, all Maver-1 peptide sequences were scored against the PSSM of the DRB1*01:01-specific motif from **(A)** with the scoring function described in the Materials and Methods section. A score greater than 3.3 was found to predict binding to DRB1*01:01, as defined by comparing scores obtained from true positive and true negative data via ROC analysis. **(E-F)** Gibbs Cluster-2.0 analysis of peptide sequences predicted to bind to DRB1*01:01. **(E)** Information content of the results between one and five clusters per solution. The solution with one cluster was proposed as the optimal result. **(F)** Sequence logo reported for DRB1*01:01 (upper panel) and respective SYFPEITHI motif (lower panel). **(G-H)** Gibbs Cluster-2.0 analysis of peptide sequences not predicted to bind to DRB1*01:01. **(G)** Information content of the results between one and five groups per solution. The solution with two clusters was proposed as the optimal result. **(H)** Sequence logos (upper panels) reported for the two clusters and comparison with the SYFPEITHI motifs (lower panels) of DRB1*13:01 (left panels) and DRB3*02:02 (right panels), respectively.



1	2	3	4	5	6	7	8	9
Y			L		A			L
V			A		G			A
L			I		S			I
F			V		T			V
I			M		C			N
A			N		P			F
M			Q					Y
W								

1	2	3	4	5	6	7	8	9
L			F			I		
V			Y			L		
I			I			V		
						M		
						F		

1	2	3	4	5	6	7	8	9
F			Q					R
Y			V					K
L			I					
M			M					

Supporting Information Figure 2: Assignment of peptides identified from DOHH2 cells to respective HLA class II alleles. (A-B) Definition of the DRB1*15:01 and DRB5*01:01 motifs from the literature. Peptide sequences reported by Ooi et al. [Ooi, J. D., et al. Nature 2017, 545, 243-250] eluted from a cell line positive for DRB1*15:01 and DRB5*01:01 were subjected to Gibbs Cluster-2.0 analysis. **(A)** Information content of the results between one and five clusters per solution. The solution with two cluster (i.e. with the largest Kullbach Leibler distance) was proposed as the optimal result. **(B)** DRB1*15:01 and DRB5*01:01-specific motifs. According to motifs reported by SYFPEITHI, clusters reported by Gibbs Cluster-2.0 could be annotated unambiguously to DRB1*15:01 (left panel) and DRB5*01:01 (right panel). **(C)** NetMHCIIpan 3.1

analysis of peptide sequences identified from DOHH2 cells. All peptide sequences from DOHH2 cells were submitted to NetMHCIIpan 3.1. HLA-specific motifs were plotted from core sequences of peptides with %Rank below 10. **(D-E)** Gibbs Cluster-2.0 analysis of the HLA DR peptide sequences identified from DOHH2 cells. **(D)** Information content of the results between one and five clusters per solution. The solution with one cluster (i.e. with the largest Kullbach Leibler distance) was proposed as the optimal result. **(E)** Sequence logos reported by Gibbs Cluster-2.0 for the solutions (from left to right) with one, two, and three clusters, respectively. **(F)** DOHH2 specific motifs as determined through scoring of the DOHH2 peptide sequences against the PSSMs from **(A)** and **Supporting Information Figure 1A**. Each peptide was assigned to the allele for which the highest score was obtained from the scoring procedure outlined in Material and Methods. Finally, core sequences (i.e. the 9mer sequence with the highest score) were submitted to Seq2Logo-2.0 to obtain respective sequence logos.