

Table S1. Anchor motifs and HLA genotypes and serotypes used to define known epitopes for each HLA allele.

HLA allele	Anchor motif ^a	HLA identifiers for which epitopes have been reported ^b
A*02	X[LVQM]XXXXXXXXZZ[LV]	A2, A02, A*0201, A*0202, A*0205, A*0206, A*0207, A*0214
A*23	X[IY]XXXXXXXXZZ[IW]	A23, A*2301
A*29	X[EM]XXXXXXXXZZ[YL]	A29, A*2902
A*30	X[YFVLMIT]XXXXXXXXZZ[LYFM]	A30, A*3002
A*3001	X[YFVLMIT]XXXXXXXXZZ[LYFM]	A30
A*3002	X[YFVL]XXXXXXXXZZ[Y]	A30, A*3002
A*68	X[TV]XXXXXXXXZZ[RKVL]	A68, A*6802
A*6802	X[TV]XXXXXXXXZZ[VL]	A68, A*6802
A*0301	X[LVMIAST]XXXXXXXXZZ[KYFR]	A3, A*0301
B*0702	X[PV]XXXXXXXXZZ[LF]	B7, B07, B*0702
B*0801	XXZ[KR]XXXXXXXXZZ	B8, B*0801
B*15	X[QKH]XXXXXXXXZZ[YFL]	B*15, B*1501, B*1503, B*1510
B*1503	X[QK]XXXXXXXXZZ[YF]	B*15, B*1503
B*1510	X[HQ]XXXXXXXXZZ[LF]	B*15, B*1510
B*42	X[P]XXXXXXXXZZ[L]	B42, B*4201, B*4202
B*4201	X[P]XXXXXXXXZZ[L]	B42, B*4201
B*44	X[EMILDV]XXXXXXXXZZ[YF]	B44, B*4402, B*4403, B*4415
B*4403	X[EMILDV]XXXXXXXXZZ[YF]	B44, B*4403
B*57	X[ATS]XXXXXXXXZZ[FWY]	B57, B*5701, B*5702, B*5703
B*5801	X[ATS]XXXXXXXXZZ[FWY]	B58, B*58, B*5801
B*5802	X[ST]XXXZ[R]XXZ[F]	B58, B*58, B*5802
B*81	X[P]XXXXXXXXZZ[L]	B81, B*8101
C*02	X[A]XXXXXXXXZZ[L]	Cw2
C*0202	X[A]XXXXXXXXZZ[L]	Cw2
C*04	X[YFP]XXXXXXXXZZ[LFM]	Cw4, Cw04
C*06	XXXXXXXXZZ[LIVY]	Cw*0602
C*0602	XXXXXXXXZZ[LIVY]	Cw*0602
C*07	X[RHKYPQ]XXXXXXXXZZ[YFLWM]	Cw7, Cw*07
C*08	X[A]XXXXXXXXZZ[LM]	Cw8, Cw08, Cw*0802, Cw*0804
C*17	X[A]XXXXXXXXZZ[L]	No documented epitopes
C*1701	X[A]XXXXXXXXZZ[L]	No documented epitopes
C*18	X[RQ]XXXXXXXXZZ[LY]	Cw18, Cw*1801

^a Amino acids in square brackets denote anchor residues, X denotes an indeterminate amino acid and Z denotes an optional, indeterminate amino acid which affects the final length of the epitope.

^b These are the HLA alleles/serotypes consistent with the HLA allele shown in column one, for which experimentally determined epitopes have been reported in Yusim et al. (2009).