

HXB2 amino acid sequence	Author wildtype sequence if different to HXB2 B-clade ancestral sequence if different to HXB2	Gene (HXB2 [S1])	Amino acid location (HXB2 [S1])	HLA	Proportion of Caucasians with HLA [36] (p)	Escape mutations ^b	Reference
KIRLRPGGK	-	p17 gag	18-26	A3	0.224	K9R K1R & K9R	Wilson 1999 [S2], Sanchez-Merino 2005 [S3] Wilson 1999 [S2]
RLRPGGKKK	-	p17 gag	20-28	A3	0.224	K9Q K9T K9R	Allen 2004 [6] Allen 2004 [6] Geels 2003 [19]
GGKKKYKL	-	p17 gag	24-32	B8	0.143	K3R K5Q K7R K7Q K5Q & K7R K3R & K5T	Reid 1996 [S4] Price 1997 [3] Reid 1996 [S4] Reid 1996 [S4] Milicic 2005 [S5] Milicic 2005 [S5]
KYKLKHIVW	-	p17 gag	28-36	A24	0.196	K1Q K3R & I7L K1R & K3R & I7L	Yokomaku 2004 [S6] Yokomaku 2004 [S6] Yokomaku 2004 [S6]
GSEELRSLY	-	p17 gag	71-79	A1	0.262	L5I & R6K Y9F	Milicic 2005 [S5] Milicic 2005 [S5]
SLYNTVATL	-	p17 gag	77-85	A2	0.438	Y3F N4I T5A T5L V6I T8V Y3H & T8V	Jamieson 2003 [S7] Brander 1998 [S8] Brander 1999 [S9] Brander 1998 [S8] Geels 2006 [S10] Jamieson 2003 [S7] Jamieson 2003 [S7]
AADTGHSNQ ^a	AADTGNSSQ AADTGNSSQ	p17 gag	119-127	Not determined		many (see paper)	Jones 2004 [S11]
HSNQVSNQYP ^a	NSSQVSNQYP NSSQVSNQYP	p17 gag	124-133	Not determined		many (see paper)	Jones 2004 [S11]
ISPRTLNAW	-	p24 gag	15-23	B57	0.057	A(-1)P I1M I1L	Draenert 2004 [S12] Pillay 2005 [S13] Pillay 2005 [S13]
KAFSPEVPMF	-	p24 gag	30-40	B57	0.057	A2G	Crawford 2007 [S14]
EVIPMFSAL	-	p24 gag	35-43	A26	0.083	E1K S7A S7T	Kawashima 2005 [S15] Kawashima 2005 [S15] Kawashima 2005 [S15]
TSTLQEQIGW	HXB2 and TSTLQEQIAGW	p24 gag	108-117	B57/B58	0.096	T3N	Leslie 2004 [17], Allen 2004 [6], Feeney 2005 [S16], Pillay 2005 [S13]
KRWILGLNK	-	p24 gag	131-140	B27	0.073	R2K R2G R2Q	Kelleher 2001 [20], Goulder 1997 [8], Geels 2006 [S10] Kelleher 2001 [20] Ammarmond 2005 [S17]
NANPDCKTI	-	p24 gag	193-201	B51	0.126	A2S & T8N	Yokomaku 2004 [S6]
DCKTILKAL	-	p24 gag	197-205	B8	0.143	K3R	Nowak 1995 [21]
CGKEGHTAR	CGKEGHAR CGKEGHAR	p7 gag	32-40	A31	0.050	R9K	Goonetilleke 2009 [5]
GPVKVQWPL	-	RT	18-26	B8	0.143	K3R K3E	Meier 1995 [S18] Meier 1995 [S18]
TAFTIPSI	-	RT	128-135	B51	0.126	I8T	Geels 2006 [S10]
AIFQSSMTK	-	RT	158-166	A3/A11	0.340	Q4L A1T & F3S	Wilson 1999 [S2] Wilson 1999 [S2]
IVLPEKDSW	-	RT	244-252	B57	0.057	V2M V2T D7G	Menendez-Arias 1998 [S19] Menendez-Arias 1998 [S19] Menendez-Arias 1998 [S19]
IPLTEEAEL ³	VPLTEEAEL	RT	293-301	B57	0.057	E5V E5A E5G	Oxenius 2004 [S20]
ETFYVDGAANR ^a	-	RT	438-448	A26	0.083	F3Y & D6N	Wilson 1999 [S2]
QVPLRPMTYK	-	nef	73-82	A3/A11	0.340	V2A L4A R5A M7A T8A	Milicic 2005 [S5] Keonig 1995 [S21] Keonig 1995 [S21] Keonig 1995 [S21] Keonig 1995 [S21]
VPLRPMTY	-	nef	74-81	B35	0.196	T7S Y8F	Oxenius 2004 [S20] Oxenius 2004 [S20]
RPMTYKAAV	RPMTYKGAAL RPMTYKAAAL	nef	77-85	B7	0.166	T4S Y5F	Oxenius 2004 [S20] Oxenius 2004 [S20]
KAVDLSHF	KAALDLSHF	nef	82-90	B57/B58	0.096	A2G	Leslie 2005 [S22]
FLKEKGGL	-	nef	90-97	B8	0.143	E4D K5N K5E L2I & K5N K3E & K5N E4G & K5N	Price 1997 [3] Price 1997 [3] Price 1997 [3] Price 1997 [3] Price 1997 [3] Price 1997 [3]
HTQGYFPDW	-	nef	116-125	B57	0.057	H1N	Pillay 2005 [S13]
TPGPGVRYPL	TPGPGIRYPL	nef	128-137	B42/B7	0.169	P2S P2Q	Pillay 2005 [S13] Pillay 2005 [S13]
RYPLTFGWY	RYPLTFGWCF RYPLTFGWCF	nef	134-143	A24	0.196	Y2F	Furutsuki 2004 [S23]
SRLAFHHVAR ^a	SRLAFHHMAR SSLAFHHVAR	nef	187-196	A31	0.050	R10Q	Goonetilleke 2009 [5]

Table S1. A summary of mutations in HIV-1 gag, RT and nef which have been reported in the literature as conferring CTL escape and confirmed by *in vitro* tests.

In some cases the HBX2 epitope sequence differs from the author wildtype sequence and/or the B-clade ancestral sequence [S24]. These changes are shown in column 2. Note that we have not included in this list any defined escape mutations that lie in sites that differ between the author wildtype sequence and the B-clade ancestral sequence. Epitopes are ordered according to their location within the genome. The proportion of Caucasians with each HLA type is estimated from allele frequencies amongst Caucasians [36], assuming mendelian genetics. This calculation ignores linkage between alleles. ^aEpitopes AADTGHSNQ and HSNQVSQNYP are excluded from analysis of the cross-sectional data (dataset 2) and the longitudinal cohort data (dataset 4) because their HLA-restrictions were not determined. In addition, sequence data were not available from the cross-sectional study for IPLTEEAEL or SRLAFHHVAR or from the longitudinal cohort study for ETFYVDGAANR. Sequence data for CGKEGHTAR were available from neither study. ^bWhere two or more different mutations are given together with the ‘&’ sign, it indicates that an epitope containing that combination of mutations was tested and found to confer escape.