

Identification of novel molecular determinants of co-receptor usage in HIV-1 subtype F V3 envelope sequences

Maximiliano Distefano, Esteban Lanzarotti, María Florencia Fernández, Andrea Mangano, Marcelo Martí, Paula Aulicino

Supplementary table 1: Comparative table of amino acid frequency and contacts between X4-F and X4-B with CXCR4 (A) and R5-F and R5-B with CCR5 (B).

A

Subtype F V3 loop-amino acid frequency	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
	C (100)	T (100)	R (100)	P (100)	N (88.46) S (11.54)	N (100)	N (100)	R (3.85) T (96.15)	R (84.62) S (7.69) V (7.69)	K (96.15) T (3.85)	A (3.85) G (19.23) R (19.23) S (57.69)	I (100)	A (3.85) H (19.23) Q (42.31) R (15.38) S (7.69) W (7.69) Y (3.85)	I (65.38) L (30.77) F (3.85)	G (100)	P (100)	G (100)	R (80.77) Q (7.69) H (7.69) K (3.85)	A (88.46) S (3.85) T (3.85) V (3.85)	F (88.46) V (11.54)	H (11.54) L (3.85) R (7.69) Y (76.92)	A (38.46) S (3.85) T (57.69)	G (3.85) I (3.85) M (3.85) N (3.85) T (84.62)	A (3.85) E (7.69) G (84.62) K (3.85)	A (7.69) D (11.54) E (23.08) G (7.69) K (19.23) Q (7.69) R (23.08)	I (100)	I (84.62) K (3.85) T (3.85) V (7.69)	G (100)	D (96.15) N (3.85)	I (100)	R (100)	K (80.77) M (3.85) Q (3.85) R (11.54)	A (100)	H (84.62) Y (15.38)	C (100)
subtype F V3 loop-CXCR4 interactions			R - D22			N - S23		T/R - Y21 R - K25	R/V/S - Y12	T-Y21 K/T - D193 K/T - L266	R - M1 R - Y7 R - Y12	I - F189 I - Y190	R/Q/H/S/W/Y - M1 W/Y - E2 R/H/Q - R30	F - E2 I/L - S178 F - A180 I/L/F - I185 F/L - D187		P - F36 P - L41	G - W64	H/K - W64 R/H/K - H113 R/H/K/Q - R188	S - H281 S/T/V - I284	F/V - R188 F/V - Y190 F/V - V196 F - Q200 F - I284	R/Y - R30 R/H/Y/L - H281	T - L266	M/T/I - M1 M - C28 M/T/I - F29	A - K25 K - P27	K/A - K25 R/Q/K - P27	I - K25 I - S23 I - P27				R - E14 R - G17 R - S18	R - G17		Y - D20		
Subtype B X4 V3 Loop-amino acid frequency	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
	C (100) A (4.76) I (2.38)	T (92.86) A (4.76) I (2.38)	R (100)	P (100)	N (85.72) G (7.14) S (2.38) Y (2.38) F (2.38)	N (80.95) K (16.67) Y (2.38)	N (71.43) Y (11.9) T (9.52) K (4.76) H (2.38)	T (69.05) K (14.29) I (11.9) V (4.76)	R (76.19) K (16.67) I (4.76) M (2.38)	K (66.67) R (19.05) N (7.14) Q (2.38) G (2.38) T (2.38)	R (59.52) G (26.19) S (11.9) K (2.38)	I (71.43) L (21.43) V (7.14)	S (26.19) Y (23.81) T (21.43) R (16.67) H (9.52) P (2.38)	I (47.62) L (28.57) V (9.52) T (7.14) M (4.76) A (2.38)	G (97.62) S (2.38)	P (80.95) Q (14.29) T (2.38) L (2.38)	G (97.62) S (2.38)	R (97.62) K (2.38)	A (64.29) V (23.81) K (7.14) Y (2.38) R (2.38)	F (45.24) V (23.81) L (14.29) W (7.14) Y (4.76) M (4.76)	Y (92.86) S (7.14)	T (80.95) A (11.9) V (4.76) R (2.38)	T (85.71) M (11.9) L (2.38)	G (40.48) R (28.57) K (14.29) E (9.52) D (7.14) K (4.76) N (2.38)	Q (54.76) G (14.29) R (9.52) E (9.52) D (4.76) K (4.76) N (2.38)	I (90.48) V (9.52)	V (23.81) T (9.52) E (4.76) R (2.38)	G (100)	D (76.19) Y (11.9) T (2.38)	I (97.62) T (2.38)	R (97.62) K (2.38)	Q (50) K (28.57) T (2.38)	A (97.62) T (2.38)	H (66.67) Y (30.95) R (2.38)	C (100)
Subtype B V3 loop-CXCR4 interactions			R - D22		F/Y - D20 Y - D22	N/K/Y - S23		I/K/T/V - Y21 I/K/V - K25	R/M - Y12	R - S18 R/T - Y21 R/N/Q/K/T - D193 R/T/K - L266	R/K - M1 R - Y7 R - Y12	I/L/V - F189 I/L/V - Y190	R/H/P/S/T/Y - M1 Y - E2 R/H - R30	I/L/M - S178 M - R183 I/L/M/T/V - I185 I/L/M - D187 M - F189	S - I185	Q/I/L/P/T - F36 T - N37 Q/I/L/P/T - L41 L - S285	G/S - W64	K - W64 K/R - H113 R - Y116 K/R - R188	A/K - R188 R/V - H281 R/K/Y/V - I284 R/K/Y - S285	L/M/F/Y/V - R188 L/M/F/V/W - Y190 L/M/F/W/Y/V - V196 F - Q200 W - D262 F - I284	Y - R30 S/Y - H281	R/T/V - L266	L/M/T - M1 LM - C28 L/M/T - F29	R/K - K25 K - P27 R/K - C28	R - M1 K/Q/R - P27	I - S23 I/V - K25 I - P27				R - E14 R - G17 R - S18	R - G17		Y - D20		

Shaded amino acids did not show contacts. Amino acids that were found to generate different contact between subtypes are marked in bold.

B

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
Subtype F V3 loop aminoacid frequency	C (100)	T (100)	R (100)	P (100)	N (100) N (6.25) P (6.25)	N (100)	N (100)	T (100)	R (100)	K (100)	G (6.25) S (93.75)	I (100)	Q (62.5) H (12.5) R (12.5) N (6.25) P (6.25)	I (62.5) L (31.25) M (6.25)	G (100)	P (100)	G (100)	R (68.75) Q (18.75) K (12.5)	A (87.5) S (12.5)	F (93.75) I (6.25)	Y (100)	A (62.5) T (37.5)	T (100)	G (93.75) E (6.25)	D (62.5) E (12.5) S (12.5) K (6.25) Q (6.25)	I (100)	I (100)	G (100)	D (100)	I (100)	R (100)	K (100)	A (100)	H (93.75) Y (6.25)	C (100)
Subtype F V3 loop-CORE interactions			R - Y15	P - Y15	N - Y14	N - E18		T - Q188 T - Y14	R - P183 R - Y184 R - Q188	K - Q188 K - K191 K - E262 K - F264		I - S179 I - S180	Q - Q170 R/H - Q172 H/N/Q/R - T177	IL - W86 IL/M - Y89 L - W94 IL/M - T177 L/M - C178 IL/M - S179	G - Y89	P - K26 P - Y89 P - Q280	G - W86 G - Y89	Q/R/K - W86 R - Y108 Q/R/K - Y251 Q/K - E283	A/S - M279	F - Y251 F - V254 F/I - L255 F - N258 F/I - M279	Y - N24 Y - V25 Y - D276 Y - M279	A/T - K22	T - Q21 T - K22	E - E18 E - C20 G/E - Q21	Q - Q21 K/E - E172		I - M1		D - Y3	I - Y3	R - Y3 R - Q4 R - D11	K - D11 K - I12		H/Y - I12	
Subtype BV3 Loop aminoacid frequency	C (100)	T (96.67) I (8.33) L (1.67) M (1.67) A (1.67)	R (95) G (3.33) Y (1.67)	P (98.33) Y (1.67)	N (83.33) S (8.33) T (3.33) G (3.33) H (1.67)	N (98.33) T (1.67)	N (98.33) S (1.67)	T (100)	R (93.33) A (1.67) I (1.67) S (1.67) V (1.67)	K (85) R (10) T (3.33) Q (1.67)	S (70) G (30)	I (98.33) V (1.67)	H (65) N (11.67) P (11.67) S (8.33) T (1.67) Q (1.67)	I (75) M (16.67) L (6.67) V (1.67)	G (96.67) A (3.33)	P (98.33) W (1.67)	G (100)	R (75) K (13.33) G (6.67) S (3.33) Q (1.67)	A (86.67) T (8.33) V (5)	F (80) L (6.67) M (3.33) V (3.33) W (3.33) Y (1.67) I (1.67)	Y (96.67) F (3.33)	A (65) T (33.33) V (1.67)	T (95) A (3.33) H (1.67)	G (96.67) T (1.67) E (1.67)	D (41.67) E (36.67) Q (8.33) A (5) R (3.33) G (1.67) K (1.67) S (1.67)	I (95) V (5)	I (86.67) V (10) T (3.33)	G (98.33) E (1.67)	D (75) N (23.33) K (1.67)	I (100)	R (96.67) G (1.67) K (1.67)	Q (85) K (11.67) R (3.33)	A (100)	H (85) Y (13.33) Q (1.67)	C (100)
Subtype BV3 loop-CORE interactions			R/K - Y15	P - Y15	N/S/T - Y14 T - Y15	T - Y14 NT - E18	NS - Y14	T - Q188 T - Y14	R/I - P183 R/IV - Y184 R/IV - Q188	R/Q/K/T - Q188 K/T - K191 R - Q261 R/Q/K - E262 Q/K - F264		I - S179 I - S180	Q/T - Q170 H - E172 N/Q/H - T177 H - S179	IL - W86 IL/MV - Y89 L - W94 IL/M - T177 L/M - C178 IL/M - S179	A/G - Y89	PW - K26 W - A30 PW - Y89 W - A90 P - Q280	G - W86 G - Y89	R/Q/G/K/S - W86 R - Y108 R/Q/K/S - Y251 S - M279 R/Q/K - E283	T/V - Y251 A/T/V - M279 T/V - Q280	L/M/FW - Y251 F - V254 IL/M/FW/Y - L255 F/W - N258 L - L275 IL/M/FW/Y/V - M279	F/Y - K22 F/Y - N24 F/Y - V25 F/Y - D276 F/Y - M279 F - Q280	A/T/V - K22	T - Q21 A/H/T - K22 H - E172	E - E18 E - C20 E/G/T - Q21	R/Q - Q21 E/K - E172	I/T/V - M1		K - M1 NDK - Y3	I - Y3	R - Y3 R - Q4 R - D11	K/Q/R - D11 K - I12		Q/H/Y - I12		

Shaded amino acids did not show contacts. Amino acids that were found to generate different contact between subtypes are marked in bold.

Supplementary table 2: table including all datasets used in the work.

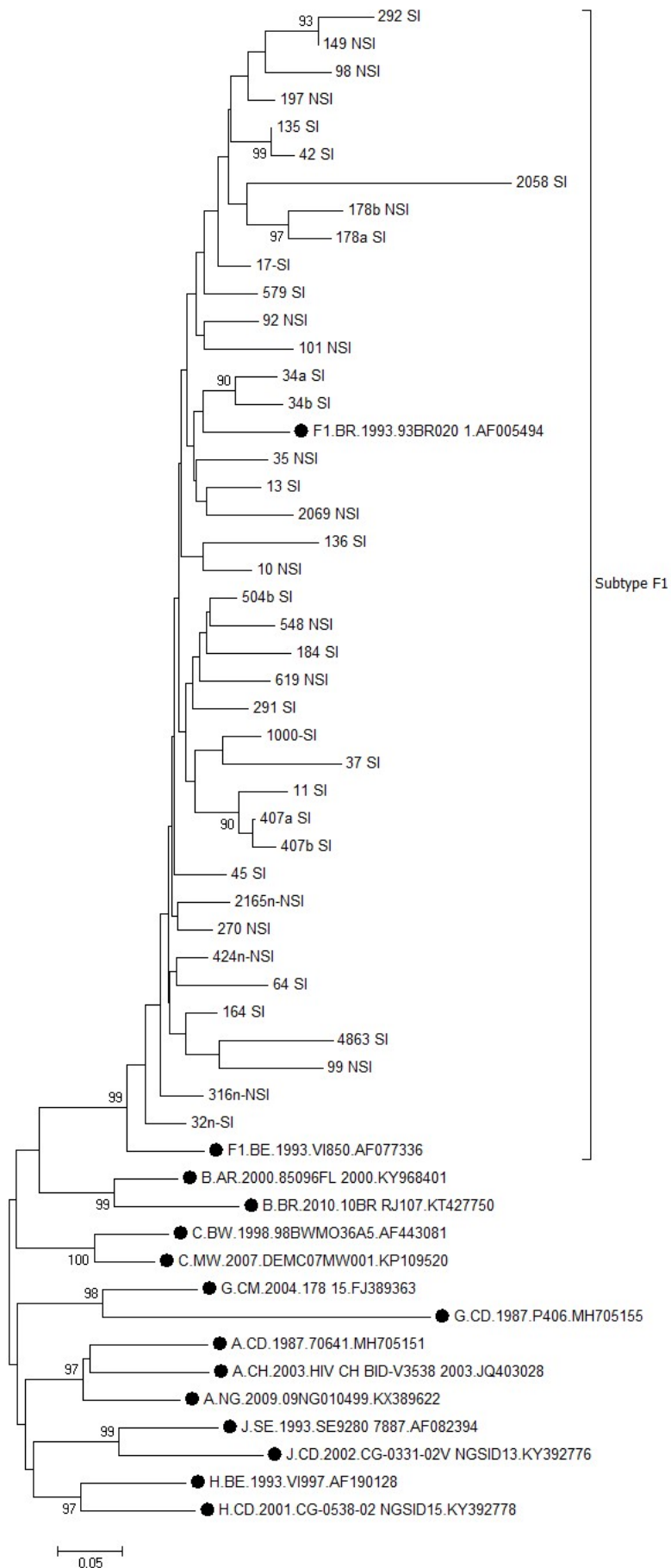
Sequence	Dataset
CTRPNNNTVKGIQIGPGRAFLSTGRIIGDIRKAHC	X4-F
CTRPNNNTRKSIQLGPGRFYATGEIIGDIRKAHC	X4-F
CTRPNNNTRKSIQLGPGRFYTTGQIIGDIRKAYC	X4-F
CTRPNNNTRKSIRIGPGRFYTTGAIIGDIRKAHC	X4-F
CTRPNNNTRKSIWIGPGRFYTTGAIIGDIRKAHC	X4-F
CTRPSNNTSKSIHIGPGRFHAIGRITGDIRKAHC	X4-F
CTRPSNNTKAIKIRIGPGKVYATGRIIGDIRRAHC	X4-F
CTRPNNNTRKSIQLGPGRFYTTGQIIGNIRKAYC	X4-F
CTRPNNNTRKSISLGPGRFYTTGDIIGDIRKAHC	X4-F
CTRPNNNTRKSIQLGPGRFYATGDIIGDIRQAYC	X4-F
CTRPNNNTRKSIQLGPGRFYATGDIIGDIRRAHC	X4-F
CTRPNNNTRKGIHIGPGQAFYTTGEIIGDIRKAHC	X4-F
CTRPNNNTRKRIQLGPGHSFHTTGKIIGDIRKAHC	X4-F
CTRPSNNRRKRIHIGPGRFYTTGGIKGDIRKAYC	X4-F
CTRPNNNTRKRIWIGPGRFYTGKGIVGDIRKAHC	X4-F
CTRPNNNTRKRIQIGPGRVYTTGKIIGDIRRAHC	X4-F
CTRPNNNTRKSIYIGPGHAFHTTGRIIGDIRKAHC	X4-F
CTRPNNNTRKSIRIGPGRVYATGRIIGDIRKAHC	X4-F
CTRPNNNTRKGIHIGPGRFYATEKIIGDIRKAHC	X4-F
CTRPNNNTVKGIQIGPGRFRTTGKIIGDIRKAHC	X4-F
CTRPNNNTRKSIHLGPGRFYTNAEIIIGDIRKAHC	X4-F
CTRPNNNTRKGIAIGPGRTFYATEKIIGDIRKAHC	X4-F
CTRPNNNTSTRISIGPGRFRTMGRIVGDIRMAHC	X4-F
CTRPNNNTRKSIRFGPGQAFYTTGEIIGDIRKAHC	X4-F
CTRPNNNTRKGIHIGPGRFYATEKIIGDIRKAHC	R5-F
CTRPNNNTRKSIPIGPGKTFYATGDIIGDIRQAHC	R5-F
CTRPNNNTRKSIQLGPGRFYATGDIIGDIRKAHC	R5-F
CTRPNNNTRKSIQIGPGRFYTTGDIIGDIRKAHC	R5-F
CTRPNNNTRKSIQIGPGRFYATGSIIGDIRKAHC	R5-F
CTRPNNNTRKSIINIGPGRFYATGDIIGDIRKAHC	R5-F
CTRPNNNTRKSIRIGPGQSFYTTGDIIGDIRKAHC	R5-F
CTRPNNNTRKSIQLGPGRAIYATGDIIGDIRKAHC	R5-F
CARPNNNTRKSIQIGPGRFYTTGEIIGDIRKAHC	R5-F
CTRPNNNTRKSIQLGPGKAFYTTGDIIGDIRKAHC	R5-F
CTRPNNNTRKSIQLGPGRFYTTGEIIGDIRKAHC	R5-F
CTRPNNNTRKSIQMGP GKAFYATGEIIGDIRKAHC	R5-F
CTRPNNNTRKSIHIGPGQAFYATGSIIGDIRKAHC	R5-F
CTRPNNNTRKSIQIGPGRFYTTGQIIGDIRKAYC	R5-F
CTRPNNNTRKSIPIGPGRFYATGDIIGDIRKAHC	R5-F
CTRPNNNTRKSIQIGPGRFYATGDIIGDIRKAHC	R5-F
CTRPNNNTRKSIISIGPGRFYAHGDIVGDIRQAHC	R5-B
CTRYNNNTRKSIHLGPGRFYATGDIIGDIRQAQC	R5-B
CTRPNNNTRKSIHIAPGRTFYATGDIIGDIRQAHC	R5-B
CTRPNNNTRKSIHMGPGSVWYATGEIIGDIRQAHC	R5-B
CTRPNNNTRKSIISIGPGRFYATGDIIGDIRQAHC	R5-B
CTGPNNNTRKSIHIGPGRFYTTGEIIGDIRQAHC	R5-B

CTRPNTNTRKSINIGPGRAFYTTGEIIGNIRQAHC	R5-B
CTRPNNNTRKGIHMGPGKVFYATGQIIGNIRQAHC	R5-B
CMRPNNNTRKSIPIGPGRAFYTTGEIIGDIRQAHC	R5-B
CTRPGNNTRKSIPIGPGRAFYATGDIIGDIRQAHC	R5-B
CTRPNNNTRRGIQMGPGRAVYATGDIIGDIRQAHC	R5-B
CTRPNNNTRKGIHIGPGGAIYATGAIIEDIRQAHC	R5-B
CTRPNNNTVKSIHIGPGRAFYTTGQIIGNIRQAHC	R5-B
CTRPNNNTRKSIHMGPGKTFATGDIIGDIRQAHC	R5-B
CTRPNNNTRKSIHIGPGRAFYTTGEIIGDIRKAHC	R5-B
CTRPNNNTRKGIHIGPGRALYATGDIIGKIRQAHC	R5-B
CLRPNNNTRKGIHIGPGRAFYTTGEIIGDIRQAHC	R5-B
CTRPSNNTRKSINIGPGRAFYTTGEIIGDIRQAHC	R5-B
CIRPNNNTRKSIPMGPGKAFYATGSIIGNIRQAHC	R5-B
CTRPNNNTRKGIHIGPGRAFYATGDIVGDIKQAHC	R5-B
CARPNNNTRKSIHIGPGRAFYTTGEIIGDIRQAHC	R5-B
CTGPNNNTRRSISIGPGGAFYTTGDIIGDIRQAYC	R5-B
CTRPNNNTRKGIHIGPGRAFYTTGEIIGDIRQAHC	R5-B
CTRPTNTRKSIHIGPGRAFYATGDIIGNIRQAHC	R5-B
CTRPNNNTRKGIHIGPGRAFYTTGEIIGNIRQAHC	R5-B
CTRPNNNTRKGIPIGP GKAFYATGEIIGDIRQAHC	R5-B
CTRPNNNTRKGIHIGPGRAFYTTGGIIGDIRQAHC	R5-B
CTRPSNNTRKSIHMGPGRAFVYTG DVIGDIRQAHC	R5-B
CTRPNNNTRKGIHIGPGKTFYATGQIIGDIRQAHC	R5-B
CTRPNNNTRKSINIGPGRAFFATGDIIGDIRQAHC	R5-B
CTRPNNSTRKSIHMGWG RAFYATGEIIGNIRQAHC	R5-B
CTRPNNNTRKSISIGPRAFYATGEIIGDIRQAYC	R5-B
CIRPNNNTRKSIHMGPGRAFYATGDVIGDIRKAYC	R5-B
CTRPNNNTIKGIHIGPGRAFYTTGQVIGDIRKAYC	R5-B
CTRPNNNTRKSINIGPGGAFYAATDIIGDIRQAHC	R5-B
CTRPTNTRKGIHIGPGRAFYTTGEIIGDIRKAHC	R5-B
CTRPNNNTAKGIHIGPGRAMYATERIVGNIRRAHC	R5-B
CTKPNNNTRRSISLGPGRAYYATGDIIGDIRQAHC	R5-B
CTRPNNNTRKSIHIGPGRAFYTTGDIIGDIRKAHC	R5-B
CTRPNNNTRRSIHIGPGSAFYATGDIIGDIRQAHC	R5-B
CTRPNNNTRKSIHIGPRAFYATGEIVGNIRQAHC	R5-B
CTRPNNNTRKSINIGPGRAWYATGKIIGNIRQAHC	R5-B
CTRPNNNTRRSIHIGPGRAFYTTGEIVGDIRQAHC	R5-B
CTRPNNNTRKGIHIGPGRTFYATGEIVGDIRQAHC	R5-B
CTRPGNNTRKSINIGPGRALYTTGDIIGDIRQAHC	R5-B
CTRPNNNTRKSIHAPGRAFYATGDIIGDIRQAYC	R5-B
CIRPSNNTRTSIHLPGQAVYATGEIIGNIRQAHC	R5-B
CIRPNNNTRKGVHLPGGALYATGAIIGDIRQAYC	R5-B
CTRPNNNTRTGIPMGPGRAMYATGDIIGNIRQAYC	R5-B
CTRPNNNTRKGIHIGPGRVYATGEIIGDIRQAHC	R5-B
CIRPNNNTRKSINIGPGRAFYTTGAITGDIRQAHC	R5-B
CTRPHNNTRKSIHIGPRAFYATGQITGDIRQAYC	R5-B
CTRPNNNTRKSIHIGPGRAFYTTGEIIGNIRQAHC	R5-B
CTRPSNNTRKSITIGPGKAFYATGEIIGDIRKAHC	R5-B

CTRPNNNTRKSIHIGPGRAFYTAGEIIGDIRQAHC	R5-B
CTRPNNNTRKSIHVGPGLTYATGDIIGDIGQAHC	R5-B
CTRPNNNTRRSIPIGPGRAFYTGDIIGDIRKAHC	R5-B
CTRPNNNTRKSIHMGPGKAFYATGDIIGDIRRAHC	R5-B
CTRPSNNTRKSIPIGPGRAFYTGDIIGNIRQAHC	R5-B
CTRPNNNTSQSIHIGPGRAFYTGRIIGDIRQAHC	R5-B
CTRPNNNIRKRIHIGPGRAFYTTRQIIGNIRQAHC	X4-B
CTRPNNNVKRISLSPGRVYTTGEIRGDIRKAYC	X4-B
CTRPKNIRKGLRLGPGRAFYTMGGIVGYIRQAHC	X4-B
CTRPNNNTRRRYIGQGRAVYTTKQIVGDIRKAYC	X4-B
CARPNNNTRKGIHMGPGRAMYATEKITGDIRQAHC	X4-B
CTRPNKTRKGLRLGPGRAFYTMGGIVGYIRQAHC	X4-B
CARPNNNTRKRIYMGTRYMSATEKITGDIRQARC	X4-B
CTRPNNNTRKRIYIGQGRAVYTTKQIVGDIRKAYC	X4-B
CTRPKNIRKGLRLGPGRAFYTMGGIVGNIRQTHC	X4-B
CTRPNKTIRKGLRLGPGRAFYTMGGIEGYIRQAHC	X4-B
CTRPNKTRKGLRLGPGRAFYTLGGIVGYIRQAHC	X4-B
CTRPNNNTRKRIYIGQGRAVYTTTRQIIGDIRKAYC	X4-B
CTRPNKTIRKGLRLGPGRAFYTMGRIEGYIRQAHC	X4-B
CTRPGNKTIRSISLGPGRFASATRQIIGDIRKAYC	X4-B
CTRPNNNTRKRITAGPGRVLYTTGQIIGDIRRAHC	X4-B
CTRPNNNTRNRiyIGQGRAVYTTKQIIGDIRKAYC	X4-B
CTRPGNKTIRSISLGPGRFSVTRQIIGDIRKAYC	X4-B
CTRPYNYKKKKITGPGRVLYTTEEIIGDIRRAHC	X4-B
CTRPNYNKRKRIHIGPGRAFYTTKNIIGTIRQAHC	X4-B
CTRPNNNTRNRiyIGQGRAVYTTKQIVGDIRKAYC	X4-B
CTRPNNNTRKRIYIGQGRAVYTTTRQIVGDIRKAYC	X4-B
CTRPGNNTRRRISIGPGRAFYTTEQIIGNIRQAHC	X4-B
CTRPNNYKKKRITGPGRVLYTTGQIIGDIRRAYC	X4-B
CTRPNNYKKKRITVGPGRVLYTTGQIIGDIRRAHC	X4-B
CTRPNNYKKKRITIGPGRVLYTTGQIIGDIRRAYC	X4-B
CTRPNNYKRKRITGPGRVLYTTGQIIGDIRRAYC	X4-B
CTRPNNHTRKRVTLGPSRVYTTGEITGDIRRAHC	X4-B
CTRPNNNTRKGIHIGLGRVYVTRQIIGDTKRAHC	X4-B
CTRPNNNTRQRISIGPGRAFYTTRQVVGDIRQAHC	X4-B
CTRPNNNTRKRISIGPGRAFYTTRQVIGDIRQAHC	X4-B
CTRPNNNTRNRISIGPGRAFYTTRQVIGDIRQAHC	X4-B
CTRPSNNTRRRISIGPGRAFYTTRQVIGDIRQAHC	X4-B
CTRPNNNTRRRLSIGPGRAFYATRDIIIGDIRQAHC	X4-B
CTRPNNNTRGRLSIGPGRAFYATRDIIIGDIRRAHC	X4-B
CTRPNNNTMKSITIGPGRAFYTGQIIGDIRQAHC	X4-B
CTRPNNNTRKRVSIGPRAWYTTKQIVGDIRQAHC	X4-B
CTRPNNNTRKRVTLGPGRVWYTTGQIIGDIRKAHC	X4-B
CIRPNNNTRKSIPLPGKAWYTTGEIIGDIRKAHC	X4-B
CTRPNNNTKKGIYVGPGRKVYTTDRIIGDIRQAHC	X4-B
CTRPNNNTKRGYVGPGRKVYTTDRIIGDIRQAHC	X4-B
CTRPNNNTKRGYVGPGRKVYTTDRIIGNIRQAHC	X4-B
CTRPFKNVRTSLRIGPGRVFYRTGGITGDIRKAYC	X4-B

Supplementary table 3: 5-fold cross-validation results at different False Positive Rate values for Geno2pheno, WebPSSM_{sinsi}, 11,22,25 X4 Energy model and mixed model.

Method	FPR 10%		FPR 15%		FPR 20%	
	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
Mixed model W=0.3	0.60	0.89	0.68	0.89	0.68	0.89
WebPSSM	0.44	0.94	0.44	0.89	0.44	0.89
Geno2Pheno	0.56	0.94	0.64	0.86	0.68	0.83
11,22,25 X4 Energy	0.36	0.72	0.40	0.72	0.40	0.72



Supplementary figure 1: Neighbor-Joining phylogenetic tree of HIV-1 env nucleotide sequences spanning positions 7089 to 7250 of HXB2 reference genome. Bootstrap support above 90% are shown at branch nodes. Subtype reference genomes retrieved from GenBank are shown with filled circles.