Appendix S1:

Positions which are predictive of PG16 resistance

Fifteen associations between amino-acid variants and resistance/sensitivity against the antibody PG16 were significant after correction for multiple testing (q-value < 0.05). As observed for the antibody PG9, the strongest association for PG16 was observed for asparagine at position 160 (p-value = 2.73e-09). The PNGs in the range between position 152 and 173 are shown in Fig. S1. Similar to PG9, the PNG at position 160 was present for 100% of the viruses sensitive to PG16 underlining the strong necessity of this glycosylation site for neutralization by PG16. Two more associations that had been found by our PG9 sensitivity analysis were also found for PG16, namely for a lysine at position 432 (p-value = 9.90e-06) and a valine at position 372 (p-value = 1.94e-05). One of the 15 polymorphisms significantly associated with PG16 sensitivity/resistance was an arginine at position 306 (p-value = 0.0001, q-value = 0.042). Another arginine at position 315 also had a strong p-value in the test of association with PG16 sensitivity/resistance (p-value = 0.0006, q-value = 0.086). The arginine at position 306 was also highly significantly associated with coreceptor usage (p-value = 1.88-07, q-value = 7.61e-05) as was an arginine at position 315 (see main manuscript).

Additional results regarding the coreceptor switch evaluation on the PGT128-treated mice

The analysis regarding the FPRs of the PGT128-treated mice and the untreated control mice in the main part of the manuscript was based on the assumption that variants emerge independently and that the clonal samples are also independent from each other. To evaluate whether the effect is still existing without this assumption, we selected only one clonal variant per (mouse,date) combination. Each chosen variant was the one with the lowest FPR. This led to 6 FPR values for the variants from the PGT128-treated mice and 17 values for the variants from the untreated controls. The Wilcoxon rank sum test still confirmed that the values are significantly different at significance level $\alpha = 0.05$ (p-value = 0.0105). A boxplot of the FPRs can be seen in Fig. S2.

Association between HIV-1 coreceptor usage and resistance broadly neutralizing antibodies with respect to clades

We also investigated, whether the strong association between coreceptor usage and PG9/PG16-sensitivity was due to clade bias in the data set. Figs S3 – S6 show that for most clades there are more resistant X4-capable variants than sensitive ones while this is reversed for the R5 viruses. Interestingly, the only clades, for which there are more X4-capable variants sensitive to PG9 and PG16 than resistant ones are clades A and CRF01_AE. It was recently suspected that for the latter clade geno2pheno[coreceptor] might overestimate CXCR4-usage[1] implying that the overall association between coreceptor usage and resistance to PG9/PG16 could be even stronger than shown in our analysis.

We did not correct for founder effects in the association analysis like Rolland *et al.*[2], since a phylogenetic tree calculated from the *env* sequences is largely influenced by the positions affecting tropism. This can be seen in the phylogenetic tree annotated with coreceptor usage information (see Fig. S7) showing that many X4-capable variants form clusters. The tree was calculated with PhyML using standard parameters[3] and visualized with TreeDyn[4]. Three variants had to be removed (GenBank:DQ187171, GenBank:DQ187240, GenBank:DQ187269), since only the V3 part of the *env* sequence was available.

Additional evaluation regarding the antibody panels

The analysis for the antibody panels is based on the hypothesis that dual-tropic viruses cannot establish new infections as effectively as R5 viruses. In order to analyze the situation in which they could, we furthermore evaluated whether the highly significant result reported in the previous section is mainly driven by dual-tropic viruses. To this end, we constructed a predictor to distinguish R5-capable (i.e. R5 and dual-tropic viruses) from X4 viruses. We trained a support vector machine on all *Env* sequences that were used for training the geno2pheno[coreceptor] model and for which the label R5, X4 or dual-tropic was available. We could not include sequences only labeled with syncytium-inducing/non-syncytium-inducing, since dual-tropic viruses are usually syncytium-inducing as are X4 viruses. After creating a multiple sequence alignment of all V3 loop sequences via MUSCLE[5] we performed a five-fold cross validation for an SVM with a polynomial kernel. The parameter that penalizes the magnitude of the slack variables (*C*) was tested for orders of magnitude covering the range 0.01, 0.1, …, 1000

and the degree (*d*) of the polynomial kernel was tested in the range 1, 2, …, 10. The performance measure was area-under-receiver-operating-characteristic curve (AUC). The training set contained 2430 R5-capable sequences and 423 X4 sequences. We applied different setting of *C* for positive and negative examples to account for this bias. In particular, the sum of the slack variables of the X4 sequences was scaled by the ratio of the number of R5-capable sequences and the number of X4 sequences in the SVM objective. The parameters resulting in best performance were $C = 1$ and $d = 7$. We trained an SVM with these parameters on the whole training set and predicted the labels for the 199 test panel sequences.

Due to the small fraction of X4 sequences in the training set, robustly estimating an FPR analogous to the FPR of geno2pheno[coreceptor] is not straightforward, when considering R5-capable viruses as positive examples and X4 viruses as negative examples. Therefore, we tested whether there is a significant difference in the signed distances of each sequence to the separating hyperplane of the SVM. The higher this value is for a certain sequence the more likely it is R5-capable and the lower it is the more likely the sequence is an X4 sequence.

According to a Wilcoxon rank sum test, the difference between the medians of these values for the resistant and sensitive viruses is significant for PG9 and PG16 (p-value = 0.0020 and p-value = 0.0012) and not significant for VRC01 and VRC-PG04 (p-value = 0.2937 and p-value = 0.8017). For PG9 and PG16 the medians of the groups of resistant viruses were smaller than the medians of the sensitive groups, meaning that there are proportionally more X4 viruses in the resistant groups, further supporting the hypothesis that accounting for coreceptor usage in HIV vaccine studies is important.

Mapping

The full mapping of the IDs from Doria-Rose et al.[11] to GenBank IDs as well as the FPRs predicted with geno2pheno[coreceptor][12] can be found in Table S3.

Additional References:

1. Mulinge M, Lemaire M, Servais J-Y, et al. HIV-1 Tropism Determination Using a Phenotypic Env Recombinant Viral Assay Highlights Overestimation of CXCR4-Usage by Genotypic Prediction Algorithms for CRRF01_AE and CRF02_AG. Chen Z, ed. *PloS one*. 2013;8(5):e60566.

2. Rolland M, Edlefsen PT, Larsen BB, et al. Increased HIV-1 vaccine efficacy against viruses with genetic signatures in Env V2. *Nature*. 2012;490(7420):417–20.

3. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology*. 2010;59(3):307–21.

4. Chevenet F, Brun C, Bañuls A-L, Jacq B, Christen R. TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC bioinformatics*. 2006;7(1):439. Available at: http://www.biomedcentral.com/1471-2105/7/439.

5. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*. 2004;32(5):1792–1797.

6. Kuiken C, Foley B, Leitner T, et al. HIV Sequence Compendium 2010. *Eds. Published by Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, NM, LA-UR* 10-03684, 2010.

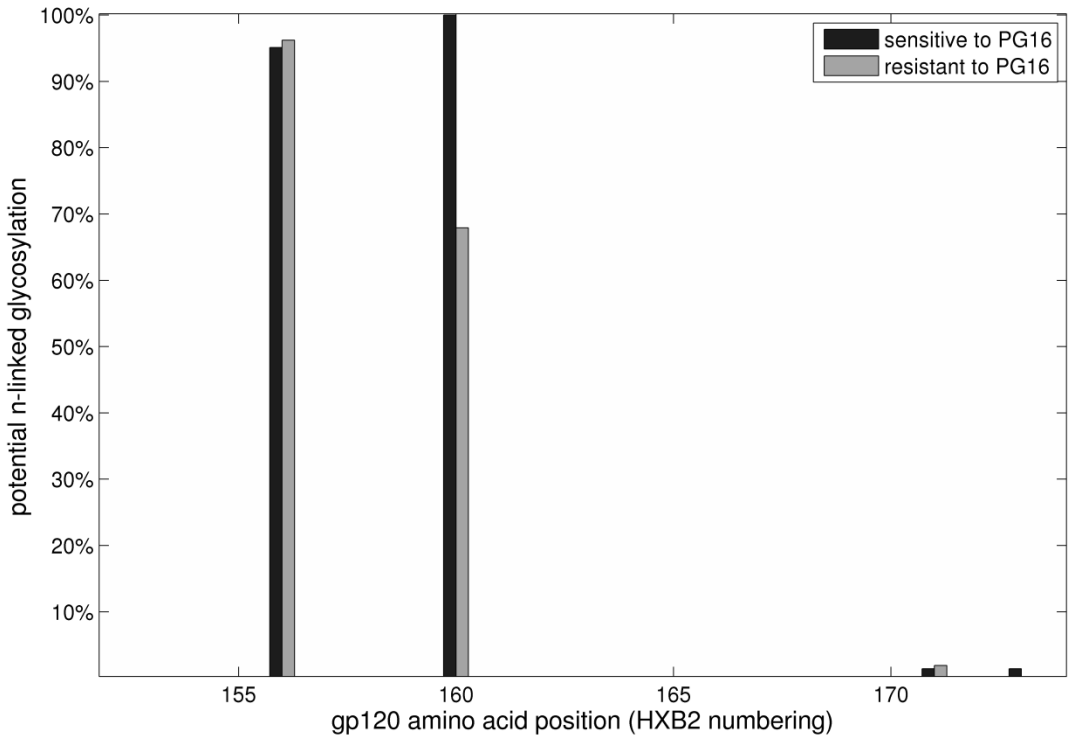

Fig. S1 Differences in glycosylation between different HIV variants regarding sensitivity to PG16. All viruses sensitive to PG16 had a potential N-linked glycosylation site at position 160 (black), while this was not the case for viruses resistant to PG16 (grey).

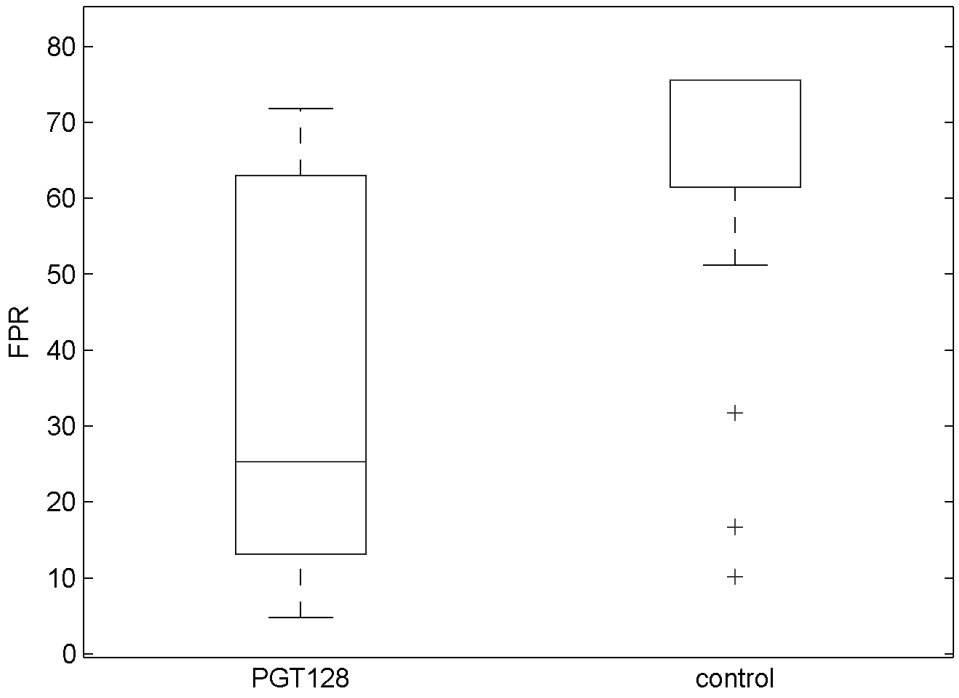

Fig. S2 Boxplot of geno2pheno[coreceptor] FPRs of the *Env* sequences of the HIV variants extracted from the PGT128-treated mice as well as the control-mice.

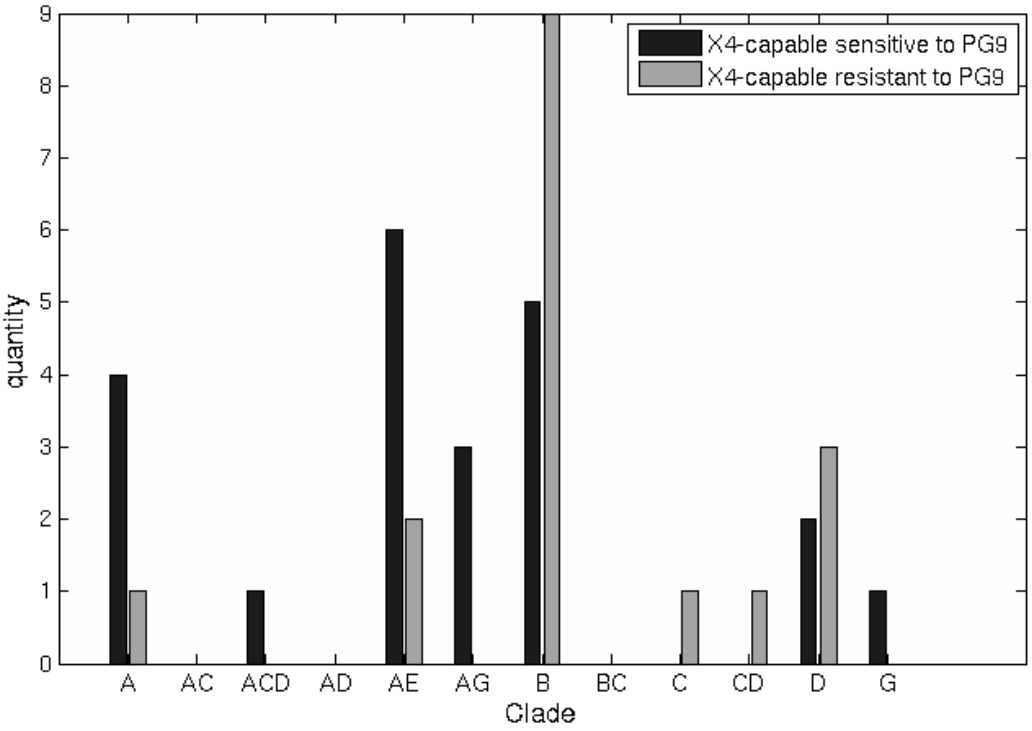

Fig. S3 Counts of X4-capable variants with respect to PG9 resistance and clade.

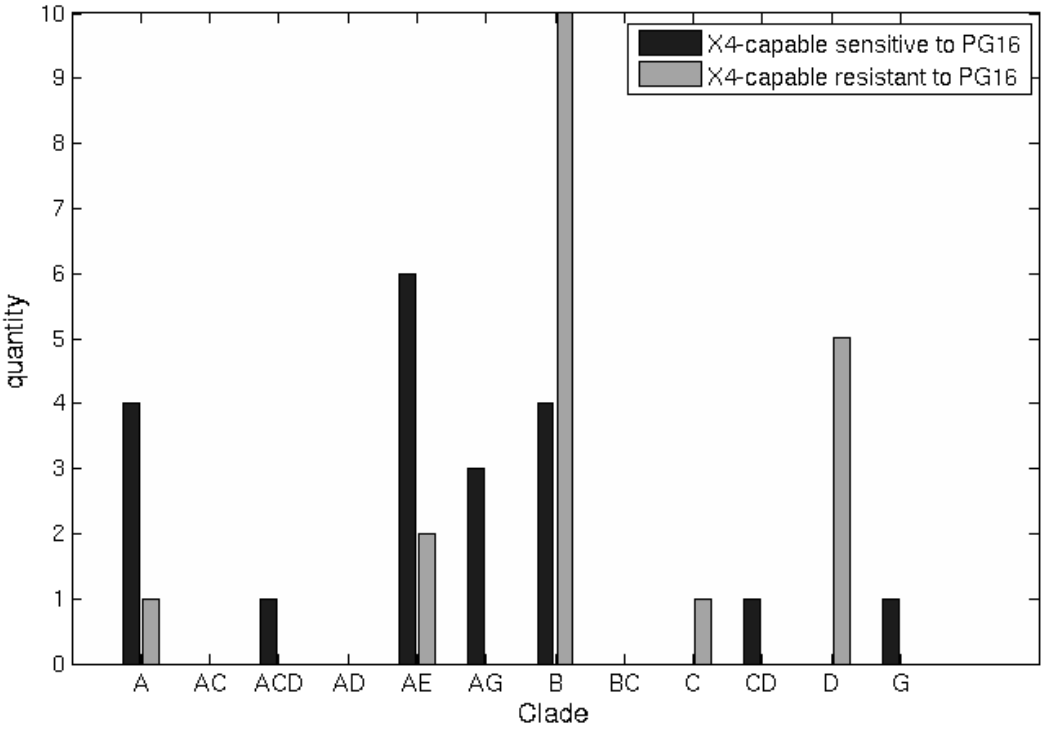

Fig. S4 Counts of X4-capable variants with respect to PG16 resistance and clade.

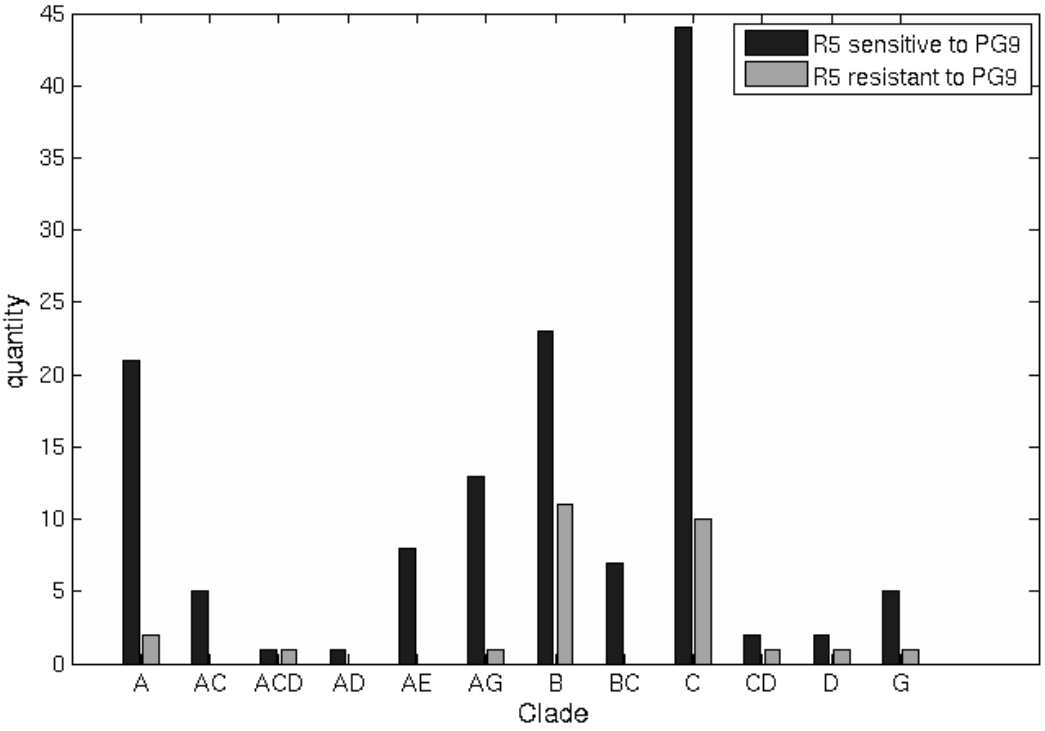

Fig. S5 Counts of R5 variants with respect to PG9 resistance and clade.

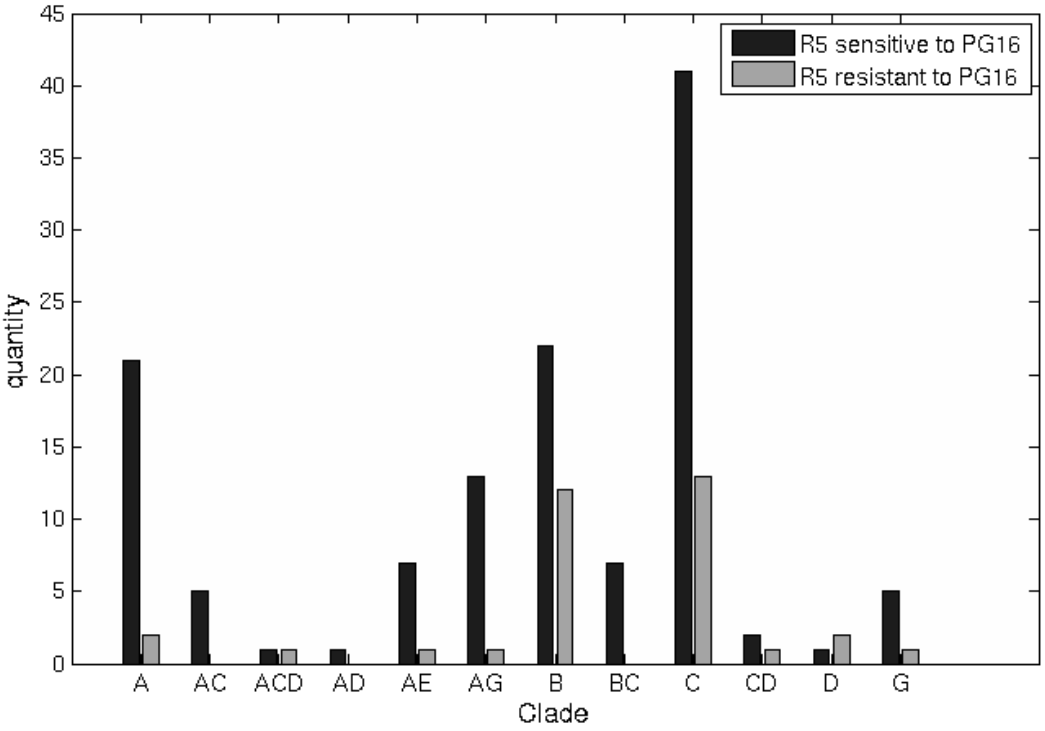

Fig. S6 Counts of R5 variants with respect to PG16 resistance and clade.

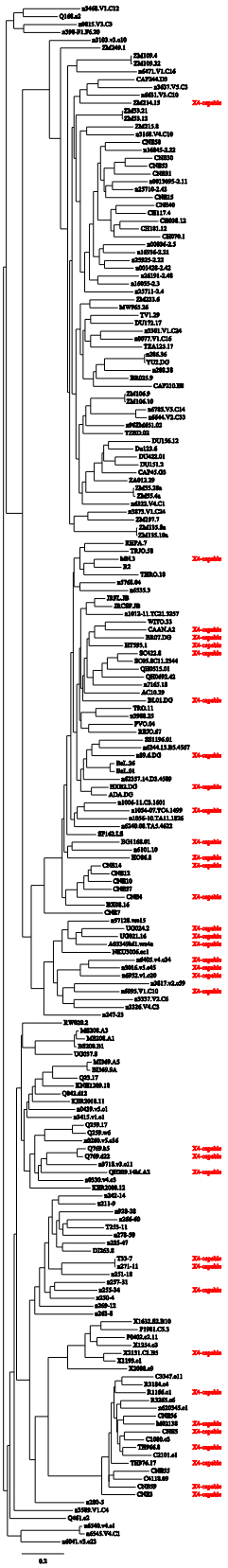

Fig. S7 Phylogenetic tree with coreceptor usage labels

Table S1. Coreceptor usage of HIV variants resistant and susceptible to four different broadly neutralizing antibodies according to the 5%/15% FPR cutoff.

antibody		X4-capable	R5	p-value
VRC01	resistant	0	17	0.2242
	sensitive	17	137	
VRC-PG04	resistant	2	32	0.5296
	sensitive	15	122	
PG9	resistant	9	28	0.0028
	sensitive	8	126	
PG16	resistant	11	33	0.0004
	sensitive	6	121	

This table shows the number of resistant and sensitive HIV variants with regard to their coreceptor usage for VRC01, VRC-PG04, PG9, and PG16. Variants were considered resistant to an antibody if the IC50 value was larger than 50 µg/ml and sensitive otherwise.

Table S2. Tropism information for variants with available information from the Los Alamos HIV data base[6].

Panel ID	GenBank ID	syncytia induction (MT2 T cell line)	tropism	geno2pheno FPR	PG9 IC50	PG16 IC50
KER2008.12	AY736809	SI	CCR5 CXCR4	12.5	0.017	0.006
KER2018.11	AY736810	NSI	CCR5	71	0.001	0.0006
KNH1209.18	AY736813	NSI	CCR5	94.5	0.367	0.678
RW020.2	EU855131	NSI	CCR5	26	0.103	0.07
M02138	AY713424	SI	CXCR4	0.6	0.122	0.022
TH966.8	U08456	NSI	CCR5	7.9	0.042	0.008
DJ263.8	AF063223	NSI	CCR5	56.9	0.1	0.048
HT593.1	U08444	SI	CCR5 CXCR4	1	0.271	0.153
BR025.9	U15121	NSI	CCR5	38.4	0.044	0.009
MW965.26	U08455	NSI	CCR5	73.4	1.99	0.961
HXB2.DG	K03455	SI	CXCR4	0	0.553	> 50
UG024.2	U43386	SI	CXCR4	0	3.94	> 50
NKU3006.ec1	AY736835	NSI	CCR5	42.6	> 50	> 50
57128.vrc15	AY736829	NSI	CCR5	24.6	0.104	0.162
UG021.16	U27399	SI	CXCR4	0	> 53	> 50

This table shows the results from the phenotypic assays available for 15 samples from the 199-isolate panel as well as the geno2pheno[coreceptor] FPR that can be used to predict the tropism. Furthermore, the IC50 values for PG9 and PG16 binding are shown. It can be seen that for the only two discordant tropism assignments (KER2008.12 and TH966.8) the IC50 values are all very small meaning that those variants are sensitive to both antibodies.

Table S3. Mapping of the IDs from Doria-Rose et al.[11] to GenBank IDs together with the FPRs predicted with geno2pheno[coreceptor][12]

Panel ID	GenBank ID	geno2pheno[coreceptor] FPR
0260.v5.c36	HM215256	89
0330.v4.c3	HM215257	89
3415.v1.c1	HM215299	90
BB201.B42	DQ187171	14
BB539.2B13	DQ187240	4.6
BI369.9A	DQ187019	43.8
BS208.B1	DQ187023	78.8
KER2008.12	AY736809	12.5
KER2018.11	AY736810	71
KNH1209.18	AY736813	94.5
MB539.2D1	DQ187269	5.8
MI369.A5	DQ187018	32.9
MS208.A1	DQ187010	83.5
MS208.A3	DQ187022	83.5
Q168.a2	AF407148	13.5
Q23.17	AF004885	44
Q259.w6	AF407151	32.6
Q461.e2	AF407156	49.7
Q769.d22	AF407158	7.4
Q769.h5	AF407159	7
Q842.d12	AF407160	94.5
RW020.2	EU855131	26
UG037.8	U09127	84.9
3301.V1.C24	HM215294	48.7
6041.v3.c23	HM215321	84.9
3103.v3.c10	HM215288	82.4
6095.V1.C10	HM215323	2.5
3468.V1.C12	HM215301	21
C1080.c3	AY945712	35.5
C2101.c1	AY945716	56.9
C4118.09	AY945722	79.9
CNE5	HM215415	5.9
CNE55	HM215418	10
CNE59	HM215422	6
M02138	AY713424	0.6
R1166.c1	AY945728	9
R2184.c4	AY945730	19
R3265.c6	AY945732	24

TH966.8	U08456	7.9
211-9	EU513187	87.5
235-47	EU513195	83.5
255-34	EU513184	1.8
257-31	EU513185	90.9
263-8	EU513182	72
271-11	EU513197	8.5
280-5	EU513183	23.4
928-28	EU513199	68
DJ263.8	AF063223	56.9
T253-11	EU513191	57
T33-7	EU513186	8.5
3988.25	AY835436	41
5768.04	AY835435	15
1012-11.TC21.3257	EU289184	48.9
1056-10.TA11.1826	EU289186	28.4
ADA.DG	AY426119	24.7
BaL.01	DQ318210	44.4
BaL.26	DQ318211	24.7
BX08.16	GQ855765	35
CAAN.A2	AY835452	6.9
CNE10	HM215397	17
HT593.1	U08444	1
JRCSF.JB	AY669726	31.7
PVO.04	AY835444	42
REJO.67	AY835449	48.9
SC05.8C11.2344	EU289200	33.9
SC422.8	AY835441	8.5
SS1196.01	AY835442	24.6
TRJO.58	AY835450	24
TRO.11	AY835445	57
WITO.33	AY835451	30
YU2.DG	M93258	75.6
CH038.12	EF042692	52.5
CH070.1	EF117255	10.5
CH117.4	EF117262	63
CH181.12	EF117259	69.8
CNE15	HM215401	48.7
CNE40	HM215414	95.5
CNE7	HM215426	31.4
286.36	JQ362420	78

288.38	JQ362421	95.5
0013095-2.11	EF117267	50.9
001428-2.42	EF117266	94.5
25710-2.43	EF117271	54.4
25711-2.4	EF117272	75.6
25925-2.22	EF117273	88.5
26191-2.48	EF117274	64
3168.V4.C10	HM215289	86
6644.V2.C33	HM215336	73
6785.V5.C14	HM215338	73
BR025.9	U15121	38.4
CAP244.D3	DQ435684	78.8
CAP45.G3	DQ435682	22.4
CNE31	HM215412	90
CNE58	HM215421	74.4
DU151.2	DQ411851	91.7
DU156.12	DQ411852	84.9
MW965.26	U08455	73.4
TZBD.02	JQ362424	63
ZA012.29	EU855133	52.5
ZM106.10	AY424164	48.6
ZM106.9	AY424163	48.6
ZM109.32	AY424141	29.4
ZM109.4	AY424138	21
ZM197.7	DQ388515	94.6
ZM233.6	DQ388517	77
ZM249.1	DQ388514	91
ZM53.12	AY423984	29.5
ZM53.21	AY423985	29.5
ZM55.4a	AY423973	58.7
P0402.c2.11	EU885759	51.8
P1981.C5.3	FJ817369	32.6
X1193.c1	EU885761	28.8
X1254.c3	EU885762	96
X1632.S2.B10	FJ817370	49.9
X2131.C1.B5	FJ817368	8.5
266-60	EU513193	97
6535.3	AY835438	64
1006-11.C3.1601	EU289183	22
6240.08.TA5.4622	EU289190	51.5
HXB2.DG	K03455	0

CNE53	HM215417	83
ZM215.8	DQ422948	38
ZM55.28a	AY423971	58.7
3016.v5.c45	HM215283	3.8
UG024.2	U43386	0
251-18	EU513196	16
RHPA.7	AY835447	17.8
3873.V1.C24	HM215311	78.7
6952.v1.c20	HM215343	1.7
3718.v3.c11	HM215306	88.8
Q259.17	AF407152	35
3589.V1.C4	HM215304	67.6
C3347.c11	AY945721	83
CNE3	HM215410	8.5
269-12	EU513194	43.6
AC10.29	AY835446	24
THRO.18	AY835448	17.8
0077.V1.C16	HM215254	15.4
16055-2.3	EF117268	83
16845-2.22	EF117269	69.8
Du123.6	DQ411850	35
3326.V4.C3	HM215296	80
DU172.17	DQ411853	95.8
3817.v2.c59	HM215310	97.8
0439.v5.c1	HM215258	97
398-F1.F6.20	HM215312	77
QH209.14M.A2	FJ866118	7.4
0815.V3.C3	HM215260	96.7
CNE56	HM215419	0
TH976.17	U08458	6.9
1054-07.TC4.1499	EU289185	7.9
6101.1	AY835434	17.8
62357.14.D3.4589	EU289189	72.7
6244.13.B5.4567	EU289191	68.6
89.6.DG	U39362	0
BG1168.01	AY835443	5
BR07.DG	AY124979	9.6
CNE12	HM215399	40
CNE4	HM215413	4
CNE57	HM215420	0.5
JRFL.JB	U63632	24.7

QH0515.01	AY835440	21
QH0692.42	AY835439	38.4
R2	AF128126	64
SF162.LS	EU123924	42.7
16936-2.21	EF117270	77
96ZM651.02	AF286224	98.4
CNE30	HM215411	67.8
ZM214.15	DQ388516	8.5
3337.V2.C6	HM215297	89
6405.v4.c34	HM215327	0.5
A03349M1.vrc4a	HM215356	4
NKU3006.ec1	AY736835	42.6
00836-2.5	EF117265	79.5
247-23	EU683891	99
6545.V4.C1	HM215332	99
TV1.29	EU855132	30
HO86.8	EF210732	5
242-14	EU513188	78
6540.v4.c1	HM215330	95
CAP210.E8	DQ435683	48.7
57128.vrc15	AY736829	24.6
TZA125.17	JQ362423	80.8
DU422.01	DQ411854	89
278-50	EU513198	80
250-4	EU513189	66.4
7165.18	AY835437	18.9
620345.c1	JQ362422	13
CNE14	HM215400	8
MN.3	HM215430	0
3637.V5.C3	HM215305	68
ZM135.10a	AY424079	63
ZM135.8a	AY424077	63
UG021.16	U27399	0
BL01.DG	AY124970	6
6322.V4.C1	HM215326	20.9
6471.V1.C16	HM215328	92
6631.V3.C10	HM215335	87
X2088.c9	EU885764	27