Supplemental Material

Figure S1. Alignment of the CD4 D1 domain for human, gorilla, and all chimpanzee alleles.

Figure S2. The glycosylation state of CD4 is recapitulated in T cells.

Figure S3. CD4 and CCR5 expression in stable cell lines utilized.

Table S1. Source of chimpanzee CD4 sequences.

Table S2. Polymorphic sites in chimpanzee CD4.

```
                   1        10        20        30        40        50        60
                   ●        ●         ●         ●         ●         ●         ●
Human              KKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRS
Gorilla            NKVVLGKKGDTVELNCTASQKKSIQFHWKNSNQMKILGNQGSFLTKGPSKLSDRADSRRS
Chimp_Allele1      KKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQTKILGNQGSFLTKGPSKLNDRIDSRRS
Chimp_Allele2      KKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQTKILGNQGSFLTKGPSKLNDRVDSRRS
Chimp_Allele3      KKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQTKILGNRGSFLTKGPSKLNDRIDSRRS
Chimp_Allele4      KKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQTKILGNQGSFLTKGPSKLKDRVDSRRS
Chimp_Allele5      KKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQTKILGNQGSFLTKGPSKLKDRVDSRRS
Chimp_Allele6      KKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQTKILGNQGSFLTKGPSKLNDRVDSRRS
Chimp_Allele7      KKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQTKILGNQGSFLTKGPSKLNDRVDSRRS
Chimp_Allele8      KKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQTKILGNQGSFLTKGPSKLNDRIDSRRS
                   :*************.****************** *****.***********.** *****

                            70        80        90      98
                            ●         ●         ●       ●
Human              LWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVF
Gorilla            LWDQGNFPLIIKNLKIEDSDTYICEVEGQKEEVQLLVF
Chimp_Allele1      LWDQGNFPLIIKNLKIEDSDTYICEVGDQKEEVQLLVF
Chimp_Allele2      LWDQGNFPLIIKNLKIEDSDTYICEVGDQKEEVQLLVF
Chimp_Allele3      LWDQGNFPLIIKNLKIEDSDTYICEVGDQKEEVQLLVF
Chimp_Allele4      LWDQGNFTLIIKNLKIEDSDTYICEVGDQKEEVQLLVF
Chimp_Allele5      LWDQGNFPLIIKNLKIEDSDTYICEVGDQKEEVQLLVF
Chimp_Allele6      LWDQGNFTLIIKNLKIEDSDTYICEVGDQKEEVQLLVF
Chimp_Allele7      LWDQGNFTLIIKNLKIEDSDTYICEVGDQKEEVQLLVF
Chimp_Allele8      LWDQGNFPLIIKNLKIEDSDTYICEVGDQKEEVQLLVF
                   *******.****************** .**********
```

Figure S1. Alignment of the CD4 D1 domain for human, gorilla, and all chimpanzee alleles. Bolded sites indicate predicted N-glycosylation motifs.

www.pnas.org/cgi/doi/10.1073/pnas.1813909116

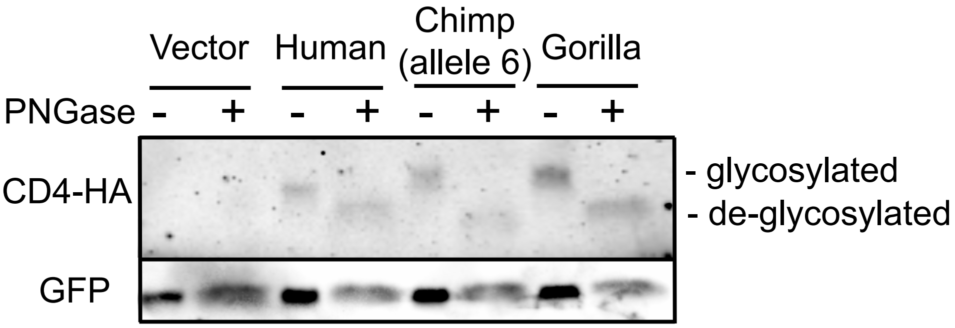

Figure S2. The glycosylation state of CD4 is recapitulated in T cells. 2x10^6 human CD4+ T cells (Hut78 cells) were nucleofected with 1 μg of the indicated HA-tagged CD4 expression constructs or an empty vector control, and 1 μg turbo GFP which served as a marker of nucleofection efficiency. Total cell lysates were harvested 48 hours post nucleofection and split into two reactions, one of which was treated with PNGase F, and then the remaining lysate (~5 μg) was loaded onto a 10% denaturing SDS-PAGE gel. Full-length CD4s were predicted to encode the following number of glycans: Human (2), Chimpanzee allele 6 (4), and Gorilla (3). This is in contrast to the sCD4 molecules shown in Figure 2B, which only encode the D1-D2 domains of CD4 and are predicted to encode 0, 2, or 1 glycans for human, chimpanzee, or gorilla, respectively. Anti-HA western blotting confirmed that CD4 is differentially glycosylated when expressed in human T cells. Further, anti-GFP western blotting confirmed efficient nucleofection, however, the weak signal for HA-CD4 may reflect poor expression of exogenous CD4 in cell lines that constitutively express CD4 natively.

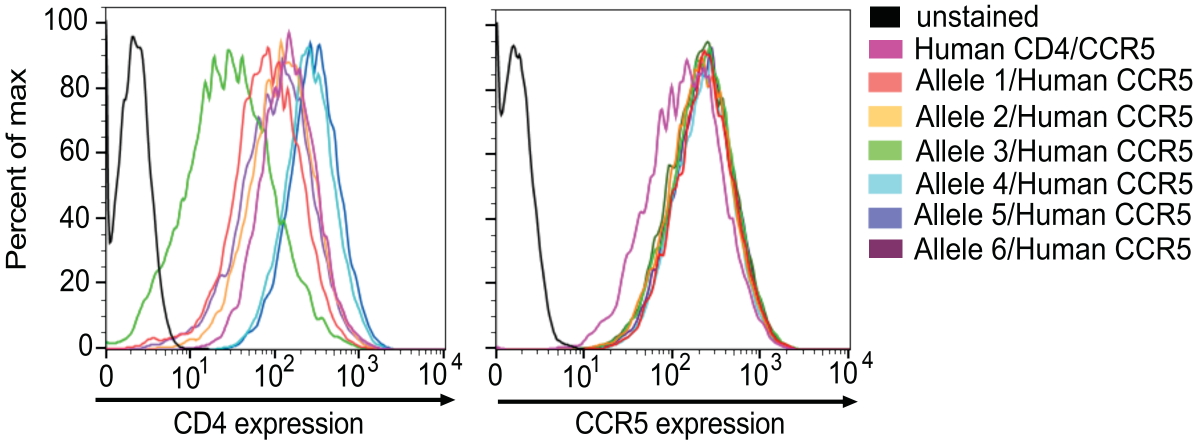

Figure S3. CD4 and CCR5 expression in stable cell lines utilized. Histograms of CD4 and CCR5 signal in Cf2th cell lines stably expressing the indicated CD4s (left) and human CCR5 (right).

Table S1. Sources for chimpanzee CD4 sequences.

	Number of individuals			Total # of unique DNA haplotypes	Total # of unique protein alleles
	This study[1]	Great Apes Genome Project[2]	Hvilsom et al.[3]		
P.t. ellioti	0	10	6	2	2
P.t. verus	2	4	6	2	1
P.t. schweinfurthii	1	6	2	5	4
P.t. troglodytes	1	4	8	9	7
Total	**4**	**24**	**22**	**14**	**8**

[1] SNPs were identified by sequencing additional chimpanzee CD4 individuals (this study).

[2] SNPs were identified in sequence data extracted from variant calls produced in Prado-Martinez et. al., 2013, whole-genome great ape sequencing survey (1). Because these were extracted from a whole-genome sequencing project, these CD4 sequences are newly reported in the present study.

[3] Data previously deposited Hvilsom et. al., 2008, (2).

Table S2. Polymorphic sites in chimpanzee CD4 (the final column corresponds to allele numbers used in manuscript)

nucleotide position	38	51	194	231	238	277	891	1056	1078	1218	DNA haplotype		Protein haplotype	
nucleotide residue ref[1]/SNP	t/c	a/g	a/g	t/a	g/a	a/c	c/t	a/g	g/a	c/g				
amino acid position	-13	-8	40	52	55	68	272	327	334	381				
amino acid residue ref[1]/SNP	V/A	A/A	Q/R	N/K	V/I	T/P	L/L	S/S	V/M	A/A				
Gene location	Signal peptide	Signal peptide	D1 domain	D1 domain	D1 domain	D1 domain	D3 domain	D4 domain	D4 domain	D4 domain				
P.t. troglodytes Hap1 (n=4)	V	A	Q	N	I	P	L	S	V	A	taataccggc	1.1	VQNIPV	1
P.t. troglodytes Hap2 (n=7)	V	A	Q	N	V	P	L	S	V	A	tgatgctggc	2.1	VQNVPV	2
P.t. troglodytes Hap3 (n=3)	V	A	Q	N	I	P	L	S	V	A	tgataccggc	1.2	VQNIPV	1
P.t. troglodytes Hap4 (n=1)	V	A	R	N	I	P	L	S	V	A	tggtaccggc	3	VRNIPV	3
P.t. troglodytes Hap5 (n=1)	V	A	Q	K	V	P	L	S	V	A	taaagctggc	5	VQKVPV	5
P.t. troglodytes Hap6 (n=2)	V	A	Q	N	I	P	L	S	M	A	tgataccgac	8	VQNIPM	8
P.t. troglodytes Hap7 (n=4)	V	A	Q	K	V	T	L	S	V	A	taaagacggc	4.1	VQKVTV	4
P.t. troglodytes Hap8 (n=2)	V	A	Q	N	V	T	L	S	V	A	taatgacggc	6.1	VQNVTV	6
P.t. troglodytes Hap9 (n=2)	V	A	Q	K	V	T	L	S	V	A	taaagacggg	4.2	VQKVTV	4
Source of identified SNP[2]	-	2,3	2	3	2,3	2,3	2,3	-	2,3	3				
P.t. schweinfurthii Hap1 (n=9)	V	A	Q	N	V	P	L	S	V	A	taatgccggc	2.2	VQNVPV	2
P.t. schweinfurthii Hap2 (n=5)	V	A	Q	N	V	P	L	S	V	A	tgatgccggc	2.3	VQNVPV	2
P.t. schweinfurthii Hap3 (n=2)	V	A	Q	K	V	T	L	S	V	A	taaagacggc	4.1	VQKVTV	4
P.t. schweinfurthii Hap4 (n=1)	A	A	Q	N	V	T	L	S	V	A	caatgacggc	7	AQNVTV	7
P.t. schweinfurthii Hap5 (n=1)	V	A	Q	K	V	T	L	S	V	A	taaagatggc	4.3	VQKVTV	4
Source of identified SNP[2]	4	2	-	2-4	-	2,3	3	-	-	-				
P.t. ellioti Hap1 (n=18)	V	A	Q	K	V	T	L	S	V	A	taaagacggc	4.1	VQKVTV	4
P.t. ellioti Hap2 (n=14)	V	A	Q	N	V	P	L	S	V	A	taatgccggc	2.2	VQNVPV	2
Source of identified SNP[2]	-	-	-	2,3	-	2,3	-	-	-	-				
P.t. verus Hap1 (n=1)	V	A	Q	N	V	T	L	S	V	A	taatgacagc	6.2	VQNVTV	6
P.t. verus Hap1 (n=23)	V	A	Q	N	V	T	L	S	V	A	taatgacggc	6.1	VQNVTV	6
Source of identified SNP[2]	-	-	-	-	-	-	-	2	-	-				

[1] Pan troglodytes CD4 (NM_001009043.1) served as the reference sequence for this dataset.

[2] SNPs were identified in sequence data extracted from variant calls produced in Prado-Martinez et. al., 2013, whole-genome great ape sequencing survey (1). Because these were extracted from a whole-genome sequencing project, these CD4 sequences are newly reported in the present study.

[3] SNPs were identified in sequence data deposited Hvilsom et. al., 2008, (2).

[4] SNPs were identified by sequencing additional chimpanzee CD4 individuals (this study).

SI Appendix references

1. Prado-Martinez J, et al. (2013) Great ape genetic diversity and population history. *Nature* 499(7459):471–475.

2. Hvilsom C, et al. (2008) Genetic subspecies diversity of the chimpanzee CD4 virus-receptor gene. *Genomics* 92(5):322–328.