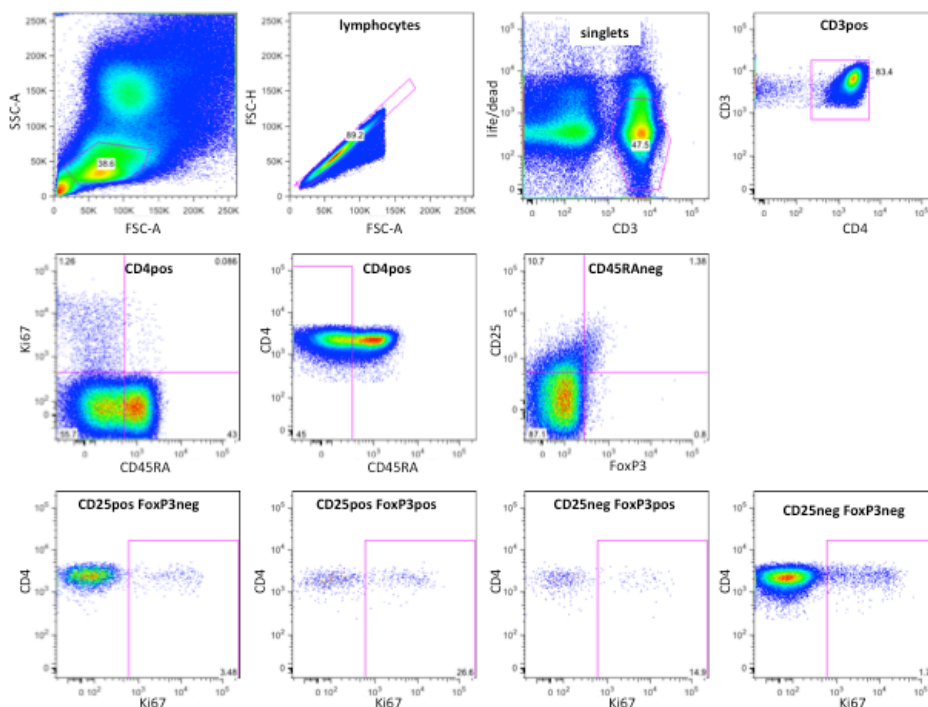


# 1 Supporting Information

supplementary figure 1



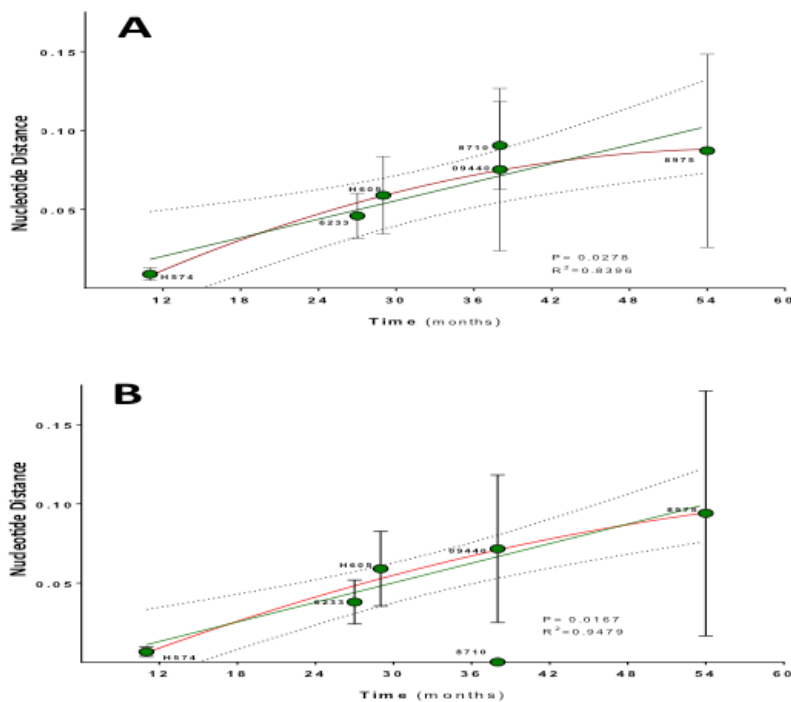
2

3 **S1 Fig: Representative gating tree for analyses of Ki67 expression in memory**

4 **CD4 T cell populations delineated by CD25 and FoxP3 expression in the**

5 **HHECO cohort**

6



7

8 **S2 Fig: Plasma- and cell-derived EnvV1V3 nucleotide sequence variability**

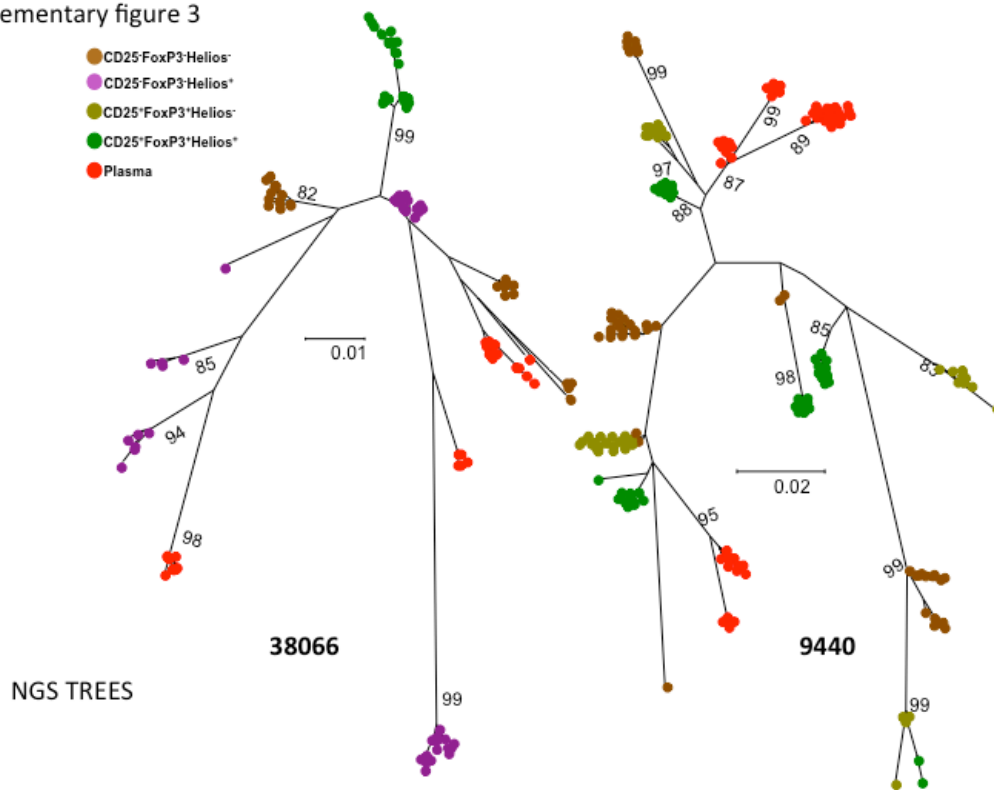
9 **increases with HIV infection duration within and between these compartments.**

10 Linear regression analysis (green line) was performed using the Prism/GraphPad  
 11 software package and P values were calculated with the Pearson two-tailed statistical  
 12 test. The red line shows a second order polynomial regression analysis indicating that  
 13 nucleotide variation may be reaching a plateau. A) Distance of the EnvV1V3  
 14 sequences derived from plasma to the sequences extracted from the corresponding  
 15 cellular fractions plotted against the estimated duration of infection. B) Plasma  
 16 sequences diversity plotted against the estimated duration of infection. The red line  
 17 indicates a non-linear analysis performed using a second order polynomial equation  
 18 taking into account the best-fit values. The evolutionary distances were computed  
 19 using the Kimura 2-parameter method (75) and are in the units of the number of base  
 20 substitutions per site including both Transitions + Transversions. The rate variation

21 among sites was modelled with a gamma distribution. The analysis was conducted in  
22 MEGA6 (70). No sequence diversity was observed in the 8710 plasma fraction and  
23 thus was not included in the linear regression analysis with the P value been not  
24 significant if 8710 had been included.

25

Supplementary figure 3



27

28 **S3 Fig: Phylogenetic analyses of HIV Envelope sequences derived from plasma**  
 29 **and sorted memory CD4 T cell populations using a using Next Generation**  
 30 **sequencing.**

31 Shown is the phylogenetic analyses of EnvV1V3 sequences from the 50 most  
 32 frequently detected sequences derived from either plasma or the different sorted  
 33 memory CD4 T cell subsets for two viremic subjects of the WHIS cohort. The  
 34 phylogenetic relationship was inferred by the Maximum Likelihood method based on  
 35 the General Time Reversible substitution model (GTR+G). EnvV1V3 amplicons were  
 36 directly subjected to next generation sequencing. Quasi-species reconstruction was  
 37 performed using the software QuasiRecomb. The applied methods are described in  
 38 detail in the material and methods section.

39