Supplementary Material and Methods

Materials and Methods:

Study cohorts

CAPRISA 004. Centre for the AIDS Programme of Research in South Africa (CAPRISA) 004 was a Phase IIb, double-blind, randomized, placebo-controlled study, assessing the effectiveness and the safety of a vaginal gel containing 1% tenofovir for the prevention of HIV infection in South African women (clinicaltrials.gov: NCT00441298). A total of 889 sexually active, HIV-uninfected women aged 18-40 years from urban and rural settings in South Africa participated in the CAPRISA 004 study between May 2007 and March 2010*(41)*. Participants with HIV infection at enrolment and those who became HIV infected during follow-up were referred to the CAPRISA 002 Acute HIV infection study*(42)*. Plasma, peripheral blood mononuclear cells (PBMC), and cervicovaginal lavage (CVL) samples were collected and cryopreserved. This was a matched case: control (1:2) study with a total of 270 participants. Pre-infection samples had a sampling median of 110 days (IQR 65 -182) pre-infection. Markers of microbial translocation were measured in cases at all available HIV positive time-points pre-ART initiation.

RV254/SEARCH010. The RV254/ South East Asia Research Collaboration in HIV (SEARCH) 010 (clinicaltrials.gov: NCT00796146) is a study conducted at the Thai Red Cross AIDS Research Centre, identifying acute HIV infection (AHI) and offering immediate ART to participants. Plasma samples from participants with AHI were screened using the 4th generation immunoassay and pooled nucleic acid testing (NAT)*(43)*. Participants identified with AHI were then offered enrollment into RV254/ SEARCH010 and following enrollment, a series of HIV clinical Ab, Ag, and Western Blot assays were performed in parallel for categorizing AHI using the Fiebig (F) and fourth-generation (4thG) staging systems *(30, 44)*. Aliquots of plasma and peripheral blood mononuclear cells (PBMC) were cryopreserved in cryovials under liquid nitrogen. Optional sampling of gut-associated lymphoid tissue (GALT) was obtained by sigmoid biopsies obtained at baseline and 6 months. The study was approved by the institutional review boards of Chulalongkorn University, Thailand and the Walter Reed Army Institute of Research, US Army, United States. Plasma, PBMC and MMC from HIV-uninfected and chronically infected (CHI) Thai volunteers were obtained from another protocol (clinicaltrials.gov NCT01397669) and were subject to the same procedures.

***Nairobi/Uganda based studies*.** Participants were recruited for the study from existing female sex worker (FSW) cohorts located in Nairobi, Kenya. Each participant was tested for HIV-1 serology, peripheral CD4 counts, bacterial vaginosis (BV) by Gram stain, *Chlamydia trachomatis* and *Neisseria gonorrhea* by PCR (Roche), and *Trichomonas vaginalis* by InPouch. Cervical mononuclear cells (CMCs) were collected and used for flow cytometry assays. The study was

approved by the Institutional Review Boards of Kenyatta National Hospital (Nairobi, Kenya) and St Michael's Hospital (Toronto, ON, Canada), as well as the Universities of Manitoba and Toronto (Canada). Informed written consent was obtained from each participant prior to enrolment.

Sample collection and processing

Peripheral blood was collected in ACD tubes. Plasma was collected and frozen at -70°C until further analysis. Peripheral blood mononuclear cells (PBMCs) were isolated by Ficoll-Hypaque density gradient centrifugation and cryopreserved in cryovials in media containing 10% dimethyl sulfoxide in fetal bovine serum overnight at 80°C before transfer to liquid nitrogen for long-term storage. Cervicovaginal lavage (CVL) samples were collected by rinsing the endocervix with 2ml of sterile 1x phosphate-buffered saline (PBS) and the lavage was collected from the posterior fornix. Samples were briefly centrifuged to remove cellular debris and stored at -70°C until further analysis. Cervical mononuclear cells (CMCs) were obtained from the endocervical cytobrush and filtered through a 100-mm filter (Becton Dickinson, BD). Tubes were vortexed several times to remove as many cells as possible from each cytobrush. After filtering, cells were washed in warm RPMI 1640 (10% FBS and 1% antibiotic/antimycotic) and used for flow cytometry analysis. Gut-associated lymphoid tissue (GALT) sampling involved the collection of 12-20 sigmoid pinch biopsies at baseline, 6 and 12 months post ART initiation in RV 254/SEARCH 010 study. Aliquots of the biopsy tissues were cryopreserved, embedded in paraffin and other aliquots subjected to isolation of mucosal mononuclear cells (MMC) and used for flow cytometry analysis.

Flow cytometry analysis

A total of 10^6 cells were suspended in fluorescence-activated cell sorting (FACS) buffer consisting of 1× phosphate-buffered saline (pH 7.2) containing 1% fetal bovine serum (Sigma-Aldrich). The cells were stained with premixed combinations of fluorochrome-labeled mAbs at titration-determined optimal concentrations. We optimized a flow cytometry panel to characterize several attributes of $CD4^+$ T cells *ex vivo* (Supplemental Figure 1). These included the activation status (CD38, HLA-DR, and Ki67), memory phenotype (CD27, CD45RA, and CCR7), and target cell (CCR5, integrin β_7) characteristics. Integrin β_7^{Hi} CD45RA- cells, consistent with previous reports, are characterized as $\alpha_4\beta_7$ cells*(20)*. All antibodies were purchased from BD Biosciences unless otherwise stated. To discriminate between dead and live cells, samples were stained with LIVE/DEAD® Fixable Aqua Dead Cell Stain Kit, for 405 nm excitation (L34957, Invitrogen). Primary labeled mAbs used were PE-Cy5-HLA-DR (555813), FITC-CD38 (555459), PE-Cy7-CD27 (560609), PE-CF594-CCR7 (562381), BV711-CD45RA (563733), BV421-CCR5 (562576), PE-Integrin β_7 (555945), BV650-CD8 (563821), APCH7-CD3 (560176), BV605-CD4 (562658) and AlexaFlour 647- Ki67 (558615). Intracellular staining for Ki67 was performed using a BD Cytofix/Cytoperm fixation/permeabilization method (554722). Representative

gating strategies are shown in Fig. 1A and Fig. S5.

Soluble biomarker analysis
Plasma I-FABP and sCd14 were measured using Human FABP2/I-FABP Quantikine ELISA, DFBP20 and Human sCD14 Quantikine ELISA, DC140, (R&D Systems Inc., Minneapolis, USA). All assays were performed following manufacturer's instructions. Samples with values below the lower detection limit were assigned the value half the lower limit of quantification, LLOQ/2.

Viral sequencing
Viral RNA was isolated from plasma (N=27), or from supernatants of PBMC cultures infected with plasma-derived virions (N=8). Briefly, for these 8 samples, virions were isolated from plasma samples using the μMACS VitalVirus HIV Isolation Kit (Miltenyi Biotec), and then spinoculated at 25°C for 60 minutes at 1200 x g on to pooled PBMCs that were previously stimulated for 2 days with either PHA at 0.5 ug/mL, PHA at 5 ug/mL, or an anti-CD3 Ab (OKT3) at 10 ug/mL. Infected cultures were maintained at 37°C in the presence of IL-2 for 25 days with IL-2-containing medium being refreshed every three days and cells being replenished every seven days, and viral RNA was then extracted. Importantly, for 23 participants, single genome sequencing was performed directly from plasma as well from this cell culture-based approach, and in all cases the $\alpha_4\beta_7$-binding motifs of the consensus sequences from both approaches were identical.

Non-human primates (NHPs). Data from 2 different NHP experiments are included in the manuscript. For the SIV acquisition analyses, 11 animals from the control arm of a published paper were studied, and the methods for these are described in more detail elsewhere*(18)*. A second cohort of animals (n=14) that were part of another study was infected intravenously with SIVmac239 in the absence of any intervention, and were followed over time to characterize disease progression markers, as described *(47)*. Polychromatic flow cytometry in NHPs was performed in a similar fashion as in humans and as previously described *(18)*. Briefly, heparinized blood was used to isolate mononuclear cells using standard Ficol-hypaque gradient centrifugation. Following washes in RPMI1640 media, one million cells were dispensed into 4 individual test tubes and each incubated with a cocktail of monoclonal antibodies for 30 min at 4C. After 3 washes in RPMI 1640 media, the cells were re-suspended in PBS containing 0.1% sodium azide and analyzed using a B-D Fortessa flow cytometer. A minimum number of 100,000 events were analyzed per sample and the data obtained were analyzed using FlowJo 9.9.4 software (Ashland, OR). The animals were maintained according to the guidelines of the Committee on the Care and Use of Laboratory Animals of the Institute of Laboratory Animal Resources, National Research Council and the Department of Health and Human Services. Animals were born and housed at Yerkes National Primate Research Centre (Emory University, Atlanta, GA) that has been fully accredited by the Association

for Assessment and Accreditation of Laboratory Animal Care International.

Table S1. Characteristics of the Ugandan and Kenyan study populations

Variable	**Uganda (n=39)**	**Kenya (n=18)**
Age (median, IQR)	29 (24-34)	30 (26-34)
Married, n/N (%)	25/39 (64)	n/a
Hormonal contraceptive users, n/N (%)	13/39 (34)	11/18 (61)
Vaginal discharge, n/N	n/a	0/18
Ulcers, n/N	0/39	0/18
HSV2 serostatus, n/N (%)	n/a	9/18 (50)
Parity (median, IQR)	3 (1.75, 5)	n/a

n/a = not available

Table S2. Cohort characteristics (RV254 study)

Characteristics		Acute HIV infected (n=36)	Chronic HIV infected (n=5)	HIV uninfected (n=9)
Age, median (IQR)		28 (19-48)	24 (19-28)	34 (23-40)
Gender (male:female)		33:3	5:0	7:2
Risk behavior, n	MSM	30	5	5
	Bisexual male	3	0	0
	Heterosexual male	0	0	2
	Heterosexual female	3	0	2
Fiebeg stage, n	I	8	n/a	n/a
	II	10	n/a	n/a
	III	18	n/a	n/a
Mean (SD) duration of HIV (days)			302 (162)	n/a
Median plasma HIV RNA (log_{10} copies/ml)		5.5 (3.3-7.7)	4.9 (4.0-5.4)	n/a
Median sigmoid colon HIV RNA (log_{10} copies/ml)		2.7 (1.3-6.5)	n/d	n/a
Median CD4 count (cells/mm^3)		438 (132-1127)	515 (316-883)	n/a
HIV subtype, n (%)	CRF01_AE	30 (83)	n/d	n/a
	CRF01_AE/B	4 (11)	n/d	n/a
	B	2 (6)	n/d	n/a

n/a = not available
n/d = not determined
SD = standard deviation
IQR = intraquartile range

Table S3. Cohort characteristics (Female Sex Worker Cohort, Nairobi, Kenya)

Variable	**Cases (n=10)**	**Controls (n=31)**
Age, years (median, IQR)	34 (30, 40)	34 (30, 38)
Duration of sex work, years (median, IQR)	5 (3, 5)	4 (2, 8)
Gravida (median, IQR)	3 (2, 4)	2 (1, 3)
Sex partners last week (median, IQR)	10 (4, 19)	7 (3, 18)
Regular clients per week (median, IQR)	6 (5, 14)	6 (4, 10)
Use hormonal contraception (oral, DMPA)	40% (4/10)	32% (10/31)
Always use condoms with casual clients	90% (9/10)	93% (28/30)
Practice vaginal douching	100% (10/10)	97% (30/31)

Table S4. Analysis of HIV acquisition and integrin β_7^{Hi} $CD4^+$ T cells using conditional logistic regression in CAPRISA 004 (n=165), in FSW cohort from Nairobi (n=41) and the combined cohort analysis (n=206)

Cohort	**β_7^{Hi} in cases (mean, SD)**	**β_7^{Hi} in controls (mean, SD)**	**P value**	**OR**	**95% CIs**
CAPRISA004 (n=165)	10.4 (3.4)	8.9 (3.1)	.007	1.17	1.05, 1.32
FSW, Nairobi (n=41)	7.4 (5.9)	5.1 (2.5)	.148	1.19	0.94, 1.52
Combined (n=205)	10.0 (3.9)	8.2 (3.4)	.002	1.18	1.06, 1.31

Table S5. Correlation of other immunological markers on bulk $CD4^+$ T cells and CD4 decline <500

Variable	P value	aHR	95.0% CI	
			Lower	Higher
CD4 CD38	.311	.99	.96	1.01
CD4 HLADR	.302	1.17	.87	1.56
CD4 HLADR/CD38	.898	.967	.58	1.61
CD4 CCR5	.172	1.03	.99	1.06
CD4 Ki67	.384	1.13	.86	1.49
CD4 TTM*	.232	.98	.95	1.01
CD4 TEM	.224	1.02	.99	1.05
CD4 TTD	.391	1.01	.99	1.02
CD4 TCM	.283	1.02	.98	1.07
CD4 TN	.673	1.01	.97	1.06

* Memory T cell subsets including TTM (transitional memory), TEM (effector memory), TTD (terminally differentiated), TCM (central memory), and TN (naïve)

Table S6. Effect of pre –infection integrin β_7^{Hi} $CD4^+$ T cell frequency on post-infection plasma I-FABP and sCD14 levels using linear mixed model analysis adjusting for VL and CD4:CD8 ratio

Dependent variable	**Estimate**	**p-value**	**95% CI**
I-FABP	-1.13	.904	-20.00, 17.75
sCD14	-94.56	.264	-264.23, 75.12

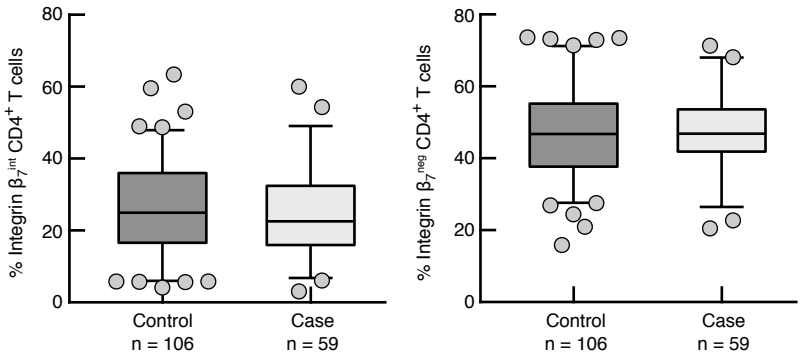

Fig. S1. Effect of pre-infection β_7^{int} CD45RA$^+$CD4$^+$ and β_7^{Neg} CD45RA$^-$CD4$^+$ T cell frequency on HIV acquisition. Differences in β_7^{Int} CD45RA$^+$CD4$^+$ and β_7^{Neg} CD45RA$^-$CD4$^+$ T cell frequency between cases (n=59) and controls (n= 106). In conditional logistic regression analyses no associations were observed for either β_7^{Int} or β_7^{Neg} populations and HIV outcome (p=0.242 and 0.882).

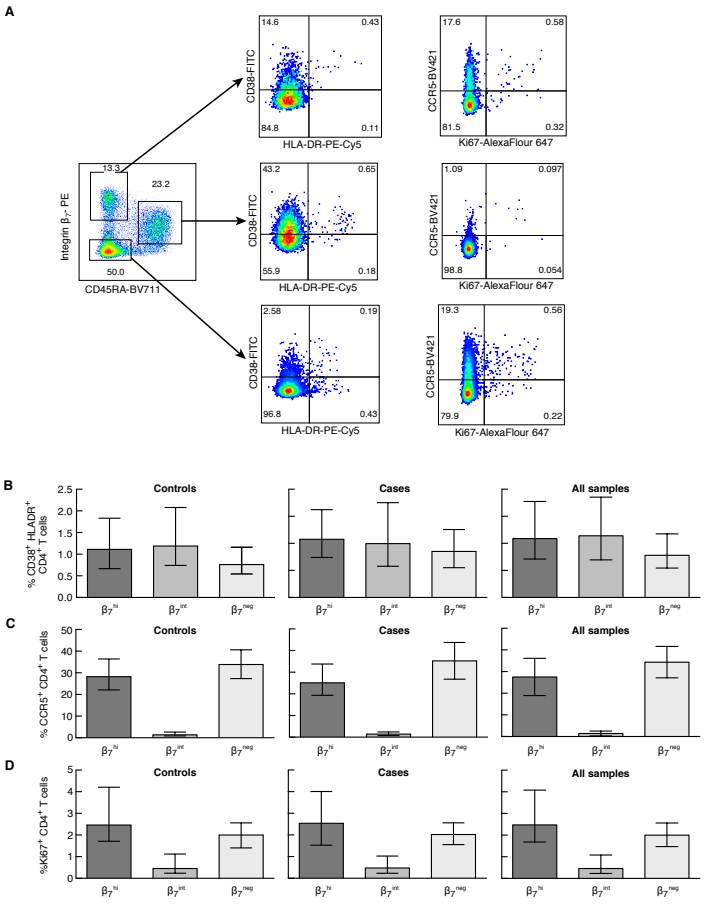

Figure S2. Phenotype of β_7^{Hi} CD45RA$^-$CD4$^+$, $\beta7^{int}$ CD45RA$^+$CD4$^+$ and $\beta7^{Neg}$ CD45RA$^-$CD4$^+$ T cells. **A)** CD4$^+$ T cells were characterized as β_7^{Hi}, β_7^{Int} and β_7^{Neg} cells using CD45RA and β_7 expression. Individual cell subsets were further characterized for the expression of activation (HLA-DR vs. CD38) and HIV co-receptor expression and cell proliferation (CCR5 vs. Ki67) markers. PBMC staining from a representative HIV-uninfected participant is shown. **B)** Frequency of CD38$^+$HLADR$^+$ double positive cells in different β_7 populations between cases (n=76, p=0.0138) and controls (n=165, p<0.0001) and all samples combined (n=241, p<0.0001) **C)** Frequency of CCR5$^+$ cells in different β_7 populations between cases (n=76, p<0.0001) and controls (n=165, p<0.0001) and all samples combined (n=241, p<0.0001) **D)** Frequency of Ki67$^+$ cells in different β_7 populations between cases (n=76, p=0.0001) and controls (n=165, p<0.0001) and all samples combined (n=241, p<0.0001). (Kruskal-Wallis test). Co-expression of immune activation markers CD38 and HLA-DR was higher on β_7^{Int} cells compared with β_7^{Hi} and β_7^{Neg} subsets (1.1 [IQR 0.7-1.9] vs. 1.1 [IQR 0.7-1.8] and 0.8 [IQR 0.6-1.2], ANOVA p<0.0001). In contrast, expression of CCR5 and Ki67 was lower on β_7^{Int} cells which is not unexpected since this T cell subset is mainly naïve (1.4 [0.8-2.5] for CCR5 and 0.5 [IQR 0.2-1.1] for Ki67). Levels of CCR5 and Ki67 were comparable between β_7^{Hi} and β_7^{Neg} cells (27.5 [IQR 20.8-36.1] vs. 34.3 [IQR 27.1-41.7] for CCR5; 2.5 [IQR 1.7-2.5] vs. 2.0 [IQR 1.5-2.6] for Ki67, respectively).

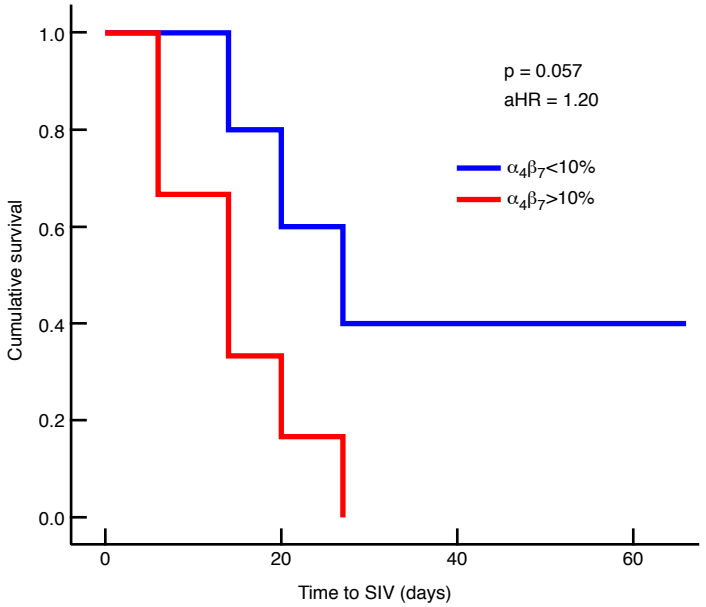

Figure S3. Effect of pre-infection $\alpha_4\beta_7^+$ $CD4^+$ T cell frequency in the blood on SIV acquisition in NHPs exposed to weekly intra-vaginal challenges with SIVmac251. Animals were grouped by $\alpha_4\beta_7$ expression > or < 10%. After adjusting for age, parity, and menses, RMs with higher $\alpha_4\beta_7$ levels acquired SIV more rapidly than RMs with lower $\alpha_4\beta_7$ frequencies (aHR 1.20, 95% CI 0.99-1.44, p=0.057).

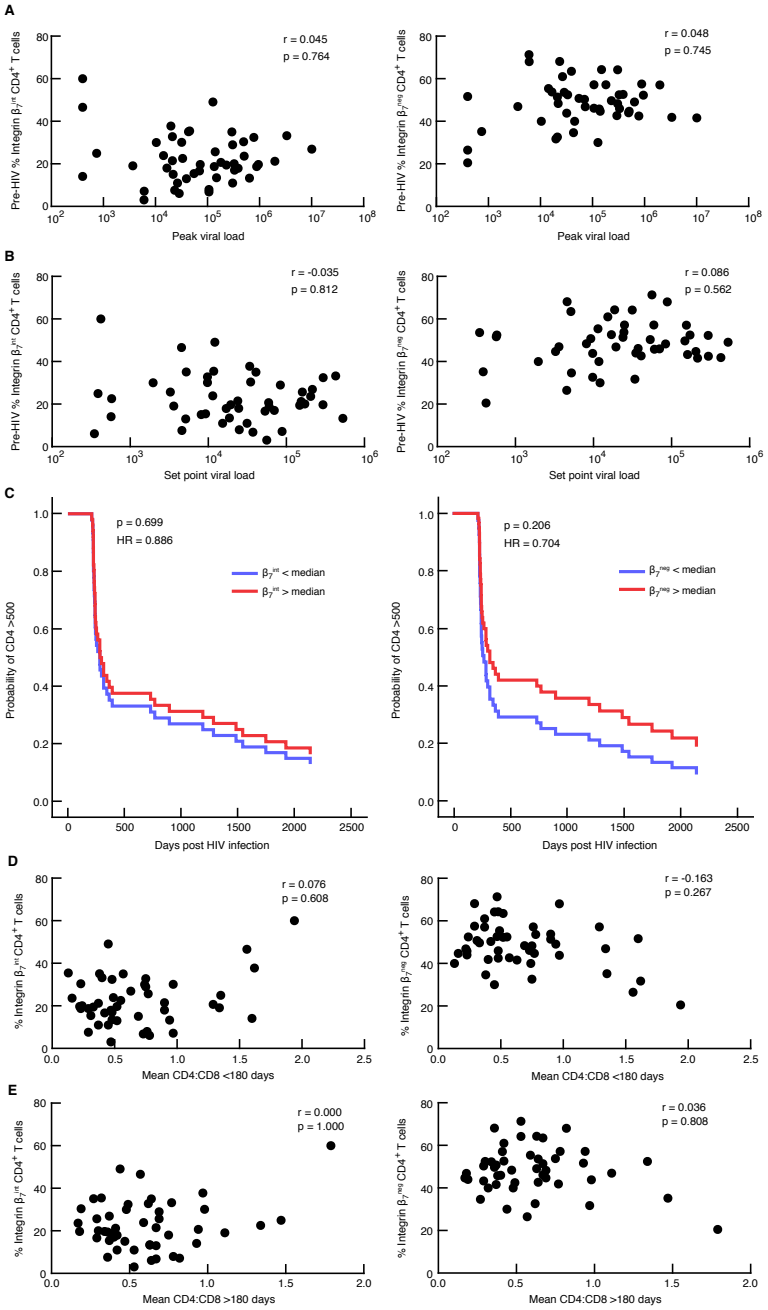

Figure S4. Effect of pre-infection β_7^{int} CD45RA$^+$CD4$^+$ and β_7^{Neg} CD45RA$^-$ CD4$^+$ T cell frequency on disease progression in patients that became infected in CAPRISA 004/002 study. A) Peak VL (n=48, r=0.045, p=0.764 for β_7^{Int}, n=48, r=0.048, p=0.745 for β_7^{Neg}) **B)** Set point VL (n=48, r=-0.035, p=0.821 for β_7^{Int}, n=48, r=0.086, p=0.562 for β_7^{Neg}) **C)** CD4$^+$ T cells decline<500 cells/µl (HR=0.886, p=0.699 for β_7Int, HR=0.704, p=0.206 for β_7^{Neg}) **D)** Mean CD4:CD8 ratio <180 days post infection (β_7^{Int}: n=48, r=0.076, p=0.608; β_7^{Neg}: n=48, r=-0.163, p=0.267) **E)** Mean CD4:CD8 ration > 180 days post infection (β_7^{Int}: n=48, r= 0.000, p=1.000; β_7^{Neg}: n=48, r=0.036, p= 0.808)

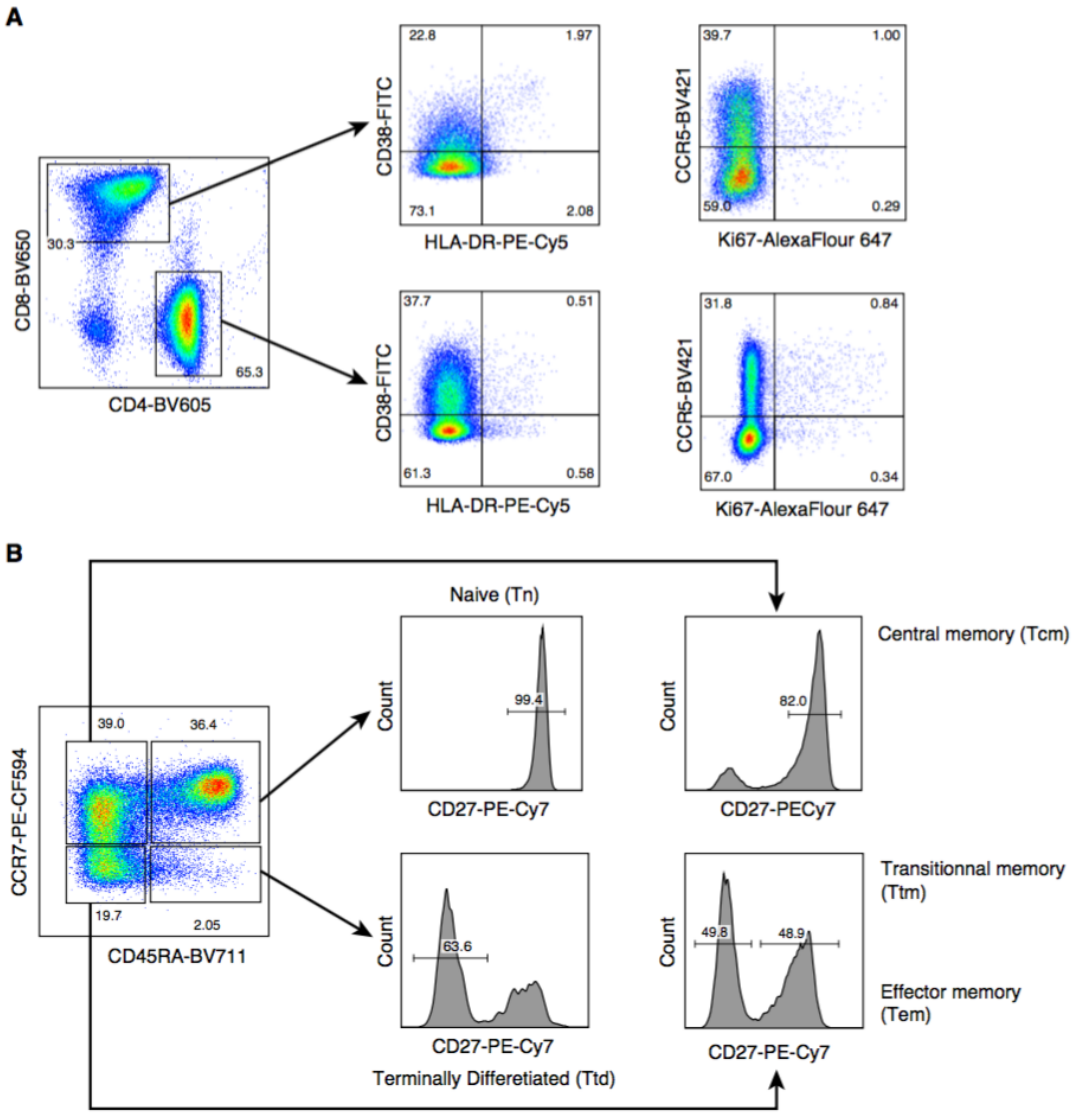

Figure S5. Gating strategy utilized for the analysis of frozen PBMCs from the CAPRISA 004 study. Cells were thawed, washed to remove cryopreservant and rested in fresh media for 3 hours at 37°C, 5% CO_2 before staining with the antibodies indicated in the figure by polychromatic flow cytometry. In the initial gating, cells were gated for singlets (FSC-H vs. FSC-A), then Live/Dead Aqua stain, followed by lymphocytes (SSC-A vs. FSC-A) and finally T cells (all $CD3^+$ lymphocytes) **A) Primary outcome measures**. $CD3^+$ T cells were separated into $CD4^+$ and $CD8^+$ T cells and further characterized for markers of activation (HLA-DR vs. CD38) and HIV co-receptor expression and a marker for cell proliferation (CCR5 vs. Ki67) **B) $CD4^+$ T cell subsets**. $CD4^+$ T cells were categorized according to the expression of CD45RA, CCR7 and CD27: T_N ($CD45RA^+CCR7^+CD27^+$), T_{TD} ($CD45RA^+CCR7^-CD27^-$), T_{TM} ($CD45RA^-CCR7^-CD27^+$), T_{EM} ($CD45RA^-CCR7^-CD27^-$), T_{CM} ($CD45RA^-CCR7^+CD27^+$). PBMC staining from a representative HIV-uninfected participant is shown.

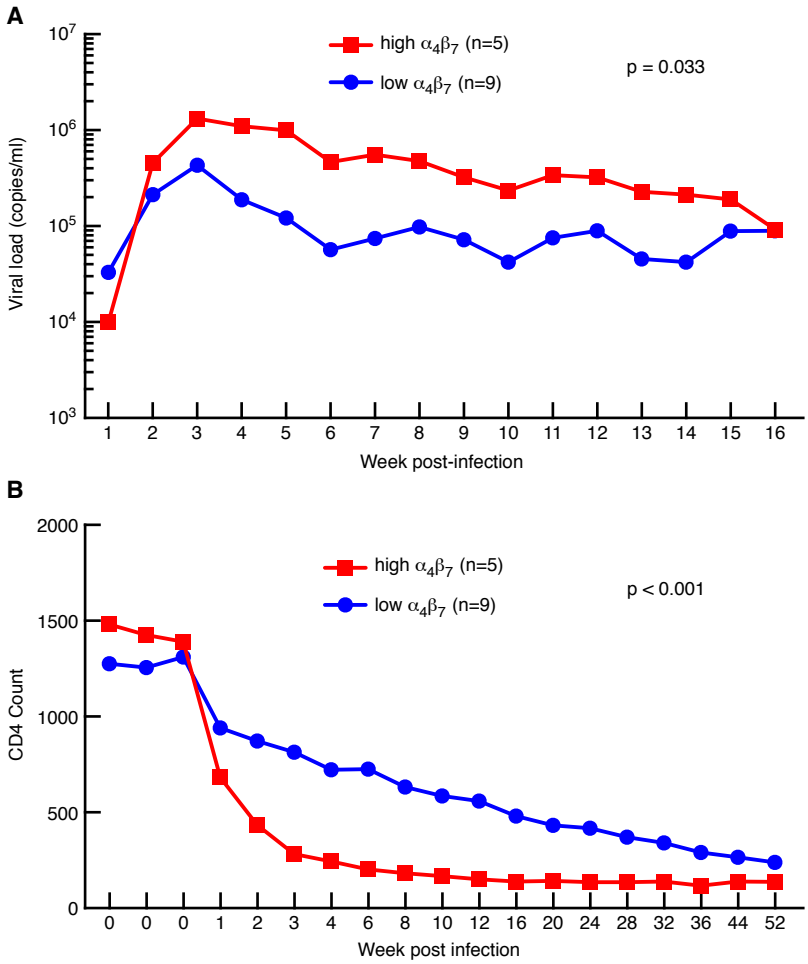

Figure S6. Effect of pre-infection $\alpha_4\beta_7^+$ $CD4^+$ T cell frequency in the gut and the blood on disease progression in RMs following IV injection with SIVmac239. A. Weekly viral load (copies/ml) measurements in RMs segregated by pre-infection $\alpha_4\beta_7$ $CD4^+$ T cell frequency below and above 30% (p=0.033). **B.** CD4 count measurement in RMs segregated by pre-infection $\alpha_4\beta_7$ $CD4^+$ T cell frequency below and above 30%. (p<0.001). Data was analyzed using linear mixed models.