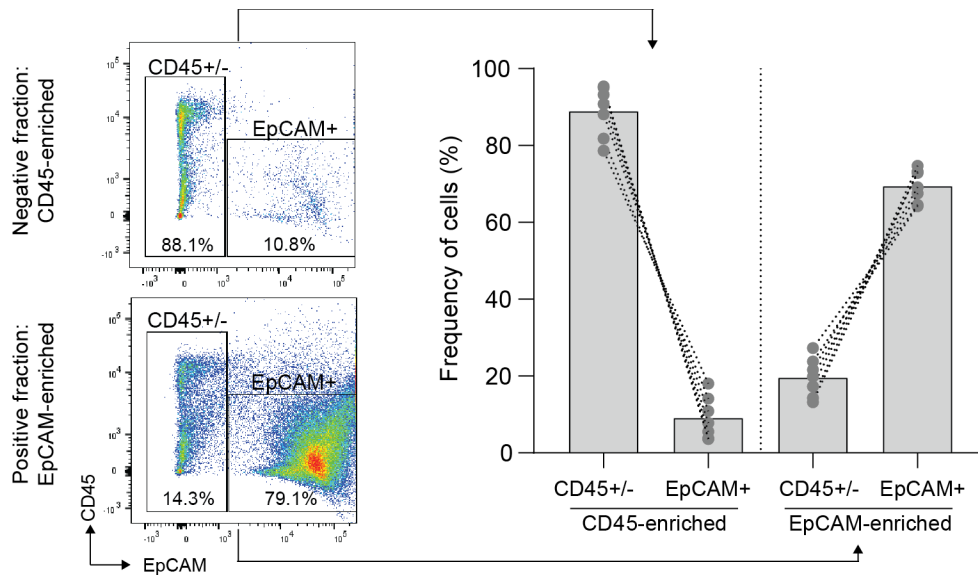
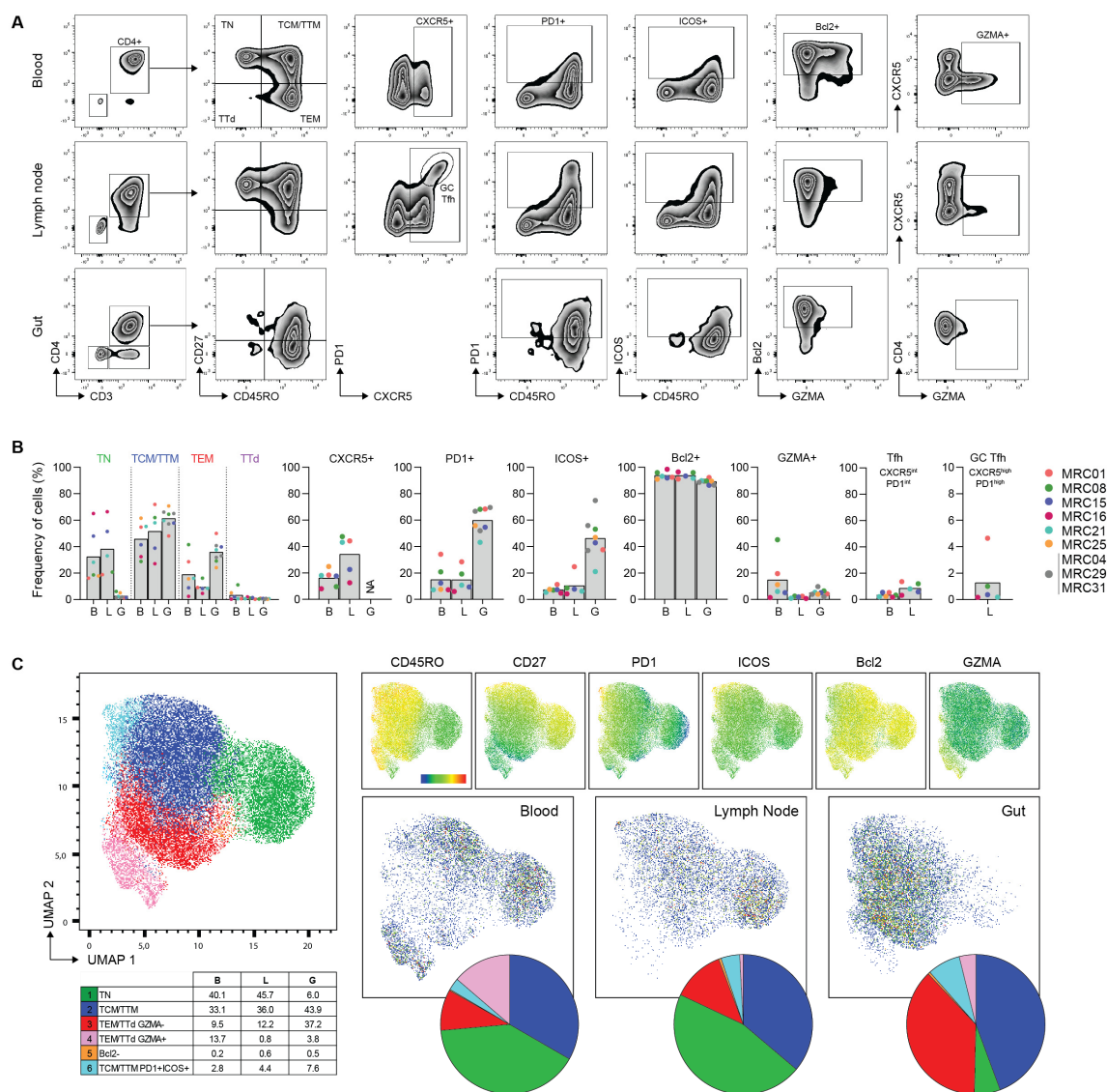
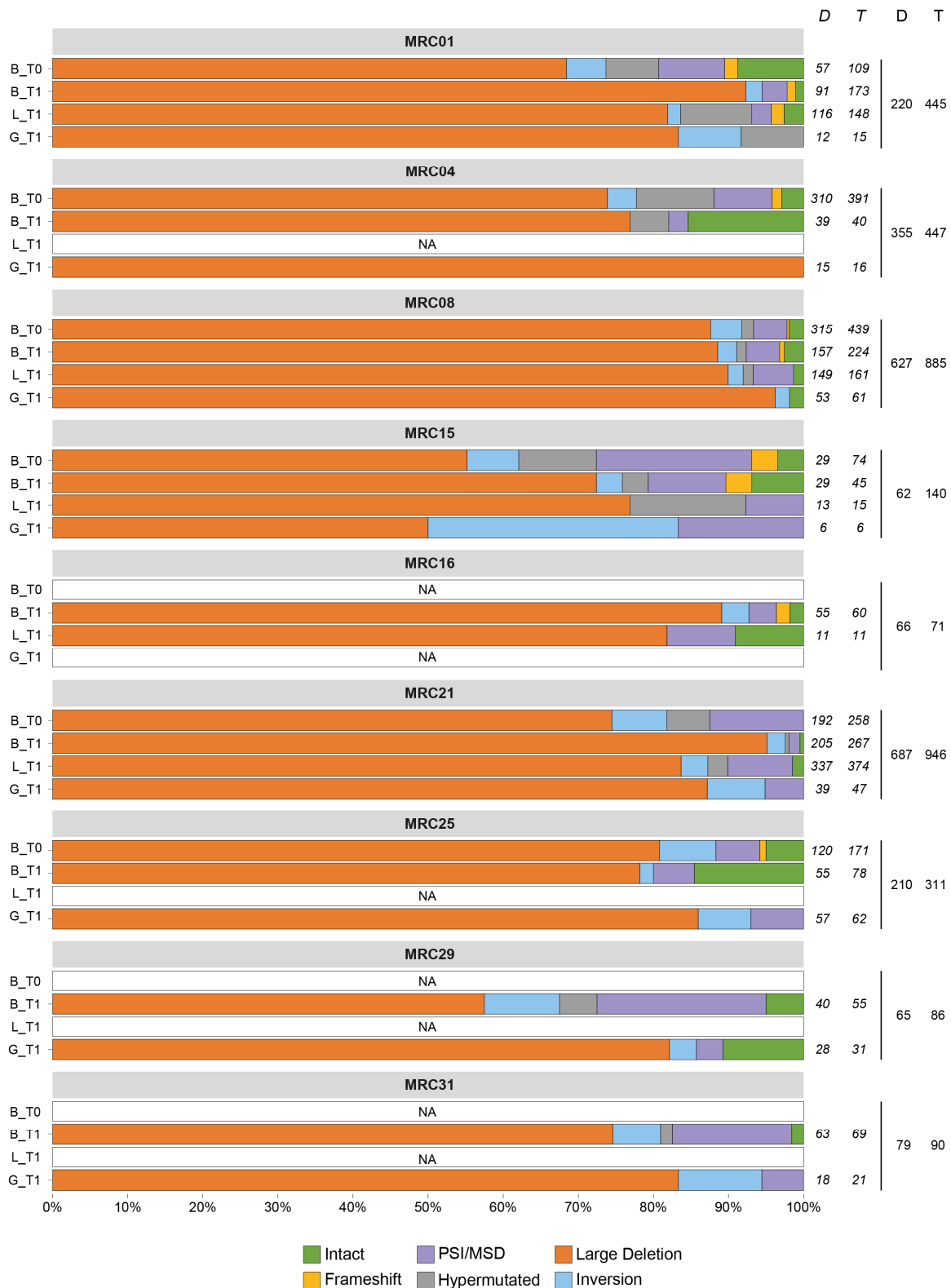




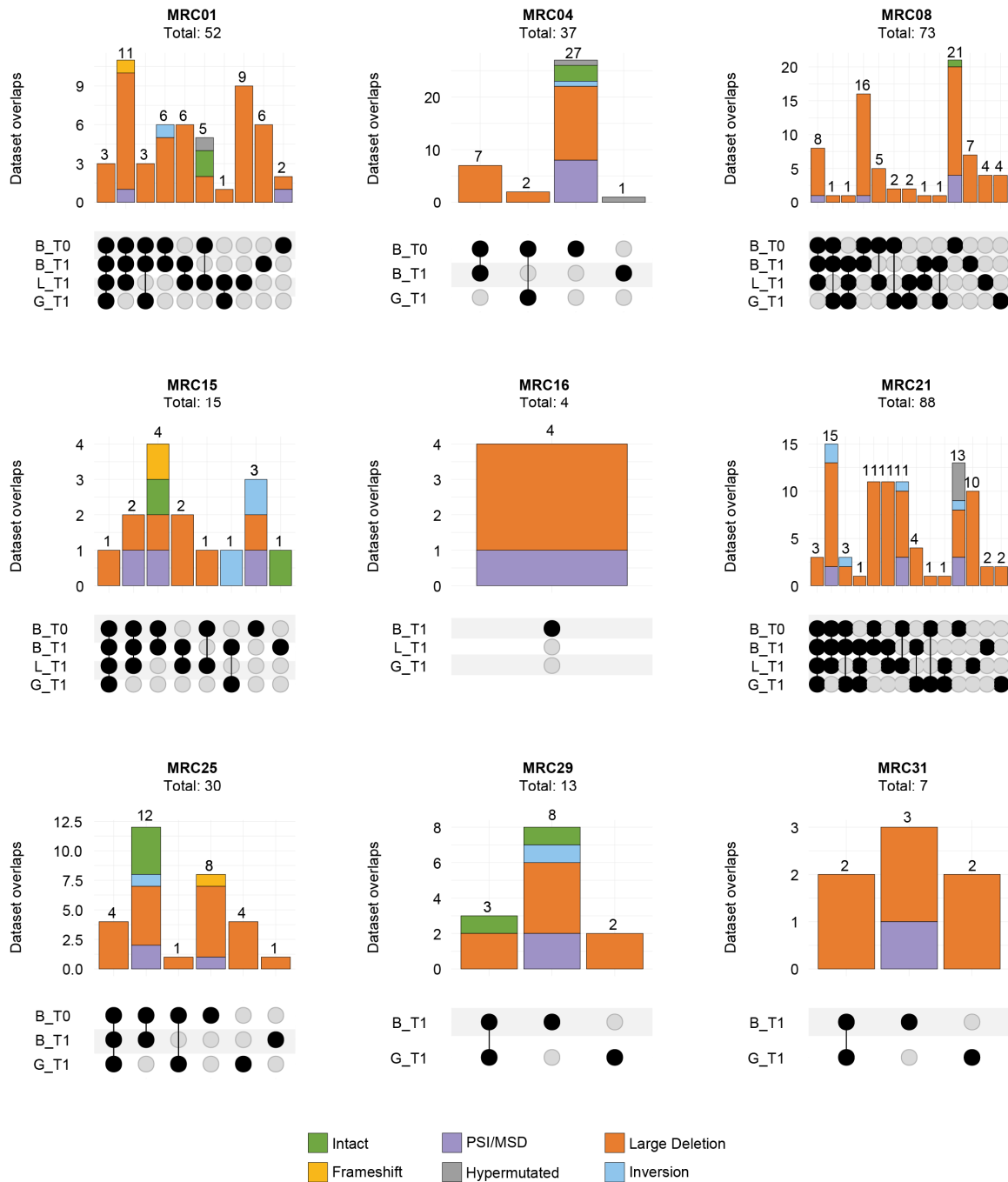
Supplementary Figure 1. CD45^{+/-} cells enrichment from gut biopsies. Magnetic cell isolation (EasySep™ Human EpCAM Positive Selection Kit) was used to enrich for immune CD45^{+/-}-EpCAM⁻ cells from gut biopsies. Left panel: representative flow cytometry dot plots showing CD45/EpCAM dual staining in the negative (CD45-enriched) versus positive (EpCAM-enriched) fractions. Right panel: percentages of CD45^{+/-} and EpCAM⁺ cells in the negative (CD45-enriched) versus positive (EpCAM-enriched) fractions (n = 7 ART-suppressed individuals). Source data are provided as a Source Data file.



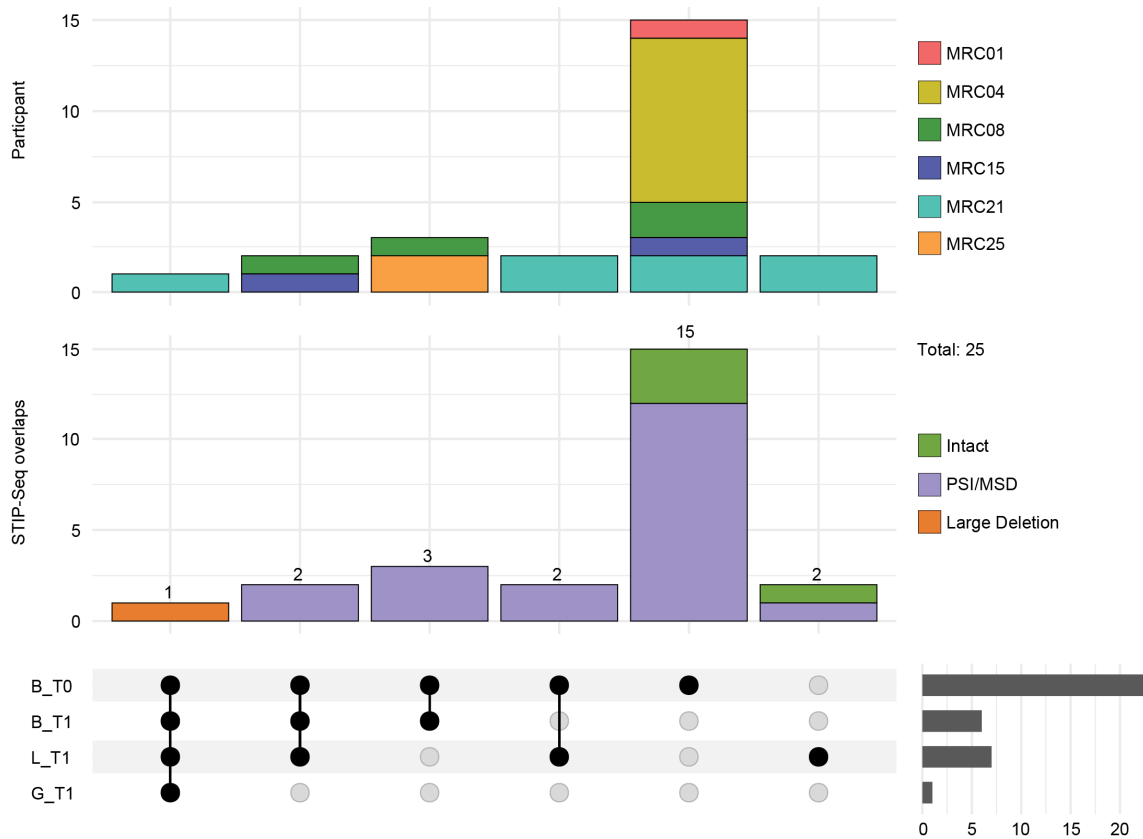
Supplementary Figure 2. Immunophenotyping of non-stimulated CD4 T cells in matched blood and tissues. A) Representative flow cytometry dot plots comparing the expression of the markers of interest between matched peripheral blood, lymph node and gut (MRC01). B) Frequency of CD4 T cells expressing the markers of interest in matched blood, lymph node and gut samples (n = 6, 5 and 8 ART-suppressed participants, respectively). Participants who provided samples from at least two anatomical sites are color-coded, whereas those with only gut samples analyzed are represented by a grey dot. Grey columns depict mean values. TN = naïve T cells, TCM/TTM = central/transitional memory T cells, TEM = effector memory T cells, TTd = terminally differentiated T cells. B = blood, L = lymph node, G = gut. C) FlowSOM analysis: Uniform Manifold Approximation and Projection (UMAP) combining CD4⁺ T cells from blood, lymph node and gut samples (n = 6, 5 and 8 ART-suppressed participants, respectively). Proportions for each cluster within each anatomical compartment are indicated in the table and in the pie charts. The levels of expression of each marker included in this analysis, as well as the UMAPs for each individual anatomical site, are also represented. Source data are provided as a Source Data file.



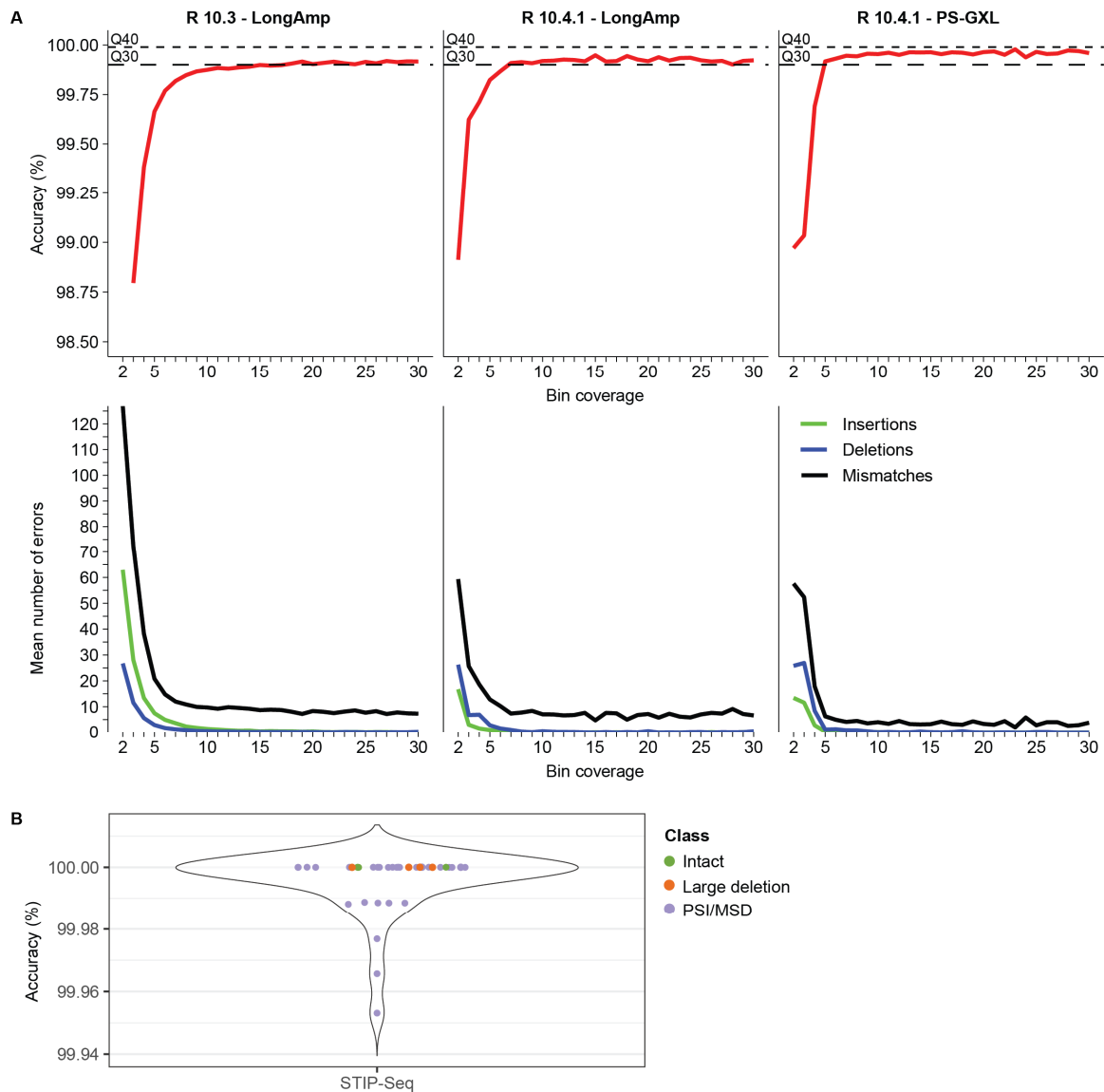
Supplementary Figure 4. Classification of HIV-1 proviruses from matched blood and tissues. The proviral reservoir was assayed by the HIV-PULSE assay. Proportions of different proviral classes observed among the distinct proviruses for each anatomical site (B = blood, L = lymph node, G = gut) and time point (T0 and T1) for each participant (n = 9). The numbers of distinct (D) and total (T) proviruses are displayed on the right, per anatomical site (in italics), and per participant (upright text). Source data are provided as a Source Data file.



Supplementary Figure 5. Clones of infected cells overlap between time points and anatomical compartments. The proviral reservoir was assayed by the HIV-PULSE assay. UpSet-plots visualizing the number of distinct HIV-PULSE sequences that overlap between anatomical compartments (B = blood, L = lymph node, G = gut) and time points (T0 and T1) for each participant (n = 9). Proviral classification is color-coded. The total number of distinct overlapping sequences is indicated on the top of each bar plot. Source data are provided as a Source Data file.



Supplementary Figure 6. Clones of infected cells carrying an inducible provirus overlap between time points and anatomical compartments. UpSet-plot combining all sequences that overlap between HIV-PULSE and STIP-Seq (n = 6 ART-suppressed participants): matches between anatomical compartments (B = blood, L = lymph node, G = gut) and time points (T0 and T1) are indicated. The contribution of each participant to these matches is shown on the upper panel. Proviral classification is color-coded. The total number of distinct overlapping sequences is indicated on the top of each bar plot. The axis of the bar graph on the right indicates the total number of distinct sequences per dataset. Source data are provided as a Source Data file.



Supplementary Figure 7. HIV-PULSE sequencing accuracy. A) Benchmarking experiments on UMI bin accuracy were conducted using the J-Lat 10.6 cell line. PCR products obtained with previous (LongAmp) and updated (PS-GXL) pre-amplification protocols were sequenced using different ONT technologies (R10.3, LSK109, Guppy v.5.0.17 vs R10.4.1, LSK114, Guppy v.6.5.7). The upper panel shows the mean bin centroid accuracy with increasing bin coverage. Mean accuracies for bins with a coverage of at least 15 reads are 99.91% with the original protocol (R10.3, LongAmp enzyme; left graph), 99.93% with the updated ONT technology (R10.4.1, LongAmp enzyme; middle graph) and 99.97% with the replacement of the enzyme (R10.4.1, PS-GXL enzyme; right graph). The lower panel shows the mean number of errors with increasing bin coverage. Insertions, deletions and mismatches are color-coded. Using the newest ONT technology (R10.4.1), combined with the use of PS-GXL, ~3.1 residual errors per 9 kb genome remain for bins with a coverage of at least 5 reads, predominantly substitutions. B) Violin plot of 'megabin' accuracy of HIV-PULSE genomes (n=37) when compared to their STIP-Seq reference counterparts (n = 6 ART-suppressed participants). Dots are colored by proviral classification. Source data are provided as a Source Data file.

ID	B T0	B T1	L T1	G T1	Age	Gender	Subtype	VL	ART duration				T1 - T0
									B T0	B T1	L T1	G T1	
MRC01					50-60	M	B	<20	18.5	20.5	20.4	20.5	2.0
MRC04					50-60	M	B	<20	18.4	19.8	x	19.8	1.4
MRC08					50-60	M	B	<20	1.3	2.7	2.8	2.7	1.4
MRC15					50-60	M	B	<20	14.7	16.6	16.6	16.7	1.9
MRC16					40-50	F	F01	<20	x	14.6	14.6	x	x
MRC21					60-70	M	B	<20	24.8	26.5	26.5	26.5	1.7
MRC25					40-50	M	B	<20	11.6	12.4	x	12.4	0.8
MRC29					60-70	M	B	<20	x	14.3	x	14.3	x
MRC31					40-50	M	B	<20	x	7.8	x	7.6	x

Supplementary Table 1. Clinical data of the participants.

B = blood, L = lymph node, G = gut; time points T0 and T1

Gender: M = male, F = female

VL = viral load

ART duration (years); T1-T0: time between the two time points (years)

Participant	Comparison	Hudson, Boos and Kaplan ¹		Slatkin-Maddison ²	
		K _{ST}	P-value	Migration events	P-value
MRC01	B_T0 vs B_T1	0.021	0.473	7	0.656
MRC01	B_T0 vs L_T1	0.011	0.634	12	0.991
MRC01	B_T0 vs G_T1	0.093	0.167	0	1
MRC01	B_T1 vs L_T1	0.029	0.08	5	0.127
MRC01	B_T1 vs G_T1	0.058	0.13	1	0.8
MRC01	L_T1 vs G_T1	0.054	0.125	1	0.91
MRC04	B_T0 vs B_T1	0.004	0.444	20	0.742
MRC04	B_T0 vs G_T1	0	0.528	1	0.047
MRC04	B_T1 vs G_T1	0.003	0.642	2	1
MRC16	B_T1 vs L_T1	0	0.995	10	1
MRC29	B_T1 vs G_T1	0.057	0.333	14	0.622
MRC31	B_T1 vs G_T1	0.039	0.348	8	0.402
MRC08	B_T0 vs B_T1	0.009	0.418	47	0.999
MRC08	B_T0 vs L_T1	0	0.969	33	0.965
MRC08	B_T0 vs G_T1	0.028	0.078	5	0.628
MRC08	B_T1 vs L_T1	0.01	0.552	4	0.114
MRC08	B_T1 vs G_T1	0.032	0.18	6	0.988
MRC08	L_T1 vs G_T1	0.116	0	3	0.947
MRC15	B_T0 vs B_T1	0.029	0.997	1	0.719
MRC15	B_T0 vs L_T1	0.251	0.099	2	0.697
MRC15	B_T0 vs G_T1	0.082	0.312	1	1
MRC15	B_T1 vs L_T1	0.234	0.12	1	0.762
MRC15	B_T1 vs G_T1	0.13	0.201	0	1
MRC15	L_T1 vs G_T1	0.065	0.555	0	0.582
MRC21	B_T0 vs B_T1	0.005	0.565	39	1
MRC21	B_T0 vs L_T1	0.014	0.206	36	0.878
MRC21	B_T0 vs G_T1	0.013	0.378	3	0.815
MRC21	B_T1 vs L_T1	0.009	0.286	37	0.939
MRC21	B_T1 vs G_T1	0.014	0.251	5	0.97
MRC21	L_T1 vs G_T1	0.015	0.175	5	0.569
MRC25	B_T0 vs B_T1	0	0.972	18	0.945
MRC25	B_T0 vs G_T1	0.039	0.191	14	0.779
MRC25	B_T1 vs G_T1	0.063	0	5	0.733

¹ Hudson, Boos and Kaplan's nonparametric test for population structure (distance-based). K_{ST} values range from 0 (no compartmentalization) to 1 (full compartmentalization).

² Slatkin-Maddison (tree-based) test. The minimum number of migration events needed to explain the observed distribution of traits on a phylogenetic tree.

The p-value indicates whether the observed distribution is significantly different from what would be expected under a random distribution.

Supplementary Table 2. Tests for panmixia based on *nef* alignments.

B = blood, L = lymph node, G = gut; time points T0 and T1

ID	HIV-Flow		IPDA			HIV-PULSE			
	B_T1	L_T1	B_T1	L_T1	G_T1	B_T0	B_T1	L_T1	G_T1
MRC01	1.9E+6	1.1E+6	2.2E+5	2.3E+5		5.4E+5	6.6E+5	1.0E+6	
MRC04						2.1E+6	3.6E+5		
MRC08	2.1E+6	7.9E+5	2.4E+5	1.8E+5		2.4E+6	1.8E+6	6.6E+5	
MRC15	2.3E+6	5.8E+5	2.6E+5	1.3E+5		1.5E+6	2.0E+6	5.6E+5	
MRC16	3.0E+6	4.5E+5	3.6E+5	9.6E+4			1.6E+6	2.6E+5	
MRC21	2.1E+6	1.8E+6	2.4E+5	1.9E+5		5.5E+5	6.9E+5	6.7E+5	
MRC25						1.1E+6	6.2E+5		
MRC29							3.0E+5		
MRC31							1.9E+6		

CD45-enriched cells from the gut were not counted

Samples not available/not tested

Note 1: HIV-Flow was only performed if B_T1 and L_T1 were available

Note 2: HIV-PULSE was performed on all samples available

Supplementary Table 3. Number of CD4 T cells analyzed per participant using HIV-Flow, IPDA and HIV-PULSE.

B = blood, L = lymph node, G = gut; time points T0 and T1

