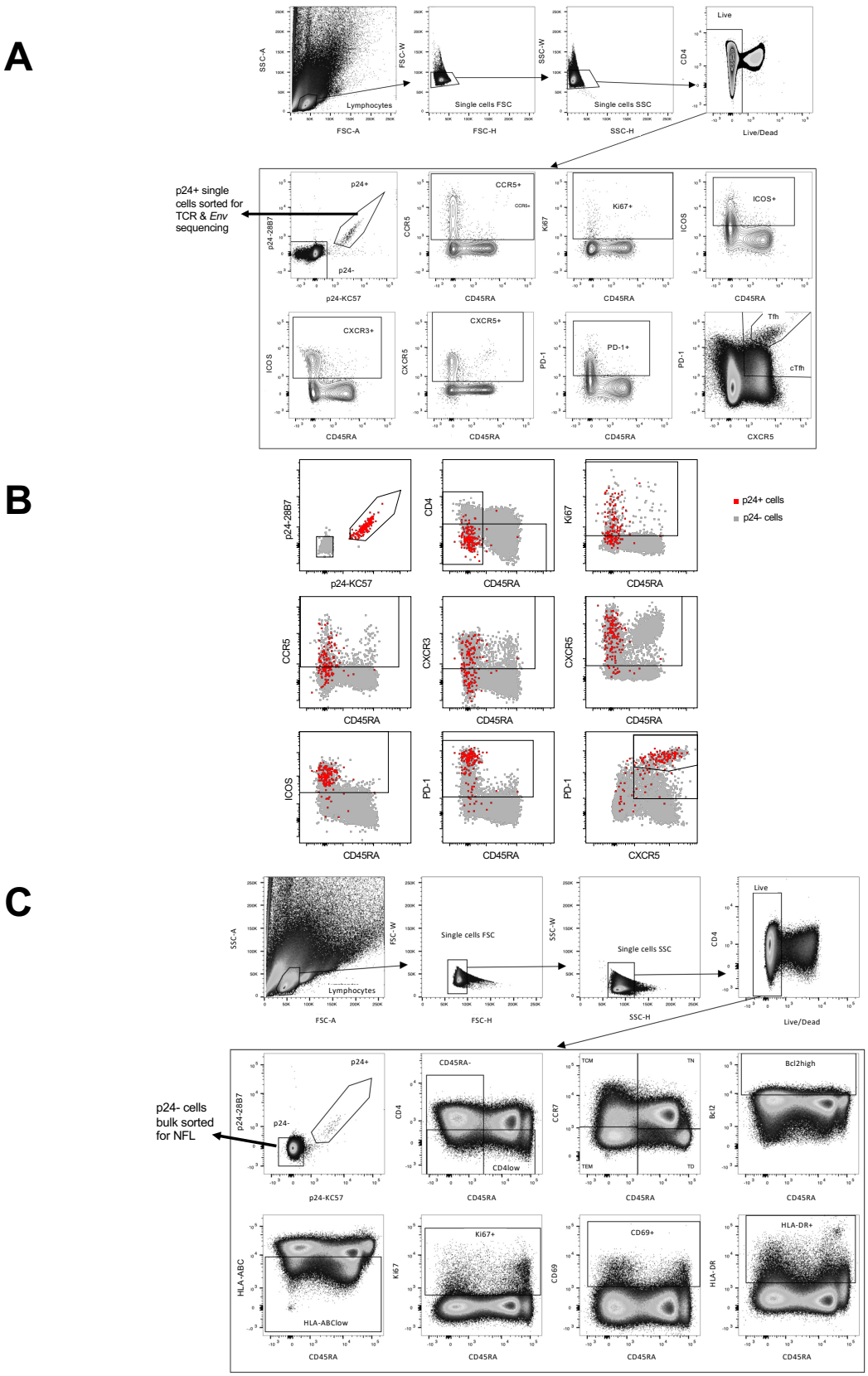


Figure S1. Gating strategy for single cell sorting, Related to Figures 1 and 6.

A. Gating strategy used for index cell sorting of p24^{+} cells for subsequent TCR and HIV Env sequencing with recording of CD45RA, CCR5, Ki67, ICOS, CXCR3, CXCR5 and PD-1 expression. **B.** Representative dot plots showing expression of CD4, CD45RA, Ki67, CCR5, CXCR3, CXCR5, ICOS and PD-1 in p24^{+} cells (in red), overlaid onto p24^{-} cells (in grey). **C.** Gating strategy used to quantify *ex vivo* and induced p24^{+} cells before and after 96 weeks of ART. Bulk p24^{-} cells were sorted for proviral HIV DNA quantification and near-full length sequencing (NFL).

Figure S2

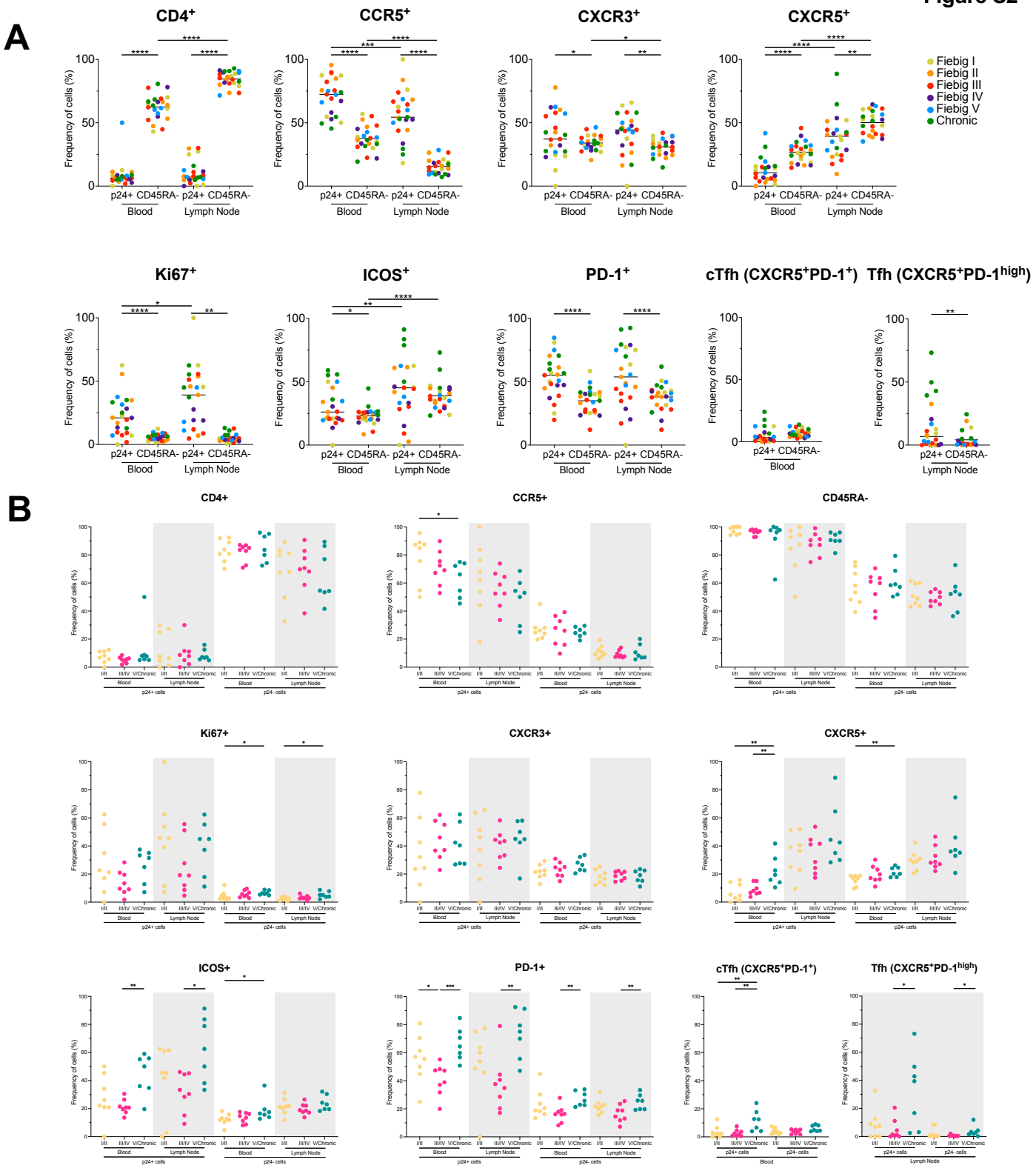


Figure S2. Expression frequencies of cellular markers on p24$^+$ cells compared to memory (CD45RA$^-$) cells, Related to Figure 1.

A. Phenotype of productively infected cells compared to memory cells. Frequencies of p24$^+$ and CD45RA$^-$ cells from blood and lymph node expressing each marker or combination of markers (CD4$^+$, CCR5$^+$, Ki67$^+$, CXCR3$^+$, CXCR5$^+$, PD-1$^+$, ICOS$^+$, circulating T follicular helpers (cTfh) cells and Tfh cells) are depicted for each participant. **B.** In each panel, the frequency of p24$^+$ (left) and p24$^-$ (right) cells from blood and lymph node expressing a given marker or combination of markers (CD4$^+$, CCR5$^+$, CD45RA$^-$, Ki67$^+$, CXCR3$^+$, CXCR5$^+$, PD-1$^+$, ICOS$^+$, circulating T follicular helpers (cTfh) cells and Tfh cells) is depicted for Fiebig stages (I/II) (acute infection), III/IV (early infection) and V/chronic (early chronic infection). (Mann-Whitney; p<0.05, *; p<0.01**; p<0.001, ***; p<0.0001 ****).

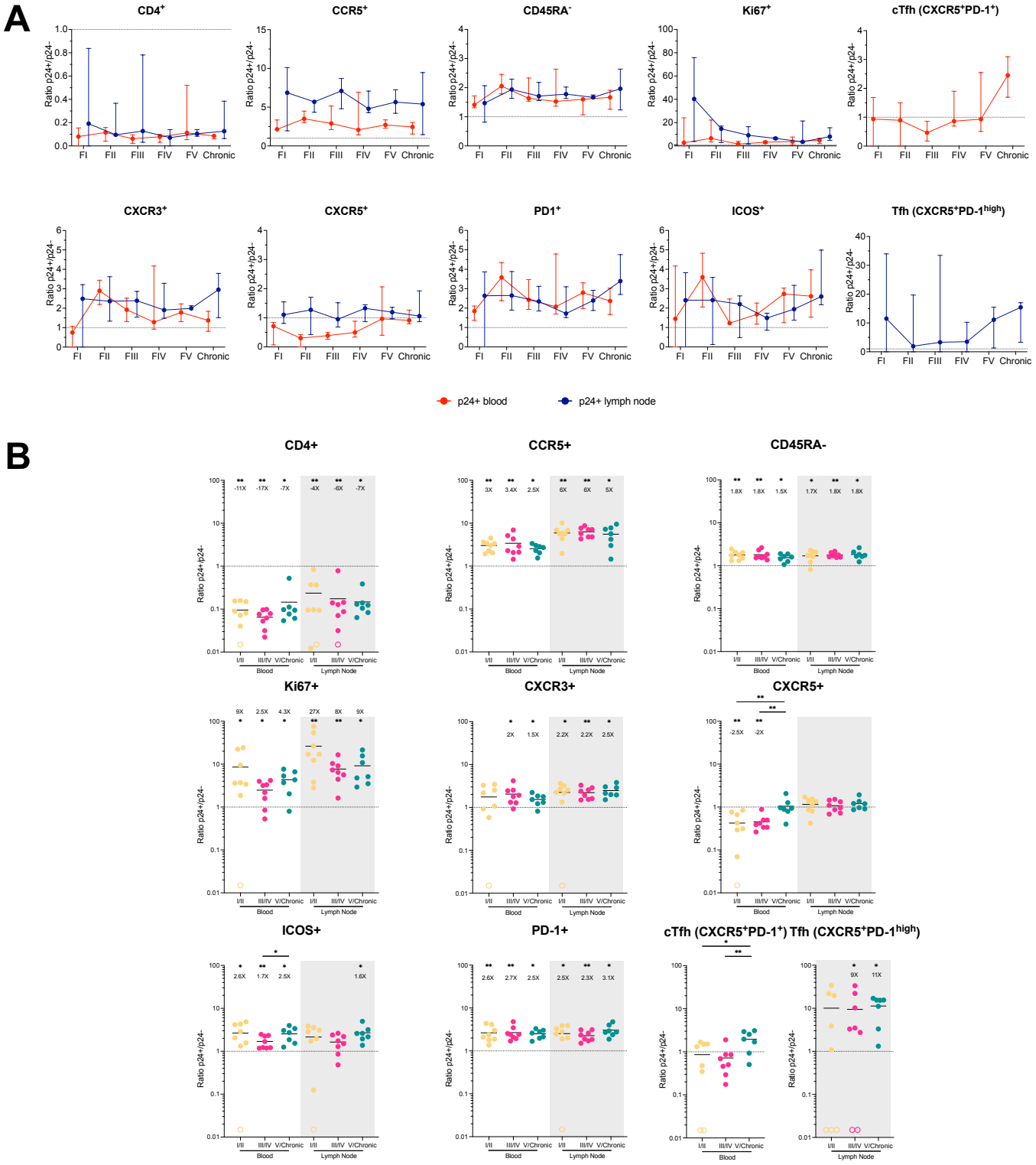
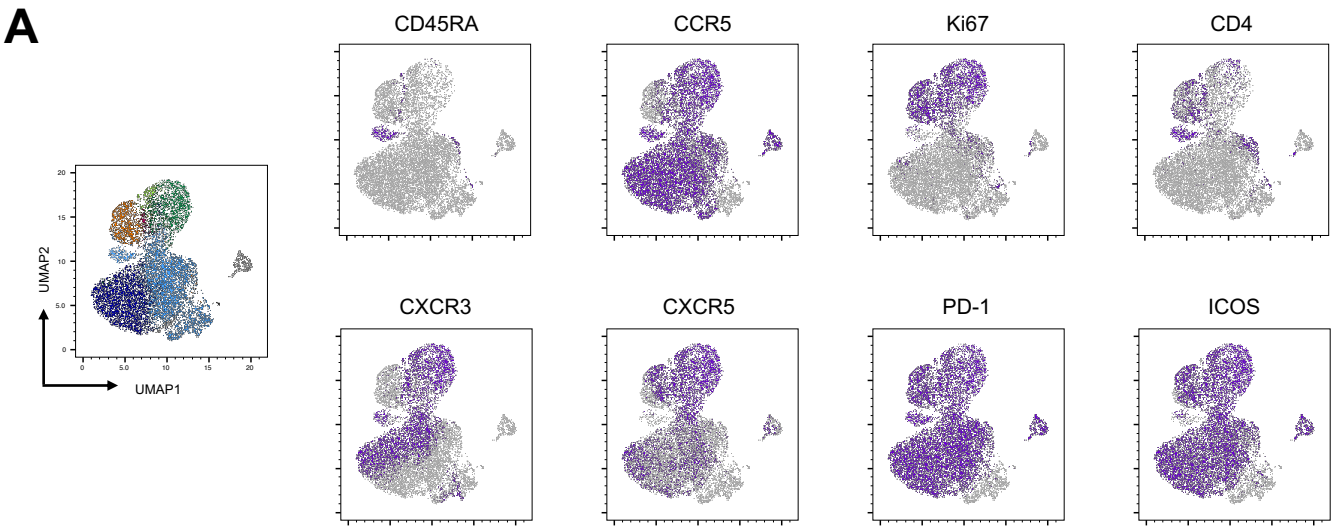
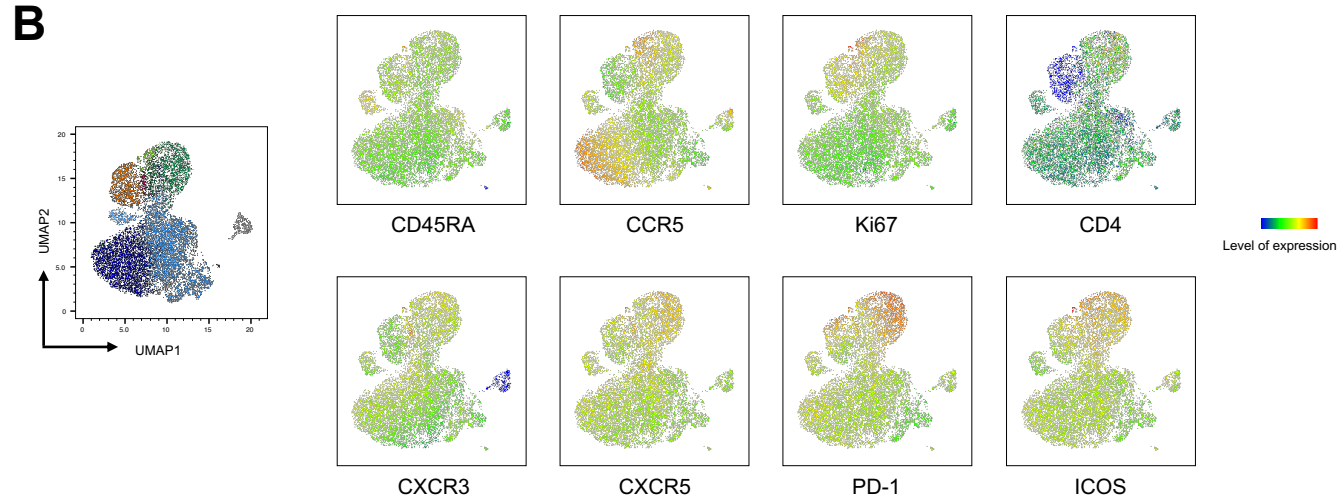


Figure S3. Productively infected and non-infected CD4⁺ T cells display distinct phenotypes, Related to Figure 1.

A. Ratio of frequencies of cells ($p24^+/p24^-$) from blood (red) or lymph node (blue) expressing each marker or combination of markers ($CD4^+$, $CCR5^+$, $CD45RA^-$, $Ki67^+$, $CXCR3^+$, $CXCR5^+$, $PD\text{-}1^+$, $ICOS^+$, circulating T follicular helpers (cTfh) cells and Tfh cells) is depicted for each Fiebig stage (I to V and Chronic infection). Median values are plotted with 95% CI. **B.** In each panel, the ratio of frequencies ($p24^+/p24^-$) of cells from blood (left) or lymph node (right) expressing each marker or combination of markers ($CD4^+$, $CCR5^+$, $CD45RA^-$, $Ki67^+$, $CXCR3^+$, $CXCR5^+$, $PD\text{-}1^+$, $ICOS^+$, circulating T follicular helpers (cTfh) cells and Tfh cells) is depicted for Fiebig stages (I/II) (acute infection), III/IV (early infection) and V/chronic (early chronic infection). Median values are shown as horizontal bars. (Mann-Whitney or Wilcoxon; $p<0.05$, *; $p<0.01$**; $p<0.001$, ***; $p<0.0001$ ****).



	n cells	% of cells	Ki67+	CCR5+	CXCR3+	CXCR5+	PD-1+	ICOS+	CD45RA-
1	1793	Tfh	43	99	88	99	99	100	100
2	293	Proliferative Tfh	99	98	94	93	98	94	86
3	124	Proliferative, Th1-like	98	54	100	90	100	99	93
4	1027	Proliferative non-Th1-like	99	33	14	63	97	90	96
5	3417	Non-Proliferative, Th1-like	3	99	50	36	98	83	100
6	4934	Non-Proliferative non-Th1-like	12	69	34	30	80	61	93
7	427	Non-Th1	8	86	0	34	91	80	98
8	52	Others	0	37	6	12	46	39	54



	n cells	Median MFI	Ki67	CCR5	CXCR3	CXCR5	PD-1	ICOS	CD45RA
1	1793	Tfh	1747	3243	2183	6882	23131	8727	839
2	293	Proliferative Tfh	13166	8465	2785	4674	9025	16851	1054
3	124	Proliferative, Th1-like	4452	752	8038	6205	4384	5280	1180
4	1027	Proliferative non-Th1-like	3597	525	494	1829	3273	2421	1127
5	3417	Non-Proliferative, Th1-like	283	8648	962	1289	1758	1524	444
6	4934	Non-Proliferative non-Th1-like	467	1029	678	1201	1250	1212	546
7	427	Non-Th1	253	1614	1	756	1308	1460	333
8	52	Others	206	4452	404	731	1382	1726	3

Figure S4. Expression of cellular markers in p24+ cells clusters, Related to Figure 2.
p24^{+} cells phenotypic data from 23 participants were integrated in a UMAP analysis and generated 8 cell clusters (left). **A.** Percentage of expression, or **B.** mean fluorescence intensity [MFI] of CD45RA, CCR5, Ki67, CD4, CXCR3, CXCR5, PD-1 and ICOS are depicted. The tables show the phenotype of each cell cluster and the percentage of expression (A) or MFI (B) of each marker.

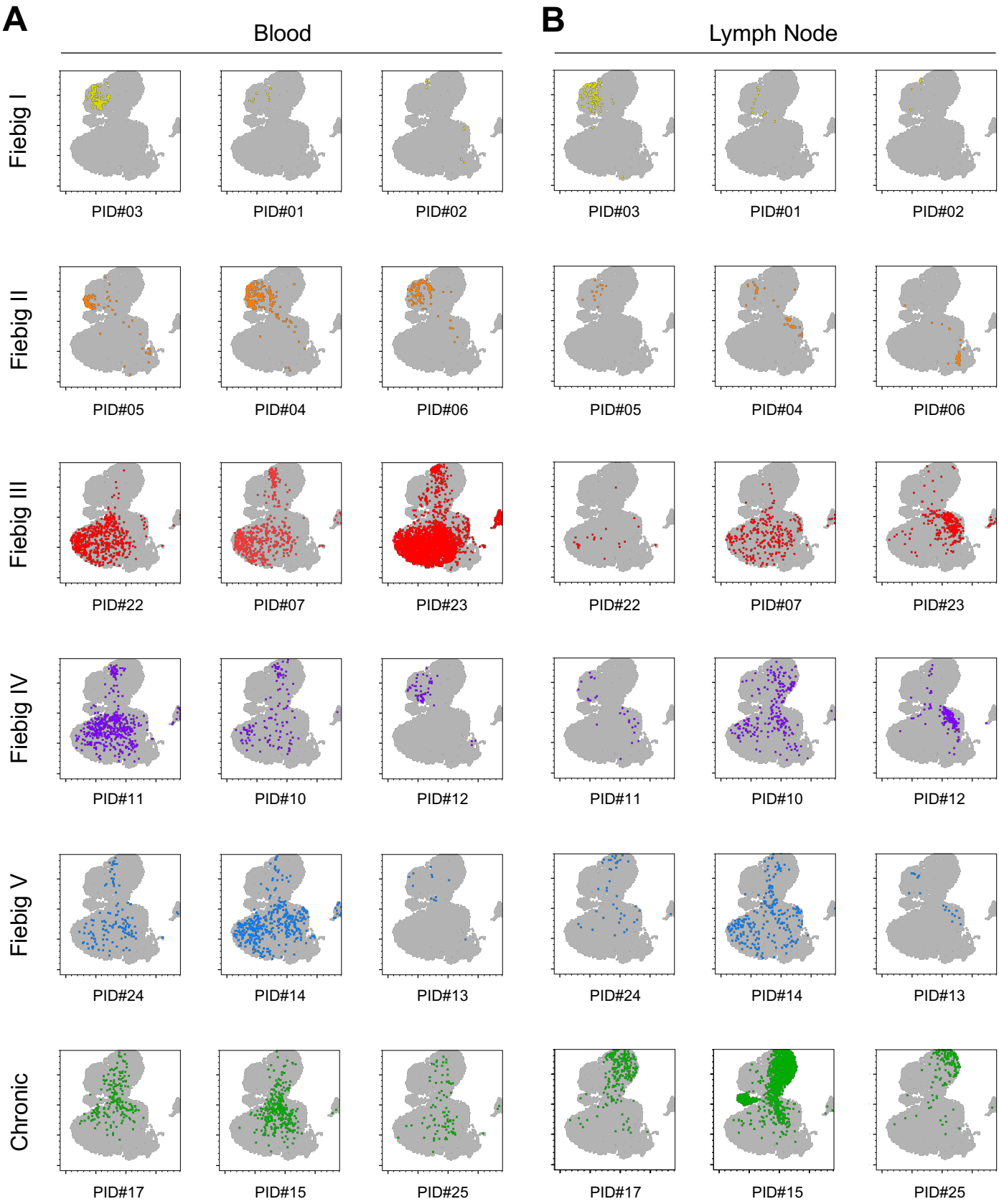


Figure S5. Clustering of p24⁺ cells in representative participants, Related to Figure 2.
Dot plots of the UMAP analysis of p24⁺ cells from individual participants and from blood (panel A) and lymph node (panel B). p24⁺ cells are overlaid in colors.

Figure S6

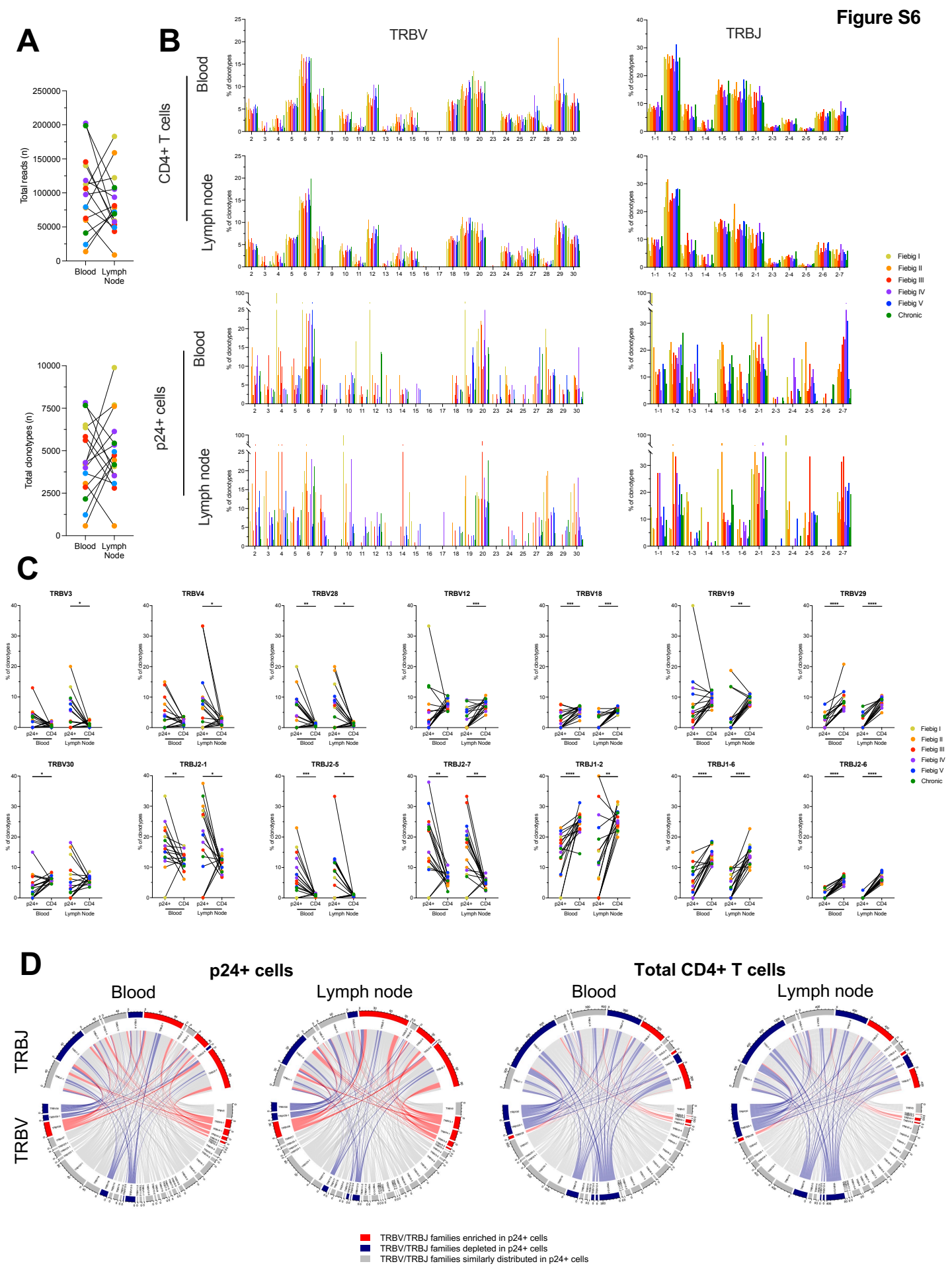

A
Total reads (n)
Blood
Lymph Node
Total clonotypes (n)
B
TRBV
TRBJ
CD4+ T cells
p24+ cells
Blood
Lymph node
% of clonotypes
Fiebig I
Fiebig II
Fiebig III
Fiebig IV
Fiebig V
Chronic
C
TRBV3
TRBV4
TRBV28
TRBV12
TRBV18
TRBV19
TRBV29
TRBV30
TRBJ2-1
TRBJ2-5
TRBJ2-7
TRBJ1-2
TRBJ1-6
TRBJ2-6
p24+
CD4
Blood
Lymph Node
D
p24+ cells
Total CD4+ T cells
Blood
Lymph node
TRBJ
TRBV
TRBV/TRBJ families enriched in p24+ cells
TRBV/TRBJ families depleted in p24+ cells
TRBV/TRBJ families similarly distributed in p24+ cells

Figure S6. The TCR repertoire of p24$^+$ cells is biased, Related to Figure 5.

A. Number of TCR reads and clonotypes in paired blood and lymph nodes samples retrieved after TCR bulk sequencing of total CD4$^+$ T cells. **B.** Frequency of TRBV and TRBJ segment usage for the clonotypes identified by TCRβ sequencing in p24$^+$ cells and in total CD4$^+$ T cells in both blood and lymph nodes. Each bar represents a single participant and is colored according to the stage of infection. **C.** Significant differences in the frequencies of TRBV and TRBJ segment usage for the clonotypes identified by TCRβ sequencing in p24$^+$ cells and in total CD4$^+$ T cells. (Wilcoxon; $p<0.05$, *; $p<0.01$**; $p<0.001$, ***; $p<0.0001$ ****). **D.** Association between the TRBV and TRBJ segments in p24$^+$ and total CD4$^+$ cells from blood and lymph nodes from all participants. The circular axis represents the number of clonotypes in each TRBV (bottom) and TRBJ gene families (top). TRBV and TRBJ genes that were significantly increased/decreased in p24$^+$ cells are highlighted with a red or blue background, respectively and so is the association link starting from the TRBV gene (Wilcoxon, as defined in C.).

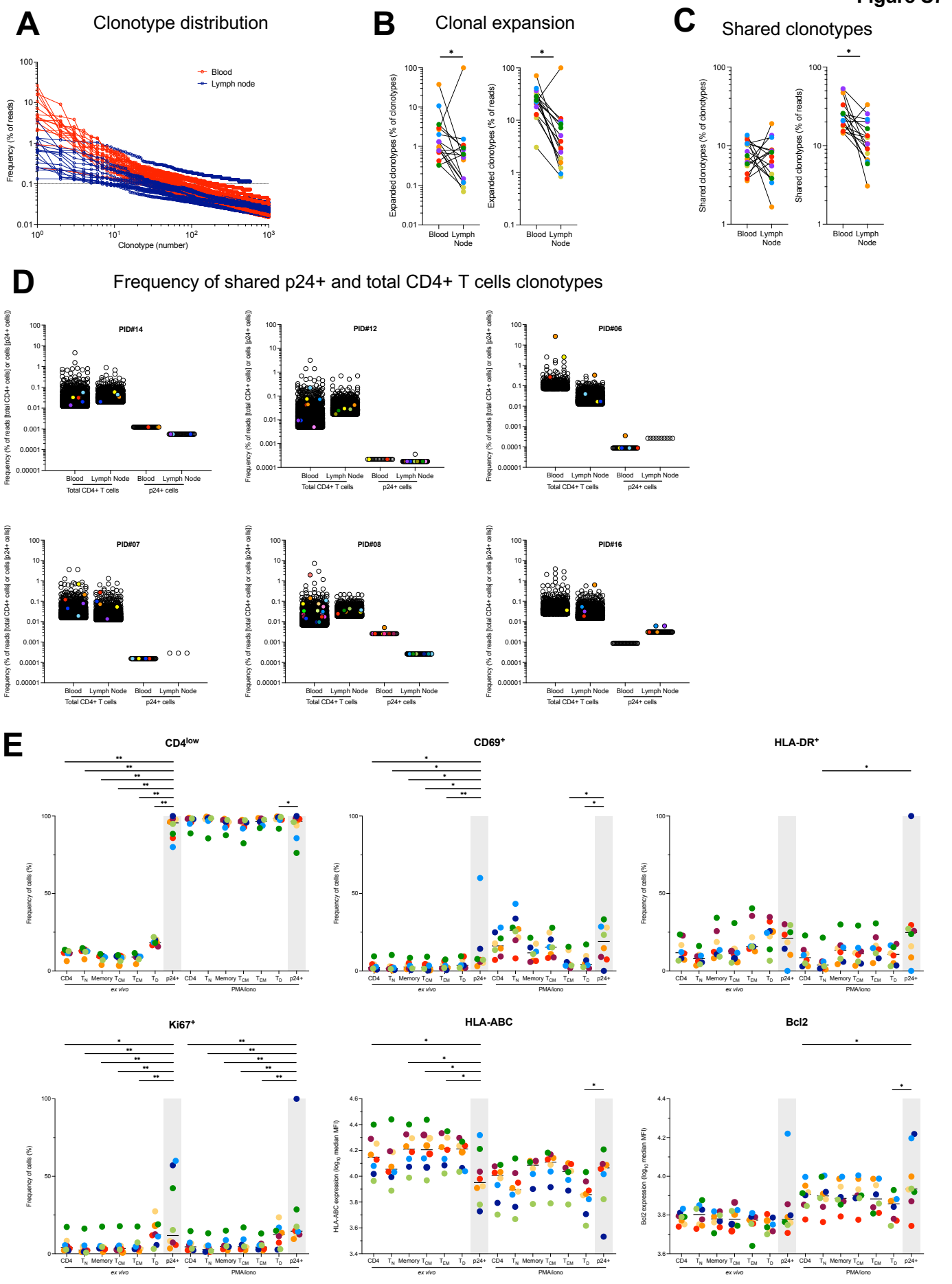

A
Clonotype distribution
Blood
Lymph node
Frequency (% of reads)
Clonotype (number)
B
Clonal expansion
Expanded clonotypes (% of clonotypes)
Expanded clonotypes (% of reads)
Blood
Lymph Node
C
Shared clonotypes
Shared clonotypes (% of clonotypes)
Shared clonotypes (% of reads)
D
Frequency of shared p24+ and total CD4+ T cells clonotypes
PID#14
PID#12
PID#06
PID#07
PID#08
PID#16
Frequency (% of reads [total CD4+ cells] or cells [p24+ cells])
Total CD4+ T cells
p24+ cells
E
CD4low
CD69+
HLA-DR+
Ki67+
HLA-ABC
Bcl2
Frequency of cells (%)
HLA-ABC expression (log10 median MFI)
Bcl2 expression (log10 median MFI)
CD4
TN
Memory
TCM
TEM
TD
p24+
ex vivo
PMA/iono

Figure S7. Distribution of p24$^+$ and total CD4$^+$ T cells TCR repertoires in blood and lymph nodes, Related to Figures 5 and 6.

A. Distribution of clonotypes in total CD4$^+$ T cells from blood and lymph nodes in all participants. Frequencies are shown as percentage of total reads. Dotted line represents the 0.1% frequency threshold. **B.** Frequency of the number and of the percentage of reads of likely expanded clonotypes (>0.1% of total reads) among all detected clonotypes in paired blood and lymph nodes samples, respectively. **C.** Common clonotypes between blood and lymph node samples. Frequency of clonotypes and of reads of common blood/lymph node clonotypes among all clonotypes detected, in blood and lymph node samples, respectively. **D.** Distribution (based on bulk deep sequencing data and single-cell sorting/Sanger sequencing) of clonotypes corresponding to the clones that were found in a single subset (empty circles) and in multiple subsets including p24$^+$ cells (colored circles). Frequencies are shown as percentage of total reads for total CD4$^+$ T cells and as the frequency of cells determined by HIV-Flow for p24$^+$ cells. (Wilcoxon; $p<0.05$, *; $p<0.01$**; $p<0.001$, ***; $p<0.0001$ ****). **E.** The frequency of CD4 expression (CD4low), activation markers expression (CD69$^+$, HLA-DR$^+$, Ki67$^+$) and median fluorescence intensity (MFI) of HLA-ABC and Bcl2 is depicted ex vivo or after stimulation with PMA/ionomycin for each participant according to the cell population (total CD4$^+$, naïve (TN), total memory (CD45RA$^-$), central memory (TCM), effector memory (TEM), terminally differentiated (TD) and p24$^+$ cells). Statistically significant comparisons with p24$^+$ cells are highlighted (Wilcoxon; $p<0.05$, *; $p<0.01$**; $p<0.001$, ***; $p<0.0001$ ****).

Table S1

PID#	Analysis				Fiebig stage	Age at Week 0 (years)	Gender	HIV subtype	Week 0 (Viremic)			Week 96 (ART)		
	Pheno	**TCR**	**Env**	**NFL**					CD4 (/mm3)	CD8 (/mm3)	HIV-RNA (copies/mL)	CD4 (/mm3)	CD8 (/mm3)	HIV-RNA (copies/mL)
PID01	X	X			1	28	Male	CRF01_AE	572	334	37142	NA	NA	NA
PID02	X	X			1	25	Male	CRF01_AE	577	346	3385	NA	NA	NA
PID03	X	X			1	26	Male	CRF01_AE	575	513	11969	NA	NA	NA
PID04	X	X			2	27	Male	CRF01_AE	158	202	2024617	NA	NA	NA
PID05	X	X			2	23	Male	CRF01_AE	283	226	6029465	NA	NA	NA
PID06	X	X			2	26	Male	CRF01_AE/B Recombinant	278	319	6046140	NA	NA	NA
PID07	X	X			3	22	Male	CRF01_AE/B Recombinant	256	352	6880660	NA	NA	NA
PID08	X	X			3	25	Male	CRF01_AE/B Recombinant	509	1045	4631002	NA	NA	NA
PID09	X	X			3	28	Male	CRF01_AE	656	820	4436997	NA	NA	NA
PID10	X	X			4	25	Male	CRF01_AE	346	677	366196	NA	NA	NA
PID11	X	X			4	34	Male	CRF01_AE	456	4556	6596826	NA	NA	NA
PID12	X	X			4	21	Male	B	236	370	110572	NA	NA	NA
PID13	X	X			5	40	Male	CRF01_AE	374	1869	686329	NA	NA	NA
PID14	X	X		X	5	39	Male	CRF01_AE	625	3936	277506	632	610	<20
PID15	X	X			Chronic	50	Male	Not Done	333	1941	21377	NA	NA	NA
PID16	X	X		X	Chronic	21	Male	Not Done	227	688	63588	Not Done	Not Done	<34
PID17	X	X			Chronic	26	Male	Not Done	616	1170	55844	NA	NA	NA
PID18	X		X		1	24	Male	CRF01_AE	457	486	146194	NA	NA	NA
PID19	X		X		1	25	Male	CRF01_AE	537	242	11674	NA	NA	NA
PID20	X		X		2	23	Male	CRF01_AE	308	941	6010006	NA	NA	NA
PID21	X		X		2	20	Male	CRF01_AE	307	297	68393	NA	NA	NA
PID22	X		X		3	36	Male	CRF01_AE	488	3269	1107430	NA	NA	NA
PID23	X		X		3	41	Male	CRF01_AE/B Recombinant	552	2036	7895089	NA	NA	NA
PID24	X		X		5	20	Male	CRF01_AE	233	2531	5541200	NA	NA	NA
PID25	X		X		Chronic	20	Male	CRF01_AE	345	865	19986	NA	NA	NA
PID26				X	2	26	Male	CRF01_AE/B Recombinant	249	430	736615	397	492	<20
PID27				X	2	22	Male	CRF01_AE	641	481	11350800	1204	1058	<20
PID28				X	3	22	Male	CRF01_AE	233	172	4416897	534	367	<20
PID29				X	3	27	Male	CRF01_AE/B Recombinant	457	640	5286145	598	781	<20
PID30				X	4	22	Male	CRF01_AE	571	552	14673	960	757	<20
PID31				X	Chronic	45	Male	CRF01_AE	30	708	723346	267	1244	<34

Table S1. Participants' characteristics, Related to Figure 1.

M, male; ART, antiretroviral therapy; TCR; T-cell receptor sequencing substudy; Env, HIV Env C2-V5 sequencing substudy; NFL, near-full length HIV genome sequencing substudy; NA, not applicable.

Name	Sequence 5'-3'
PCR1 : Forward primers (tagged with M13F)	
VB2	GTAAAACGACGGCCAGTATACTTCTATTGGTACAGACAAATCTTGG
VB3	GTAAAACGACGGCCAGTCTATGTATTGGTATAAACAGGACTCTAAG
VB4	GTAAAACGACGGCCAGTCAYARSGCTATGTATTGGTACAAGC
VB5/9	GTAAAACGACGGCCAGTCACTGTGTCCTGGTACCAACAG
VB6	GTAAAACGACGGCCAGTTACATGTACTGGTATCGACAAGACC
VB7	GTAAAACGACGGCCAGTTACCCTTTATTGGTACCGACAGAGCCTGG
VB11	GTAAAACGACGGCCAGTCTTTACTGGTACCGGCAGAWCYTGG
VB12	GTAAAACGACGGCCAGTTTTTCTGGTACAGACAGACCATGATG
VB13	GTAAAACGACGGCCAGTCACTGTCTACTGGTACCAGCAGG
VB14	GTAAAACGACGGCCAGTTGGACATGATAATCTTTATTGGTATCGAC
VB15	GTAAAACGACGGCCAGTCATGTACTGGTACCAGCAGAAGTC
VB16	GTAAAACGACGGCCAGTGTTATGTTTTTTGGTACCAACAGGTCC
VB17	GTAAAACGACGGCCAGTCATGTTTGTTCACTGGTACCGACAGAATC
VB18	GTAAAACGACGGCCAGTAGTCATGTTTACTGGTATCGGCAGC
VB19	GTAAAACGACGGCCAGTTGCCATGTACTGGTACCGACAG
VB20	GTAAAACGACGGCCAGTCACAACTATGTTTTGGTATCGTCAG
VB21	GTAAAACGACGGCCAGTTAGTTATGTTTACTGGTATCATAAGACGC
VB23	GTAAAACGACGGCCAGTATACTTTTTGTTTATTGGTATCAACAGAATCAG
VB24	GTAAAACGACGGCCAGTATGTACTGGTATCGACAAGACCC
VB25	GTAAAACGACGGCCAGTTGACAAAATGTACTGGTATCAACAAGATC
VB29	GTAAAACGACGGCCAGTTGATGTTCTGGTACCGTCAGCAAC
VB30	GTAAAACGACGGCCAGTCAACCTATACTGGTACCGACAGG
PCR1 : Reverse primers (tagged with M13R)	
JB1-1	CAGGAAACAGCTATGACCAACTGTGAGTCTGGTGCCTTGTCCAAAG
JB1-2	CAGGAAACAGCTATGACAACCTGGTCCCCGAACCGAAGG
JB1-3	CAGGAAACAGCTATGACAACAGTGAGCCAACTTCCCTCTCCAAAATA
JB1-4	CAGGAAACAGCTATGACCAGAGAGCTGGGTTCCACTGCCAAAAAACA
JB1-5	CAGGAAACAGCTATGACAGAGTCGAGTCCCATCACCAAAATGC
JB1-6	CAGGAAACAGCTATGACCTGGTCCCATTCCCAAAGTGGAGG
JB2-1	CAGGAAACAGCTATGACAGCCGTGTCCCTGGCCCGAAGAAC
JB2-2	CAGGAAACAGCTATGACCGTTTTTTGGAGAAGGCTCTAGGCTGACC
JB2-3	CAGGAAACAGCTATGACCAGCCGGGTGCCTGAGCCAAAATAC
JB2-4	CAGGAAACAGCTATGACCGGGTCACGGCGCCGAAGTAC
JB2-5	CAGGAAACAGCTATGACAGCCGCGTCCTGGCCCGAAG
JB2-6	CAGGAAACAGCTATGACCTGCCGGCCCCGAAAGTCAGG
JB2-7	CAGGAAACAGCTATGACCCTGGTGCCCGACCCGAAG
PCR2 and Sequencing: Forward and Reverse primers	
M13F	GTAAAACGACGGCCAGT
M13R	CAGGAAACAGCTATGAC
PCR3 (MiSeq adaptors) : Forward and Reverse primers	
CS1-M13F	ACACTGACGACATGGTTCTACAGTAAAACGACGGCCAGT
CS2-M13R	TACGGTAGCAGAGACTTGGTCTCAGGAAACAGCTATGAC
PCR1 : HIV Env AE C2-V5	
Env7	AATGGCAGTCTAGCAGAAG
OutV5R_AE	TCCTAGTGGTTCAATTTGTACTAC
PCR2 and Sequencing : HIV Env AE C2-V5	
Env7	AATGGCAGTCTAGCAGAAG
InV5R_AE	ACTTCTCCAATTGTCCTTTA
PCR1: near full-length HIV AE	
263-AE-F	AGGGACTCGAAAGCGRAAGT
BlouterR	TGAGGGATCTCTAGTTACCAGAGTC
PCR2: near full-length HIV AE	
652-AE-F	ACTCGAAAGCGRAAGTTCCAGAG
280-AE-R	CTAGTTACCAGAGTCCTAACACAGAYG

Table S2. Primers used for amplification and sequencing of TCRβ, HIV env and near-full length HIV genomes, Related to Figures 3, 4 and 7.