Science Translational Medicine
AAAS

stm.sciencemag.org/cgi/content/full/11/520/eaax0904/DC1

# Supplementary Materials for

## Overexpression of T-bet in HIV infection is associated with accumulation of B cells outside germinal centers and poor affinity maturation

James W. Austin, Clarisa M. Buckner, Lela Kardava, Wei Wang, Xiaozhen Zhang, Valerie A. Melson, Ryan G. Swanson, Andrew J. Martins, Julian Q. Zhou, Kenneth B. Hoehn, J. Nicholas Fisk, Yiannis Dimopoulos, Alexander Chassiakos, Sijy O'Dell, Margery G. Smelkinson, Catherine A. Seamon, Richard W. Kwan, Michael C. Sneller, Stefania Pittaluga, Nicole A. Doria-Rose, Adrian McDermott, Yuxing Li, Tae-Wook Chun, Steven H. Kleinstein, John S. Tsang, Constantinos Petrovas, Susan Moir*

*Corresponding author. Email: smoir@niaid.nih.gov

Published 27 November 2019, *Sci. Transl. Med.* **11**, eaax0904 (2019)
DOI: 10.1126/scitranslmed.aax0904

**The PDF file includes:**

Fig. S1. Additional histocytometric analyses of HIV-infected and HIV-uninfected LN sections.
Fig. S2. Additional analyses of LN B cell populations.
Fig. S3. Additional RNA-seq analyses.
Fig. S4. Resident LN B cell populations have differences in BCR mutation frequency and diversity.
Fig. S5. Characterization of intraclonal relationships.
Table S1. Differential gene expression analysis for selected genes and populations.
Table S2. Antibodies used in imaging analyses.
Table S3. HIV-infected participants for additional phenotyping.
References (*72*–*80*)

**Other Supplementary Material for this manuscript includes the following:**

(available at stm.sciencemag.org/cgi/content/full/11/520/eaax0904/DC1)

Data file S1 (Microsoft Excel format). Primary data.

## Materials and Methods

### Phenotypic analyses

Phenotyping was performed on LNMC of individuals described in Table 1 and for certain stains, on an additional three HIV-infected participants described in table S3 and identified by the color-coding indicated. The majority of stains were performed on freshly isolated LNMC with exception of certain T-bet stains that were performed on LNMC that had been cryopreserved in liquid nitrogen. Cell-surface stains were performed using combinations of the following anti-human mAbs: CD20-APC-H7 (clone 2H7), IgD-FITC (clone IA6-2), CD38-BV421 (clone HIT2), CD11c-BUV395 (clone B-ly6), CD95-PerCP-Cy5.5 (clone DX2), CXCR3-APC (clone 1C6), IgG-PE-Cy7 (clone G18-145), CD3-APC-H7 (clone SK7), CXCR4-PE (clone 12G5) (BD Biosciences); CD3-BV650 (clone OKT3), CD21-FITC (clone BU32), IgD-BV605 (clone IA6-2), CD4-BV510 (clone OKT4) (BioLegend); CD19-PerCP-Cy5.5 (clone SJ25C1), CD27-PECy7 (clone O323) (ThermoFisher); and CXCR5-APC (clone 51505) (R&D Systems). For intracellular staining, cells were fixed with Lysing Solution (BD Biosciences), permeabilized with Permeabilizing Solution 2 (BD Biosciences) and then stained with T-bet-PE (clone eBio4B10) (ThermoFisher), Blimp-1-PE (clone 3H-2-E8) (Novus) or Bcl6-PECy7 (clone K112-91) (BioLegend). HIV-specific B cells were identified using streptavidin-APC (ThermoFisher) conjugated trimeric HIV envelope gp140 from YU2, as previously described (*53*). For $IgD^+$ cells isolated from tonsil or LN, purities, typically >90%, were assessed using antibodies listed above as well as the following anti-human mAbs: CD27-BUV395 (clone L128) (BD Biosciences) and IgM-BV510 (clone MHM-88) (BioLegend). At baseline and 48 hr post-stimulation, cells were also stained for T-bet and Bcl-6 (as described above) and CXCR5-BV605 (clone J252D4) (BioLegend).

**Multiplex confocal imaging and histo-cytometry**

Formalin-fixed, paraffin-embedded LN tissue blocks were cut into 5 µm sections and mounted onto slides. Several panels were developed for these analyses (panel details shown in table S2). For staining, slides were deparaffinized and tissue rehydrated in deionized water. Antigen retrieval was performed using a decloaking chamber (Biocare Medical) (110$^{o}$C, 15 min) with Borg Decloaker buffer (Biocare Medical) followed by a cooling step to room temperature (RT). After blocking and permeabilization for 1 hour (0.1M tris, 0.3% TritonX-100, 1% Bovine Serum Albumin), sections were stained with titrated amounts of non-conjugated antibodies, followed by overnight incubation at 4$^{o}$C. Slides were then washed with PBS (3 times, 20 min each) and stained with the appropriate secondary antibodies for 2 hours at RT. Slides were washed and after another blocking step for 1 hour at RT with a 1:10 dilution of normal mouse and rabbit or goat serum, slides were stained for 2 hours RT with titrated amounts of directly conjugated antibodies. After a final washing step and staining with nuclear marker JoJo (ThermoFisher), slides were then mounted with Fluoromount G (Southern Biotech) and sealed with a glass coverslip. Confocal images were obtained using a SP5 X-WLL confocal microscope (Leica) using a 40X 1.30NA objective with a 1.5x optimal zoom and 1,024 x 1,024 pixel density. Images are presented as maximum projections of tiled z-stacks.

Multiparameter confocal images were processed analytically using histo-cytometry (*72, 73*). Briefly, fluorophore spectral spillover was corrected by imaging single-stained tissue controls and creating a compensation matrix using the Channel Dye Separation module within the LASX software (Leica) that was then used to correct for crosstalk. Corrected images were analyzed with Imaris software (Bitplane Scientific). Threshold identification, voxel gating,

surface creation and masking were performed as previously described (*72, 73*). The nuclear marker JoJo (ThermoFisher) was used to create surfaces. Average voxel intensities for all channels of interest within the surfaces, along with the X,Y positioning of the cell centroids were exported into Excel, converted to a csv file and imported into FlowJo v10 (TreeStar, Inc.) for direct visualization and gating of distinct LNMC populations. To define zones, gates for B cell follicles and GC were established using positional data on $CD20^+$ and $CD20^+/Ki67^+$ or $CD20^+/Bcl6^+$surfaces, respectively. To calculate the density of $CD20^+$ and $CD20^+$T-$bet^+$ cells per GC, the surface function in Imaris was used to create volumetric surfaces. The number of cells were then divided by the areas obtained from these regions. The number of cells in the GC zone was determined by gating on $Ki67^+IgD^-$ cells using FlowJo. Non-GC zones were defined by subtracting the area of the GC from the area of the follicle. Similarly, the number of cells in the non-GC zone was determined with FlowJo by subtracting the number of cells in the GC from that of the follicle. The extrafollicular zone was calculated by adding together non-follicular areas concentrated in CD4- or CD8-expressing cells and using Boolean gating in FlowJo. For evaluation of various markers, frequencies of $CD20^+$T-$bet^+$ and $CD20^+$T-$bet^-$ cells within follicles were first determined by GC and non-GC area, as well as among cells that were $IgD^+$ for the latter area. The percent expression of each marker was extracted from defined areas/populations and displayed in a heatmap using the R program. Distance analysis was performed for each tissue using a Python script implementing the ImarisXT license and ImarisLib module to overlay the coordinates of the positional follicular and germinal center histo-cytometry-derived gates onto the original image in Imaris. The overlay was then manually traced over in Imaris and used to generate surfaces demarcating the germinal center and follicular non-germinal center regions of the image. Another Python script using ImarisXT

license and ImarisLib module was used to create points on the original image for the positional coordinates of the entire $CD20^{+}$T-$bet^{+}$ population. Additionally, the Imaris XT Distance Transform module was used to generate the shortest distance in micrometer of each cell in the population to any part of two zones of interest. Cells with centers located within or on the border of a zone were set to have a shortest distance value of zero micrometers. All additional statistical analysis and graphing were performed using Prism (GraphPad).

**Post-hoc analyses for multiple comparisons of the conventional flow cytometric and histo-cytometric data**

For three group analyses of dependent data, multiple testing was performed as previously described (*32*). Briefly, a global Friedman test was performed and if significant, it was followed by the Wilcoxon signed rank. This procedure controls the familywise error rate (FWER), the probability of making at least one false declaration of difference, regardless of which groups are equal. For four group analyses of dependent data (for Blimp-1 in Fig. 2C and in Fig. 3A), a similar approach was taken as for the three group global test, and if significant, six pairwise comparisons were performed and the pairs were declared statistically significant if three times the pairwise p-value was less than or equal 0.05. The significant three-fold adjusted p-values were reported in figures and text accordingly. This procedure controls the FWER regardless of which pairs were equal.

**Antibody production**

Antibodies were produced and tested for specificity as previously described (*53*). Briefly, HIV $gp140^{+}IgG^{+}$ LN MBC and GCBC were single-cell-sorted into 96-well PCR plates and the heavy-

and light-chain Ig genes cloned into IgG1 expression vectors. Following production in Expi293F cells, antibodies were purified using protein G Sepharose 4 Fast Flow beads (GE Healthcare Life Sciences), according to the manufacturer's instructions. Antibodies were quantitated using a FACSarray Bioanalyzer (BD Biosciences) and tested for specificity by ELISA.

**Quantitative RT-PCR**

Total RNA was reversed transcribed using qScript cDNA SuperMix (Quantabio). Real-time PCR was performed using TaqMan gene expression master mix (ThermoFisher) on a 7500 Real-Time PCR System (ThermoFisher) with TaqMan probe/primers (ThermoFisher) for *TBX21* (Hs00203436_m1), *BACH2* (Hs00222364_m1), and *S1PR2* (custom, AI39R4J). Data are presented as relative expression using the ΔΔCt method by normalization to *POLR2A* (Hs00172187_m1).

**Preprocessing of raw BCR sequencing reads**

Demultiplexed pair-end reads were first assigned to different subtypes (IgG1, IgG3, IgG2/4, IgM, or undetermined) based on a 12-bp identifier read using a custom Python script (doi:10.5281/zenodo.3405436) as described in Schanz et al. (*62*). IgG2/4 reads were further resolved to be either IgG2 or IgG4 based on the fourth nucleotide of Read 2. A perfect match was required for identifying the IgG subtypes and a single mismatch was allowed for identifying IgM. Subsequent preprocessing was performed using pRESTO v0.5.8 (http://presto.readthedocs.io/) (*63*). Briefly, pair-end reads with a mean Phred quality score below 30 were filtered. Read 1s were aligned against V region primers and Read 2s against C region primers corresponding to their previously identified isotypes, allowing a maximum

mismatch rate of 0.1. Nucleotides aligned to the primers were then removed. Next, pair-end reads were assembled by first attempting *de novo* assembly, which, if unsuccessful (e.g. due to long CDR3 preventing overlap), was followed by reference-guided assembly. For *de novo* assembly, a minimum overlap length of 8 bps and a maximum error rate of 0.3 were allowed; and a p-value threshold of $1x10^{-5}$ was used. For reference-guided assembly, alignment was performed using USEARCH v9.2.64 (http://www.drive5.com/usearch/) (*74*), against the July 12, 2018 version of the International Immunogenetics Information System (IMGT) V-segment germlines (https://www.imgt.org/genedb/) (*65*), requiring a minimum identity of 0.5 and an E-value threshold of $1x10^{-5}$. Duplicate reads were then collapsed into unique sequences, allowing a maximum of 10 missing nucleotides. Only sequences with at least 2 duplicate reads were used subsequently.

**V(D)J gene annotation and additional quality control**

After preprocessing, initial assignment of V(D)J germline gene annotations was performed using the standalone IgBLAST v1.7.0 (ftp://ftp.ncbi.nih.gov/blast/executables/igblast/release/1.7.0/) (*64*), against the July 12, 2018 version of the IMGT human immunoglobulin gene database (https://www.imgt.org/genedb/) (*65*). IgBLAST output was processed using Change-O v0.4.1 (http://changeo.readthedocs.io/) (*66*). Additional quality control was then performed using custom Python 3 scripts (doi:10.5281/zenodo.3405436). Specifically, only sequences that successfully aligned exclusively to heavy-chain V genes were retained, removing those that aligned to light-chain V genes or to both heavy- and light-chain V genes. Furthermore, sequences were filtered by requiring a minimum V-segment coverage from nucleotide position 1 to 312 under the IMGT unique numbering scheme (*75*), and allowing a maximum of 10 Ns in the V-

segment. Next, individual genotypes, including novel V gene alleles missing from the IMGT database, were computationally inferred using TIgGER v0.2.11 (http://tigger.readthedocs.io/) (*67*), and used to finalize V(D)J annotations. Additionally, to remove chimerics, sequences containing more than 5 mismatches from the germline in any 10-bp window were identified using SHazaM v0.1.9 (http://shazam.readthedocs.io/) (*66*), and removed. Finally, non-productively rearranged sequences (labeled "non-functional" by IgBLAST) were removed from further analysis.

**Clonal clustering**

Productively rearranged heavy-chain sequences were clustered into clonally-related lineages. First, sequences were partitioned based on common *IGHV* gene annotations, *IGHJ* gene annotations, and junction region lengths. Within each partition, pairwise Hamming distances between the junction regions (defined from IMGT codon 104 encoding the conserved cysteine to codon 118 encoding phenylalanine or tryptophan) of the sequences were calculated and normalized by the junction region length. In a given partition, sequences whose junction regions were within a normalized Hamming distance of 0.1 from each other were defined as clones via hierarchical clustering with single linkage (*49*). The clonal distance threshold of 0.1 was determined by manual inspection in conjunction with kernel density estimates using a subsample containing 15,000 sequences (*66*), in order to identify the local minimum between the two modes of the within-individual bimodal distance-to-nearest distribution (*49*) (fig. S1C). Following clonal clustering, full-length clonal-consensus germline sequences were reconstructed for each clone with D-segment and N/P regions masked with ambiguous (N) nucleotides, resolving any ambiguous gene assignments by majority rule (*66*). Lastly, using Alakazam v0.2.10

(http://alakazam.readthedocs.io/) (*66*), duplicate IMGT-gapped V(D)J sequences within each clone were collapsed, with the exception of duplicates derived from different cell subsets and/or isotypes. The percentage of clonal overlap was calculated by dividing the number of shared clones or unique sequences by the total in the smaller of the two samples.

**Calculation of mutation frequency**

For each processed sequence, mutation frequency for the IMGT-numbered V-segment from nucleotide position 1 through 312 was calculated using SHazaM v0.1.9 (http://shazam.readthedocs.io/) (*66*), by counting the number of mismatches from the germline sequence.

**Calculation of clonal diversity**

Clonal diversity analysis was performed using the "rarefyDiversity" function from Alakazam v0.2.10 (https://alakazam.readthedocs.io/) (*66*). Generalized Hill diversity index ($^{q}D$) (*76*) was calculated over a range of diversity orders (q) for each participant-cell-type combination. To correct for varying sequencing depths (*77*) and allow for comparison across participants, for each participant-cell-type combination, 1000 bootstraps were performed to repeatedly and uniformly subsample 33,626 sequences, including duplicates.

**Phylogenetic analysis of cellular lineage relationships**

The BuildTrees.py script within Change-O v0.4.3 (http://changeo.readthedocs.io/) was used for each clone to convert Change-O formatted tab files into individual FASTA-formatted files (*66*). Sequences labeled as naïve were removed, leaving only GCBC and MBC types. All IMGT-

aligned sequences that differed only by N nucleotides and had the same cell type were collapsed to the sequence with the fewest ambiguous characters. All clones with only one cell type or with fewer than 10 unique sequences were excluded from the analysis. Phylogenetic tree topologies for each of the remaining clones were estimated using the Fitch parsimony algorithm (*68*), and "pratchet" optimization function implemented in the R package *phangorn* v2.2.0 (*69*).

A maximum parsimony algorithm was used to label the internal nodes of each lineage tree with a particular cell type. For a given phylogenetic tree labelled with the cell type states at the tips, the set of internal node labels are solved for which minimize the number of label switches along the tree. This was accomplished using the Sankoff parsimony algorithm with equal weights between states (*78*). Furthermore, the parsimony score was used to resolve uncertainties in the given lineage tree. Namely, all clusters of internal nodes separated by zero length branches (polytomies), were re-ordered using nearest-neighbor interchange moves to minimize the number of label switches along the tree beginning with the germline sequence (naïve) and moving down. This approach was based on the PMH-TR problem identified previously (*79*). It is important to note that this procedure resolves each polytomy to a single maximum parsimony configuration which is kept constant, and does not take into account the multiple, equally parsimonious ways in which a polytomy can be resolved. The number of switches between cell types in the internal node labels of these rearranged trees were then counted. For a given tree topology and set of tip labels, it is possible for multiple sets of internal node labels to have the minimum number of switches. In these cases, the backtrace step of the Sankoff algorithm was modified to recursively generate all maximally parsimonious internal node labels for a given tree. 1000 maximally parsimonious labelings were randomly sampled in cases where more than 1000 sequences were present in a tree or more than 1000 maximally

parsimonious labelings were found. The average number of each switch type across all ambiguous labelings was then determined. All features have been implemented in a new release of the B cell phylogenetic software package IgPhyML v1.1 (bitbucket.org/kbhoehn/igphyml) (*80*).

The following steps, similar to those described (*50*), were taken to quantify the significance of the number of observed switch events under the null hypothesis that all B cell lineages are independent instances of the same random, unstructured cell interconversion process: i) randomize tip labels among trees within a repertoire, ii) count the total number of switches of each type, iii) repeat 1000 times for each subject, and iv) calculate the p value for each cell type switch as the proportion of permuted repertoires with either at least or at most the observed number of switches. These two-sided p values were adjusted using the Benjamini-Hochberg procedure, implemented in p.adjust in R (v3.2.3) (*71*).

Figure S1

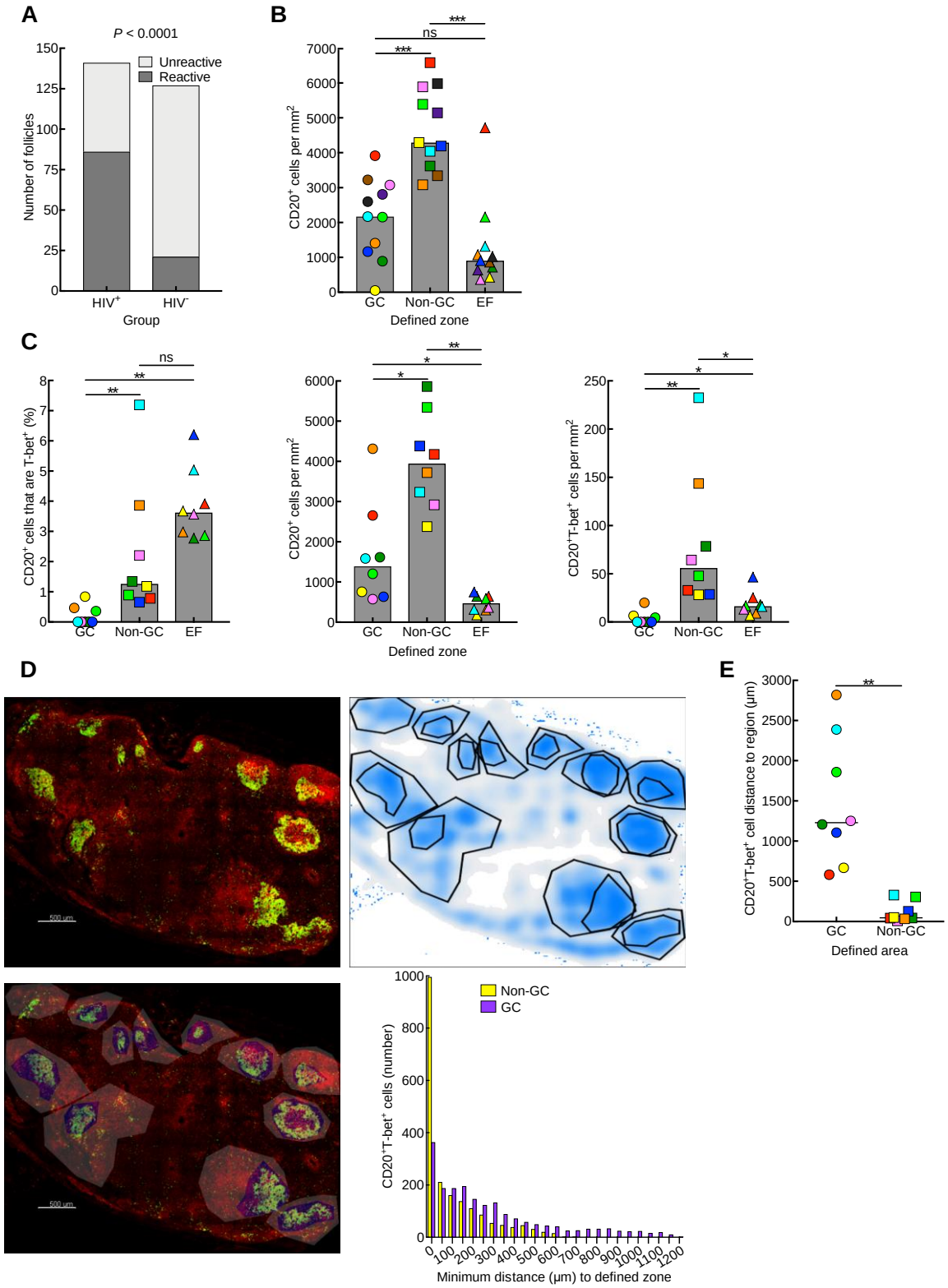

A
P < 0.0001
Unreactive
Reactive
Number of follicles
HIV+
HIV-
Group
B
ns
***
***
CD20+ cells per mm2
GC
Non-GC
EF
Defined zone
C
ns
**
**
CD20+ cells that are T-bet+ (%)
GC
Non-GC
EF
**
*
*
CD20+ cells per mm2
Defined zone
*
*
**
CD20+T-bet+ cells per mm2
D
500 um
500 um
Non-GC
GC
CD20+T-bet+ cells (number)
Minimum distance (μm) to defined zone
E
**
CD20+T-bet+ cell distance to region (μm)
GC
Non-GC
Defined area

**Fig. S1. Additional histocytometric analyses of HIV-infected and HIV-uninfected LN sections.** (**A**) Number of unreactive and reactive follicles in the LN of HIV-infected ($n$ = 11) compared to uninfected ($n$ = 8) individuals. The $P$ value was determined by Fisher's exact test. **(B)** Number of $CD20^+$ cells in each defined zone ($n$ = 11 HIV-infected) quantified by histo-cytometry. **(C)** Frequencies of $CD20^+$ cells expressing T-bet (left graph), and number of $CD20^+$ (middle graph) or $CD20^+$T-$bet^+$ (right graph) cells in each zone ($n$ = 8 HIV-uninfected) quantified by histo-cytometry. **(D)** Strategy for calculating distance of $CD20^+$T-$bet^+$ cells to defined zones using Python script and ImarisXT license in conjunction with ImarisLib and ImarisXT Distance Transform modules showing representative confocal image (top left), followed by definition of areas (top right) and overlay on confocal image (bottom left) and graphing of number of $CD20^+$T-$bet^+$ cells by incremental distance to defined zone. **(E)** Distance of $CD20^+$T-$bet^+$ cells to GC and non-GC zones ($n$ = 8 HIV-uninfected). In (B), each HIV-infected individual is color-coded per Table 1 and shaded bars represent medians. *$P$ < 0.05, **$P$ < 0.01, ***$P$ < 0.001 by Wilcoxon matched-pairs signed rank test after obtaining significance by Friedman ANOVA test on full set; ns, not significant.

Figure S2

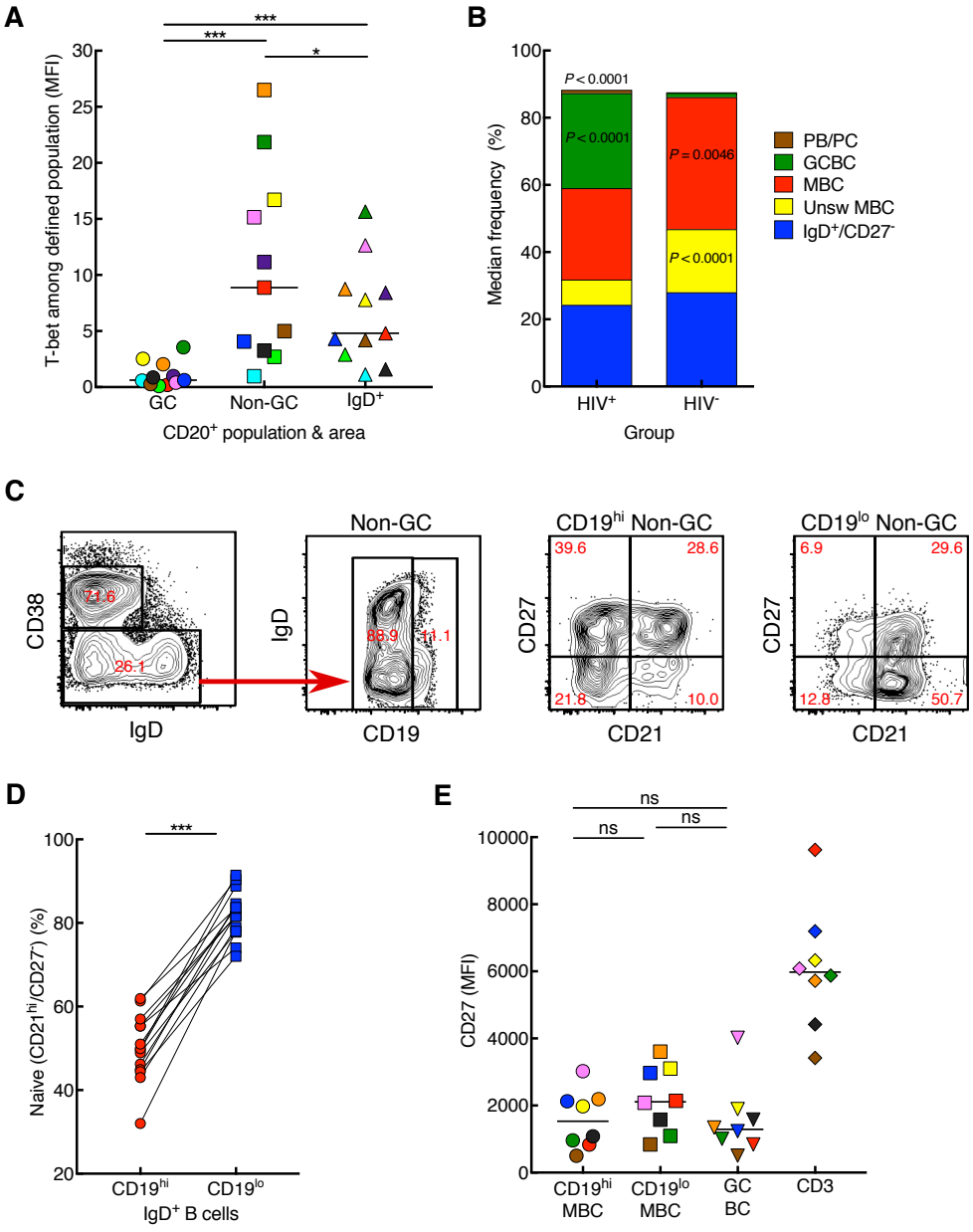

A
***
***
*
T-bet among defined population (MFI)
30
25
20
15
10
5
0
GC
Non-GC
IgD+
CD20+ population & area
B
100
80
60
40
20
0
Median frequency (%)
P < 0.0001
P < 0.0001
P = 0.0046
P < 0.0001
HIV+
HIV-
Group
PB/PC
GCBC
MBC
Unsw MBC
IgD+/CD27-
C
CD38
IgD
73.6
26.1
Non-GC
IgD
CD19
88.9
11.1
CD19hi Non-GC
CD27
CD21
39.6
28.6
21.8
10.0
CD19lo Non-GC
CD27
CD21
6.9
29.6
12.8
50.7
D
***
Naive (CD21hi/CD27+) (%)
100
80
60
40
20
CD19hi
CD19lo
IgD+ B cells
E
ns
ns
ns
CD27 (MFI)
10000
8000
6000
4000
2000
0
CD19hi MBC
CD19lo MBC
GC BC
CD3

**Fig. S2. Additional analyses of LN B cell populations. (A)** Frequencies of $CD20^{+}$T-$bet^{+}$ cells among defined population and area in LN from HIV-infected individuals ($n$ = 11) as measured by histo-cytometry. (**B**) Distribution of B cell subpopulations (defined as in Fig. 3B where PB/PC are $CD38^{++}IgD^{-}$) in HIV-infected and uninfected individuals evaluated by conventional flow cytometry on LNMC. The *P* values, shown for each subpopulation in the group with the higher frequency, were determined by the Mann-Whitney test. **(C)** Conventional flow cytometry performed on LNMC isolated from an HIV-infected individual depicts delineation of $CD19^{hi}$ and $CD19^{lo}$ non-GC B cells used to evaluate expression of CD21 by CD27. **(D** Frequencies of $CD21^{hi}/CD27^{-}$ naïve cells among $CD19^{hi}$ and $CD19^{lo}$ $IgD^{+}$ B cells in HIV-infected individuals evaluated by conventional flow cytometry on LNMC. ***$P < 0.001$ by Wilcoxon matched-pairs signed rank test. **(E)** MFI for CD27 by B cell population or $CD3^{+}$ cells performed by conventional flow cytometry on LNMC ($n$ = 12). In (A) and (E), each HIV-infected individual is color-coded per Table 1 and black horizontal bars represent medians. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$ by Wilcoxon matched-pairs signed rank test after obtaining significance by Friedman ANOVA test on full set; ns, not significant.

Figure S3

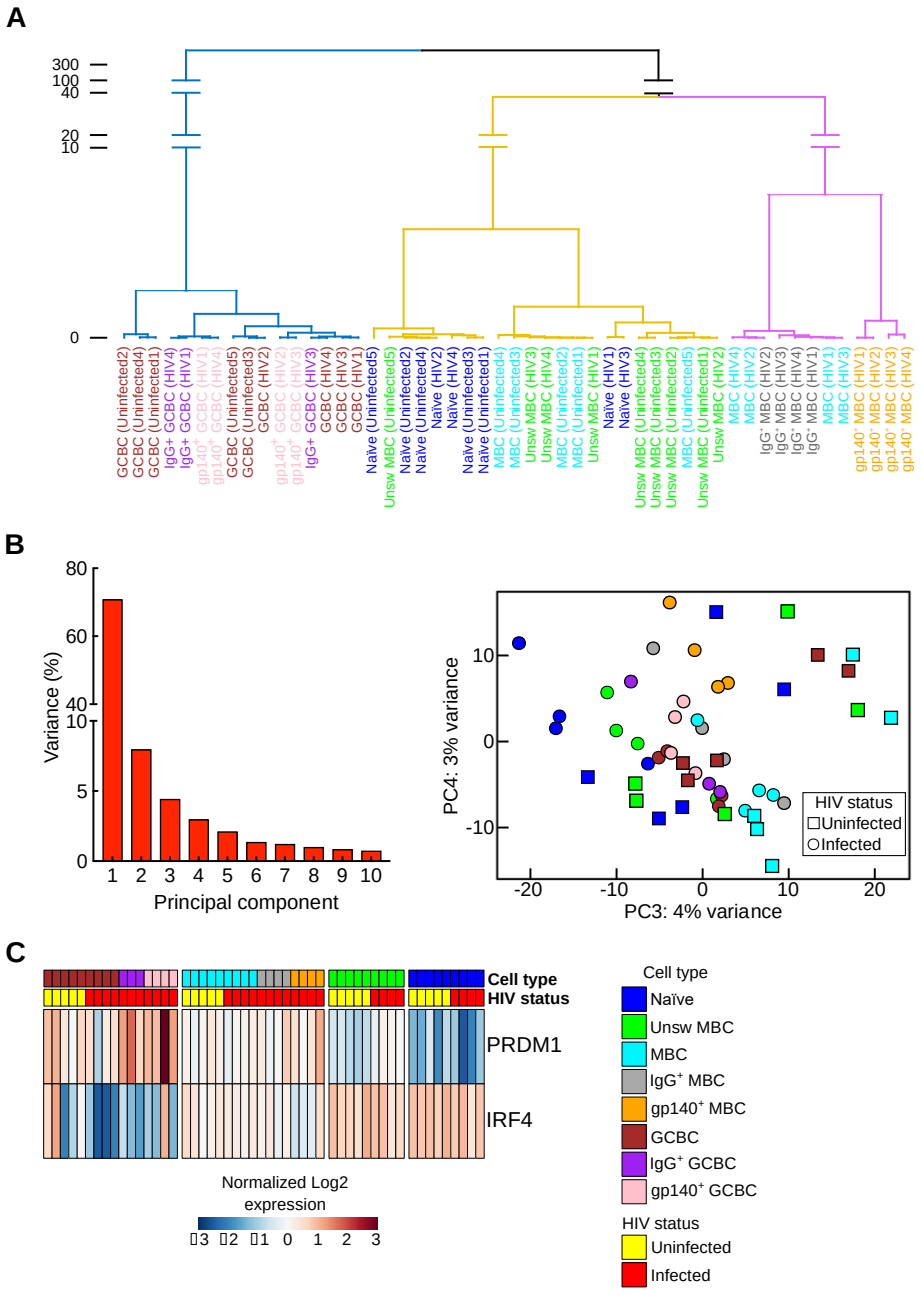

A
300
100
40
20
10
0
GCBC (Uninfected2)
GCBC (Uninfected4)
GCBC (Uninfected1)
IgG+ GCBC (HIV4)
IgG+ GCBC (HIV1)
gp140+ GCBC (HIV1)
gp140+ GCBC (HIV4)
GCBC (Uninfected5)
GCBC (Uninfected3)
GCBC (HIV2)
gp140+ GCBC (HIV2)
gp140+ GCBC (HIV3)
IgG+ GCBC (HIV3)
GCBC (HIV4)
GCBC (HIV3)
GCBC (HIV1)
Naïve (Uninfected5)
Unsw MBC (Uninfected5)
Naïve (Uninfected2)
Naïve (Uninfected4)
Naïve (HIV2)
Naïve (HIV4)
Naïve (Uninfected3)
Naïve (Uninfected1)
MBC (Uninfected4)
MBC (Uninfected3)
Unsw MBC (HIV3)
Unsw MBC (HIV4)
MBC (Uninfected2)
MBC (Uninfected1)
Unsw MBC (HIV1)
Naïve (HIV1)
Naïve (HIV3)
Unsw MBC (Uninfected4)
Unsw MBC (Uninfected3)
Unsw MBC (Uninfected2)
MBC (Uninfected5)
Unsw MBC (Uninfected1)
Unsw MBC (HIV2)
MBC (HIV4)
MBC (HIV2)
IgG+ MBC (HIV2)
IgG+ MBC (HIV3)
IgG+ MBC (HIV4)
IgG+ MBC (HIV1)
MBC (HIV1)
MBC (HIV3)
gp140+ MBC (HIV1)
gp140+ MBC (HIV2)
gp140+ MBC (HIV3)
gp140+ MBC (HIV4)
B
Variance (%)
80
60
40
10
5
0
1 2 3 4 5 6 7 8 9 10
Principal component
PC4: 3% variance
10
0
-10
-20 -10 0 10 20
PC3: 4% variance
HIV status
Uninfected
Infected
C
Cell type
HIV status
PRDM1
IRF4
Normalized Log2
expression
3 2 1 0 1 2 3
Cell type
Naïve
Unsw MBC
MBC
IgG+ MBC
gp140+ MBC
GCBC
IgG+ GCBC
gp140+ GCBC
HIV status
Uninfected
Infected

**Fig. S3. Additional RNA-seq analyses. (A)** Hierarchical clustering of all sorted LN B cell populations used in RNA-Seq by principal components 1 and 2. **(B)** Percent of the variance between RNA-Seq samples explained by the first 10 principal components (left) and principal component analysis using PC3 and PC4 (right). **(C)** Heatmap depicting scaled, normalized $\log_2$ expression values of *PRDM1* (Blimp-1) and *IRF4*.

Figure S4

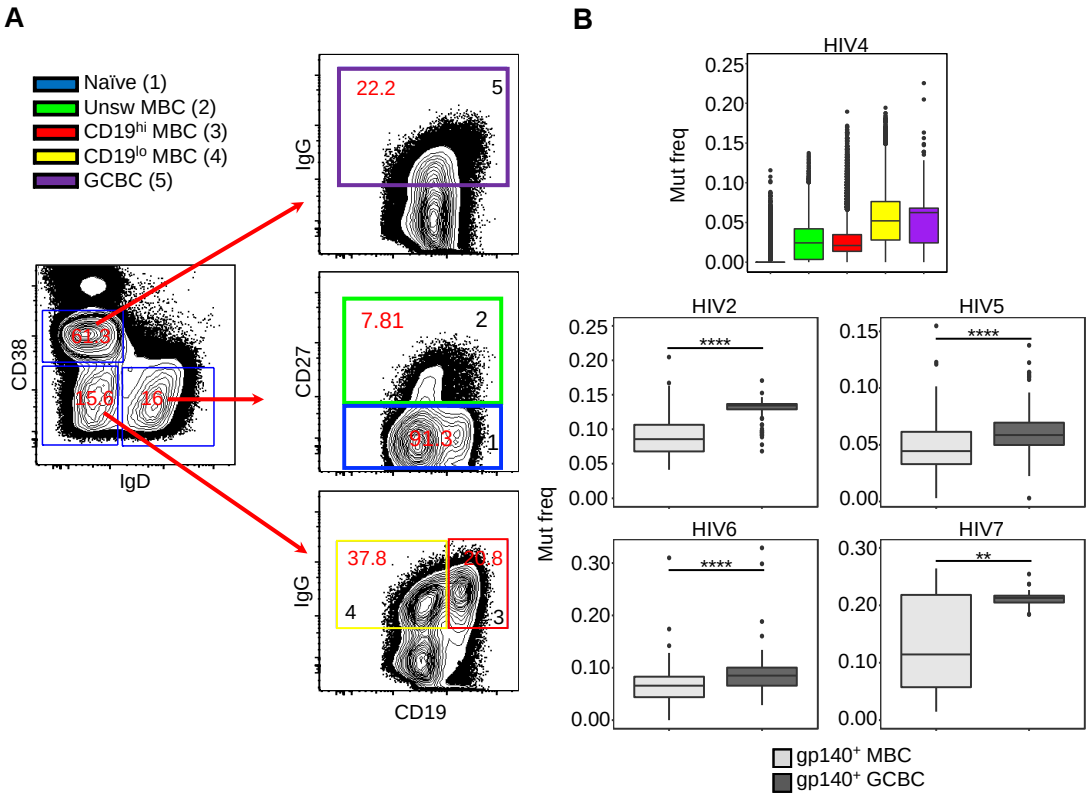
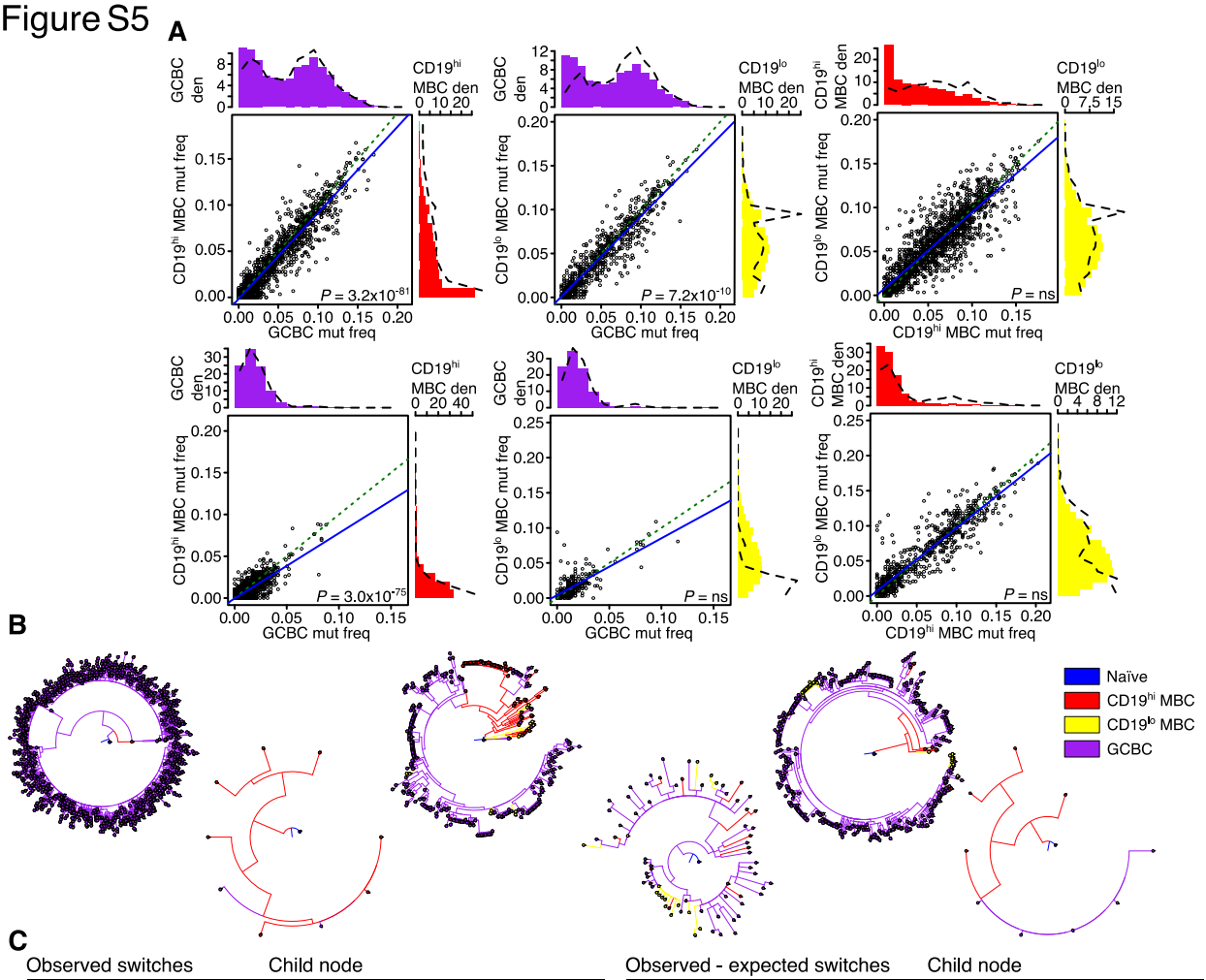

A
Naïve (1)
Unsw MBC (2)
CD19hi MBC (3)
CD19lo MBC (4)
GCBC (5)
CD38
IgD
IgG
CD27
CD19
22.2
5
7.81
2
91.3
1
37.8
4
20.8
3
B
HIV4
Mut freq
HIV2
HIV5
HIV6
HIV7
****
**
gp140+ MBC
gp140+ GCBC
C
HIV1
HIV2
HIV3
HIV4
Density
0.1
D
HIV1
HIV2
HIV3
qD
q
Naïve
Unsw MBC
CD19hi MBC
CD19lo MBC
GCBC

**Fig. S4. Resident LN B cell populations have differences in BCR mutation frequency and diversity.** (**A**) Sorting strategy for isolating LN naïve B cells, class-unswitched MBC, $IgG^+CD19^{hi}$ MBC, $IgG^+CD19^{lo}$ MBC, and $IgG^+$ GCBC used for BCR sequencing. Numbers in legend refer to populations gated and color-coded in flow plots. (**B**) *IGHV* mutation frequencies (mut freq) of LN B cell populations sorted from HIV-infected participant HIV4 (top) and of HIV $gp140^+$ MBC and GCBC single-cell sorted from four HIV-infected individuals (bottom), described in Table 1. Data are presented as box-and-whiskers plots showing the two inner quartiles (box) with median (black line), 1.5 times the inner quartile range (whiskers) and outliers (black circles). $^{**}P < 0.01$, $^{****}P < 0.0001$ by Mann-Whitney U test. (**C**) Distance-to-nearest plots of Ig heavy chains for four HIV-infected individuals (Table 1). After partitioning sequences based on common *IGHV* gene, *IGHJ* gene, and junction region length, the minimum Hamming distance of the junction (normalized by junctional length) to a non-identical sequence within the same partition was calculated for each sequence within a given partition. The fixed threshold used to partition BCR sequences into clones is indicated by the dashed vertical line. (**D**) Generalized Hill diversity index ($^qD$) over a range of diversity orders (q) for different B cell populations in three HIV-infected individuals (Table 1). The curves visualize means from 1000 bootstraps and the bands indicate 95% confidence intervals.

Figure S5




C

Observed switches — Child node

| HIV1 | Naïve | Unsw MBC | CD19hi MBC | CD19lo MBC | GCBC |
|---|---|---|---|---|---|
| Naïve | - | 100.0 | 203.7 | 68 | 387.7 |
| Unsw MBC | 0 | - | 20.0 | 104.0 | 14.0 |
| CD19hi MBC | 0 | 4.0 | - | 63.3 | 167.3 |
| CD19lo MBC | 0 | 12.0 | 43.3 | - | 14.0 |
| GCBC | 0 | 26.0 | 493.0 | 109.7 | - |

Observed - expected switches — Child node

| HIV1 | Naïve | Unsw MBC | CD19hi MBC | CD19lo MBC | GCBC |
|---|---|---|---|---|---|
| Naïve | - | 84.96 | 181.78 | 62.25 | -201.55 |
| Unsw MBC | 0 | - | 11.45 | 90.55 | 5.17 |
| CD19hi MBC | 0 | -7.95 | - | 52.55 | 147.65 |
| CD19lo MBC | 0 | 12.86 | 40.07 | - | 15.27 |
| GCBC | 0 | -1322.62 | -1172.50 | -803.17 | - |

Parent node

| HIV2 | Naïve | Unsw MBC | CD19hi MBC | CD19lo MBC | GCBC |
|---|---|---|---|---|---|
| Naïve | - | 22.0 | 366.0 | 137.0 | 488.0 |
| Unsw MBC | 0 | - | 13.0 | 4.0 | 3.0 |
| CD19hi MBC | 0 | 14.0 | - | 272.0 | 290.0 |
| CD19lo MBC | 0 | 0.0 | 174.0 | - | 58.0 |
| GCBC | 0 | 15.0 | 809.0 | 348.0 | - |

| HIV2 | Naïve | Unsw MBC | CD19hi MBC | CD19lo MBC | GCBC |
|---|---|---|---|---|---|
| Naïve | - | 19.0 | 307.3 | 92.0 | -279.7 |
| Unsw MBC | 0 | - | 11.9 | 3.0 | 2.8 |
| CD19hi MBC | 0 | 7.5 | - | 200.9 | 204.2 |
| CD19lo MBC | 0 | -4.4 | 118.6 | - | 8.5 |
| GCBC | 0 | -167.2 | -1881.6 | -1841.8 | - |

| HIV3 | Naïve | Unsw MBC | CD19hi MBC | CD19lo MBC | GCBC |
|---|---|---|---|---|---|
| Naïve | - | 6.0 | 182.0 | 41.0 | 485.0 |
| Unsw MBC | 0 | - | 2.0 | 0.0 | 1.0 |
| CD19hi MBC | 0 | 2.0 | - | 40.0 | 193.0 |
| CD19lo MBC | 0 | 2.0 | 40.0 | - | 9.0 |
| GCBC | 0 | 25.0 | 872.0 | 193.0 | - |

| HIV3 | Naïve | Unsw MBC | CD19hi MBC | CD19lo MBC | GCBC |
|---|---|---|---|---|---|
| Naïve | - | 3.89 | 141.61 | 24.74 | -129.37 |
| Unsw MBC | 0 | - | 1.92 | -0.08 | 0.97 |
| CD19hi MBC | 0 | -0.20 | - | 20.91 | 131.59 |
| CD19lo MBC | 0 | 1.69 | 28.71 | - | 5.02 |
| GCBC | 0 | -48.90 | -1053.85 | -577.91 | - |

**Fig. S5. Characterization of intraclonal relationships.** (**A**) Scatterplots show the median mutation frequency (mut freq) of the two indicated LN B cell populations for each shared clone from HIV-infected participants HIV2 (top) or HIV3 (bottom), described in Table 1. Green equivalence and blue linear regression lines are shown. The density (den) histograms above and right of the scatterplot show the mutation frequency of all sequences for the indicated population. The overlaid line on each histogram is the mutation frequency for sequences within shared clones between the two populations. *P* values by Wilcoxon signed rank tests with normal approximation and adjusted for multiple testing using Bonferroni correction. (**B**) Examples of phylogenetic trees containing multi-population clonal families with predicted internal node types. (**C**) Table of observed switches in which internal node states have been predicted using maximum parsimony (left) and differences in observed versus expected switches within BCR lineage trees (right). Numbers highlighted in blue and orange are statistically higher or lower (adjusted $P < 0.05$) than expected, respectively. *P* values by permutation test (see Supplemental Materials and Methods, section on phylogenetic analysis of cellular lineage relationships) and adjusted using the Benjamini-Hochberg procedure.

**Table S1. Differential gene expression analysis for selected genes and populations.**

| PRDM1 | log2FoldChange | lfcSE | padj |
|---|---|---|---|
| HIV- MBC vs gp140+ MBC | 0.11 | 0.53 | 1.00 |
| HIV+ MBC vs gp140+ MBC | -0.77 | 0.55 | 0.47 |
| IgG+ MBC vs HIV+ gp140+ MBC | -0.39 | 0.55 | 0.75 |
| IRF4 | | | |
| HIV- MBC vs gp140+ MBC | 0.42 | 0.37 | 0.48 |
| HIV+ MBC vs gp140+ MBC | 0.34 | 0.39 | 0.69 |
| IgG+ MBC vs HIV+ gp140+ MBC | 0.27 | 0.39 | 0.75 |
| TBX21 | | | |
| HIV- Naïve vs HIV+ Naïve | -4.07 | 0.57 | $1.35 \times 10^{-9}$ |
| HIV- Unsw MBC vs HIV+ Unsw MBC | -4.36 | 0.57 | $9.27 \times 10^{-11}$ |
| HIV- MBC vs HIV+ MBC | -3.05 | 0.52 | $1.09 \times 10^{-5}$ |
| HIV- GCBC vs HIV+ GCBC | -1.95 | 0.53 | 0.01 |
| CXCR3 | | | |
| HIV- Naïve vs HIV+ Naïve | -3.70 | 0.74 | 0.00 |
| HIV- Unsw MBC vs HIV+ Unsw MBC | -2.06 | 0.69 | 0.13 |
| HIV- MBC vs HIV+ MBC | -1.18 | 0.69 | 0.89 |
| HIV- GCBC vs HIV+ GCBC | -2.57 | 0.70 | 0.01 |
| FCRL4 | | | |
| HIV- Naïve vs HIV+ Naïve | -4.85 | 0.77 | $3.16 \times 10^{-7}$ |
| HIV- Unsw MBC vs HIV+ Unsw MBC | -1.01 | 0.67 | 0.64 |
| HIV- MBC vs HIV+ MBC | -3.99 | 0.68 | $8.50 \times 10^{-6}$ |
| HIV- GCBC vs HIV+ GCBC | -1.57 | 0.69 | 0.25 |
| ITGAX | | | |
| HIV- Naïve vs HIV+ Naïve | -3.80 | 0.81 | 0.00 |
| HIV- Unsw MBC vs HIV+ Unsw MBC | -2.90 | 0.79 | 0.03 |
| HIV- MBC vs HIV+ MBC | -1.68 | 0.77 | 0.59 |
| HIV- GCBC vs HIV+ GCBC | 1.00 | 0.80 | 0.67 |
| CR2 | | | |
| HIV- Naïve vs HIV+ Naïve | -0.36 | 0.24 | 0.61 |
| HIV- Unsw MBC vs HIV+ Unsw MBC | 0.58 | 0.24 | 0.31 |
| HIV- MBC vs HIV+ MBC | 0.35 | 0.24 | 1.00 |
| HIV- GCBC vs HIV+ GCBC | -0.08 | 0.24 | 0.94 |
| FCRL5 | | | |
| HIV- Naïve vs HIV+ Naïve | -1.20 | 0.38 | 0.08 |
| HIV- Unsw MBC vs HIV+ Unsw MBC | -1.62 | 0.38 | 0.00 |
| HIV- MBC vs HIV+ MBC | -1.65 | 0.38 | 0.00 |
| HIV- GCBC vs HIV+ GCBC | 0.08 | 0.38 | 0.97 |

lfcSE, log fold change standard error
padj, Benjamini-Hockberg adjusted *P* value

**Table S2. Antibodies used in imaging analyses.**

| Marker | Source | Species/Isotype | Fluorochrome | Clone or Cat# | Company |
|---|---|---|---|---|---|
| Panel 1 | | | | | |
| CD20 | direct | | eF 615 | L26 | Thermofisher |
| CD4 | direct | | Alexa 700 | FAB8165N | R&D |
| Ki67 | direct | | BV605 | Ki-67 | Biolegend |
| JoJo | direct | | 529/545 | cat#J11372 | Thermofisher |
| IgD | direct[a] | | Alexa 488 | EPR6146 | Abcam |
| CD11c | indirect | mIgG1 | Alexa 647 | CL1831 | Novus |
| CD8 | indirect | mIgG2b | Alexa 546 | 4B11 | Thermofisher |
| T-bet | indirect | rabbit | BV421 | polyclonal/cat#LS-B14387 | LSBio |
| Panel 2 chemokine | | | | | |
| CD20 | direct | | eF 615 | L26 | Thermofisher |
| T-bet | direct | | Alexa 647 | 4B10 | Biolegend |
| JoJo | direct | | 529/545 | cat#J11372 | Thermofisher |
| CD8 | indirect | mIgG2b | Alexa 546 | 4B11 | Thermofisher |
| CXCL13 | indirect | goat | Alexa 488 | polyclonal/cat#PA5-47035 | Thermofisher |
| Panel 3 | | | | | |
| CD20 | direct | | eF 615 | L26 | Thermofisher |
| IgD | direct | | Alexa 488 | EPR6146 | Abcam |
| T-bet | direct | | Alexa 647 | 4B10 | Biolegend |
| JoPro | direct | | 530/546 | cat#J11373 | Thermofisher |
| CD27 | indirect | rabbit | BV421 | EPR8569 | Abcam |
| Bcl2 | indirect | mIgG1 | Alexa 700 | 124 | Thermofisher |
| Bcl6 | indirect | mIgG2b | Alexa 546 | LN22 | Abnova |
| Panel 4 | | | | | |
| CD20 | direct | | eF 615 | L26 | Thermofisher |
| IgD | direct | | Alexa 488 | EPR6146 | Abcam |
| T-bet | direct | | Alexa 647 | 4B10 | Biolegend |
| JoPro | direct | | 530/546 | cat#J11373 | Thermofisher |
| CXCR3 | direct | | Alexa 700 | 1C6/CXCR3 | BD Pharmingen |
| BACH2 | indirect | rabbit | BV421 | D3T3G | Cell Signaling |
| Bcl6 | indirect | mIgG2b | Alexa 546 | LN22 | Abnova |
| Blimp1 | indirect | rat IgG2a | BV605 | 6D3 | Thermofisher |
| Secondary antibodies | | | | | |
| mIgG1 | indirect | goat α mouse | Alexa 647 | cat#A21240 | Thermofisher |
| mIgG1 | indirect | rat α mouse | Alexa 700 | cat#406631 | Biolegend |
| mIgG2b | indirect | goat α mouse | Alexa 546 | cat#A21143 | Thermofisher |
| rabbit | indirect | donkey α rabbit | BV421 | cat#406410 | Biolegend |
| goat | indirect | chicken α goat | Alexa 488 | cat#A21467 | Thermofisher |
| rat IgG2a | indirect | goat α rat | BV605 | cat#405430 | Biolegend |

[a]in house conjugation

**Table S3. HIV-infected participants for additional phenotyping.**

| Participants | Age (years) | Gender (M/F) | $CD4^+$ T-cell % | $CD4^+$ T-cell (count/µl) | $CD8^+$ T-cell (count/µl) | $CD19^+$ B-cell (count/µl) | HIV RNA (copies/ ml) | Years infected |
|---|---|---|---|---|---|---|---|---|
| [a,]HIV14 | 30 | M | 27 | 407 | 649 | 256 | 62024 | <1 |
| HIV15 | 41 | M | 49 | 920 | 732 | 113 | 23263 | <1 |
| HIV16 | 30 | M | 30 | 408 | 789 | 68 | 12167 | 1-5 |

[a]Colors match with those in Figures 2B, 2C, and 3A