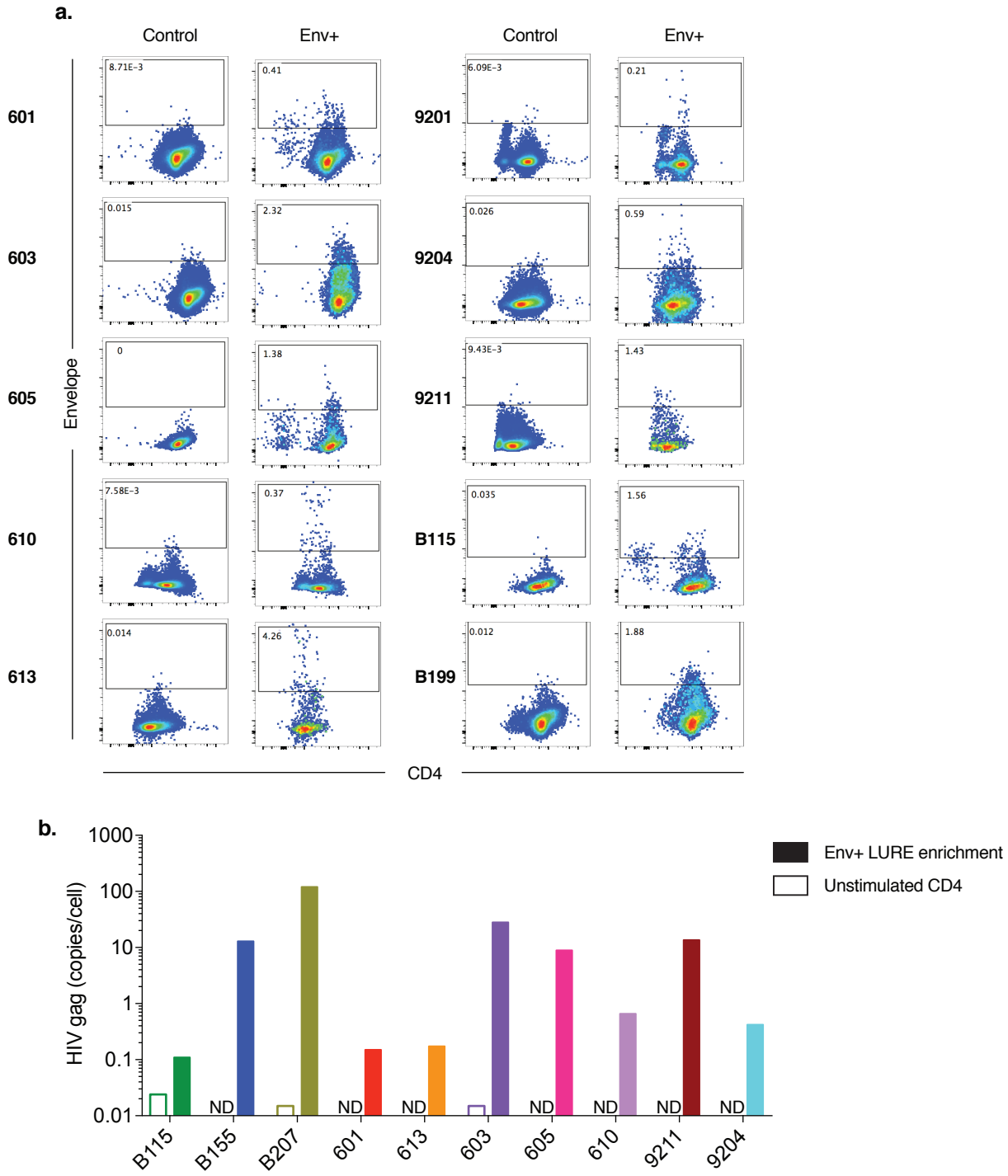
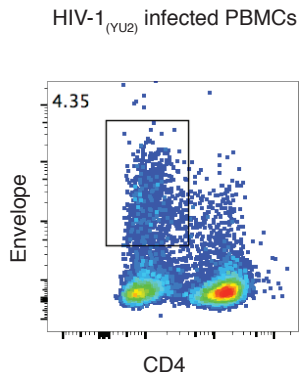
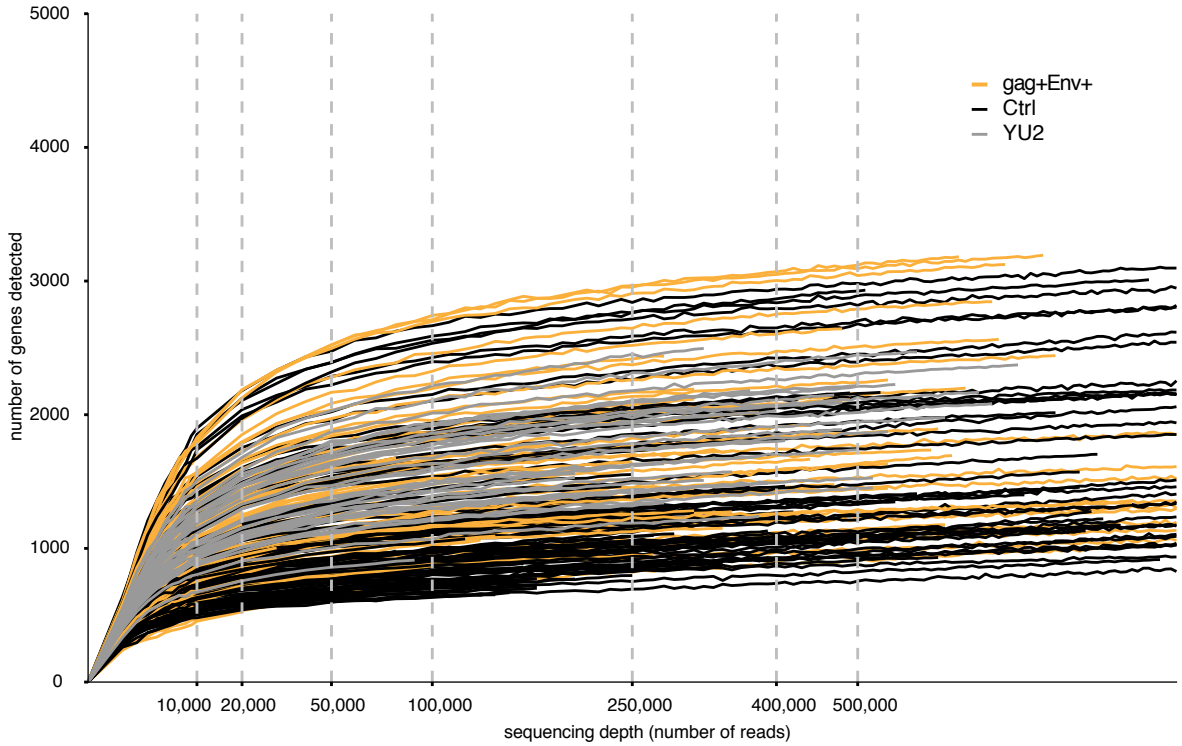


Supplemental Data Figure 1. LURE enrichment and Gag mRNA expression upon PHA stimulation. a) Envelope-expressing cell enrichment. Dot plots show Env vs. CD4 staining on pre-enrichment control, and positively selected cells. Gate shows frequency of Env+ cells in each population. Shown is representative data from more than 30 independent LURE capture experiments. b) HIV-gag mRNA was measured in equivalent numbers (300-3000 cells, depending on the individual) of Env+ and unstimulated CD4+ cells. Graph shows results of qPCR (12.8-copy limit of detection) for HIV-gag mRNA, normalized to the number of sorted cells.

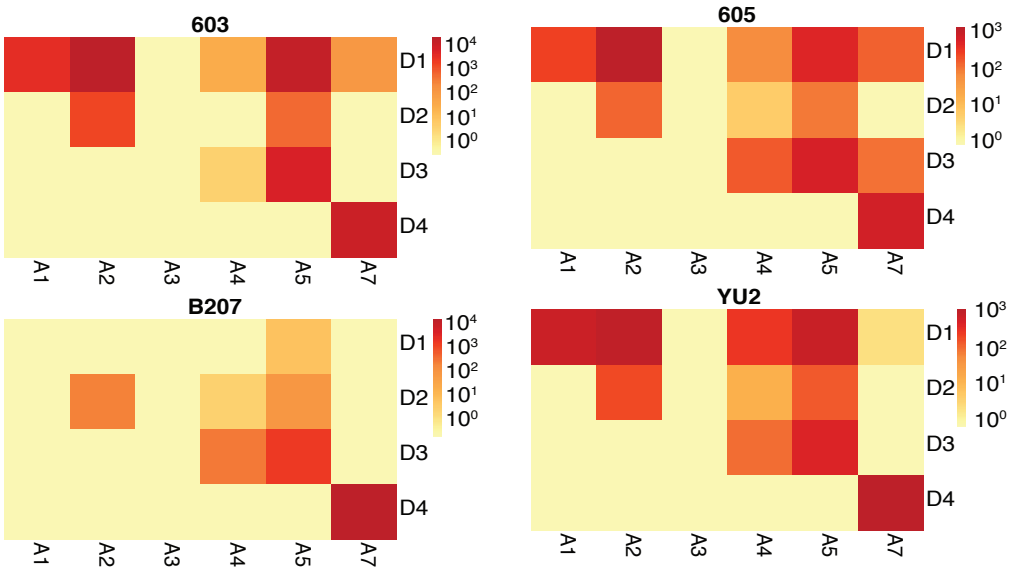


Supplemental Data Figure 2. Gating strategy for HIV-1(YU2) infected cells. Cells infected in vitro with HIV-1(YU2) for 2 days were FACSorted by gating on Env+CD4lo cells.

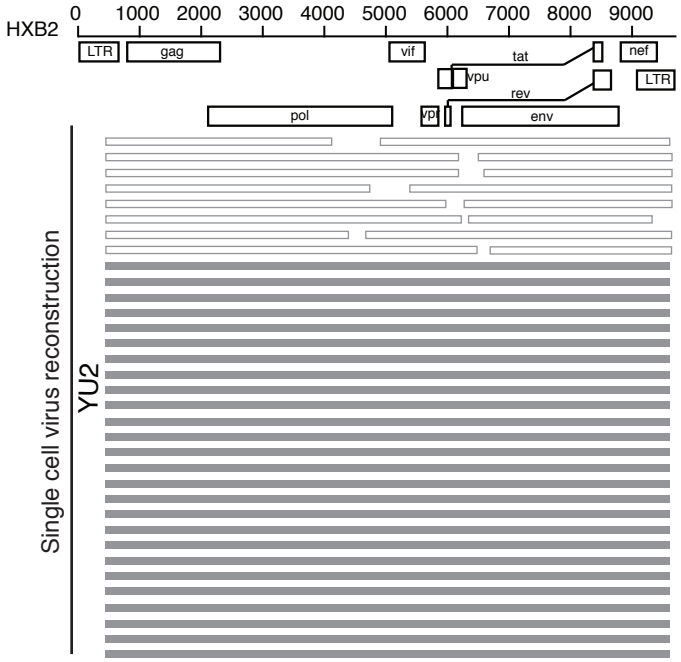


Supplemental Data Figure 3. Number of genes detected per cell. Results of single cell RNASeq showing saturation of genes detected.

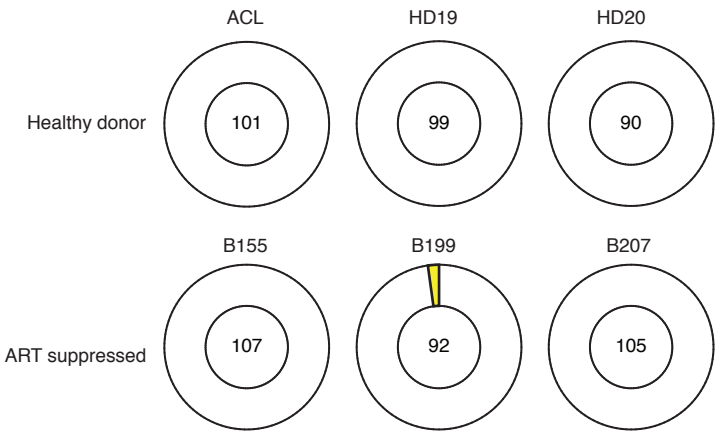
a.



b.



Supplemental Data Figure 4. HIV splice junctions and HIV-1(YU2) single cell virus reconstruction. a) Junctions between HIV splice donors and acceptors observed in RNASeq data. Acceptors are shown as the columns and donors as the rows with the coloring indicating the frequency of reads identified containing indicated splice junction. b) Map of individual viruses recovered by scRNASeq in HIV-1(YU2) infected cells. Each horizontal bar represents a virus from an individual cell. Solid bars indicate that the entire virus was reconstructed from the scRNASeq reads. Outlined bars indicate incomplete genome reconstruction.



Supplemental Data Figure 5. Control TCR Sequences. TCR sequences amplified by PCR in single sorted CD4+ T cells. The number in the center of the pie denotes the number of cells sequenced; yellow slice is a unique clone consisting of two members. The single clone in B199 was identified by shared TCR alpha and beta sequence.

603

TCRBV7-2*01
TCRBD2*01
TCRBJ2-3*01

TCTCCACTCTGAAGATCCAGCGCACACAGCAGGAGGACTCGGCCGTGTATCTCTGT
GCCAGCAGCTATAGAGGACTAGCGGGCACAGATACGCAGTATTTTGGCCCAGGCAC
CCGGCTGACAGTGCTCG

605

TCRBV11-2*01
TCRBD2*01
TCRBJ2-5*01

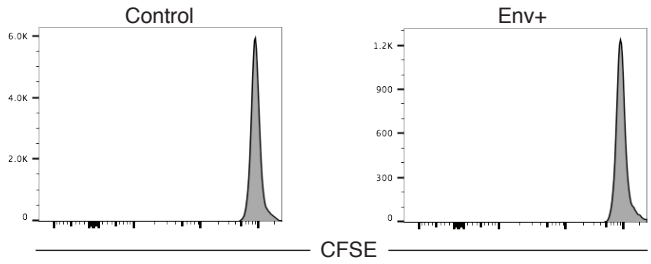
GCAGAGAGGCTCAAAGGAGTAGACTCCACTCTCAAGATCCAACCTGCAAAGCTTGA
GGACTCGGCCGTGTATCTCTGTGCCAGCAGCTTAGGAGCGGCTCAAGAGACCCAGT
ACTTCGGGCCAGGCACGCGGCTCCTGGTGCTCG

B207

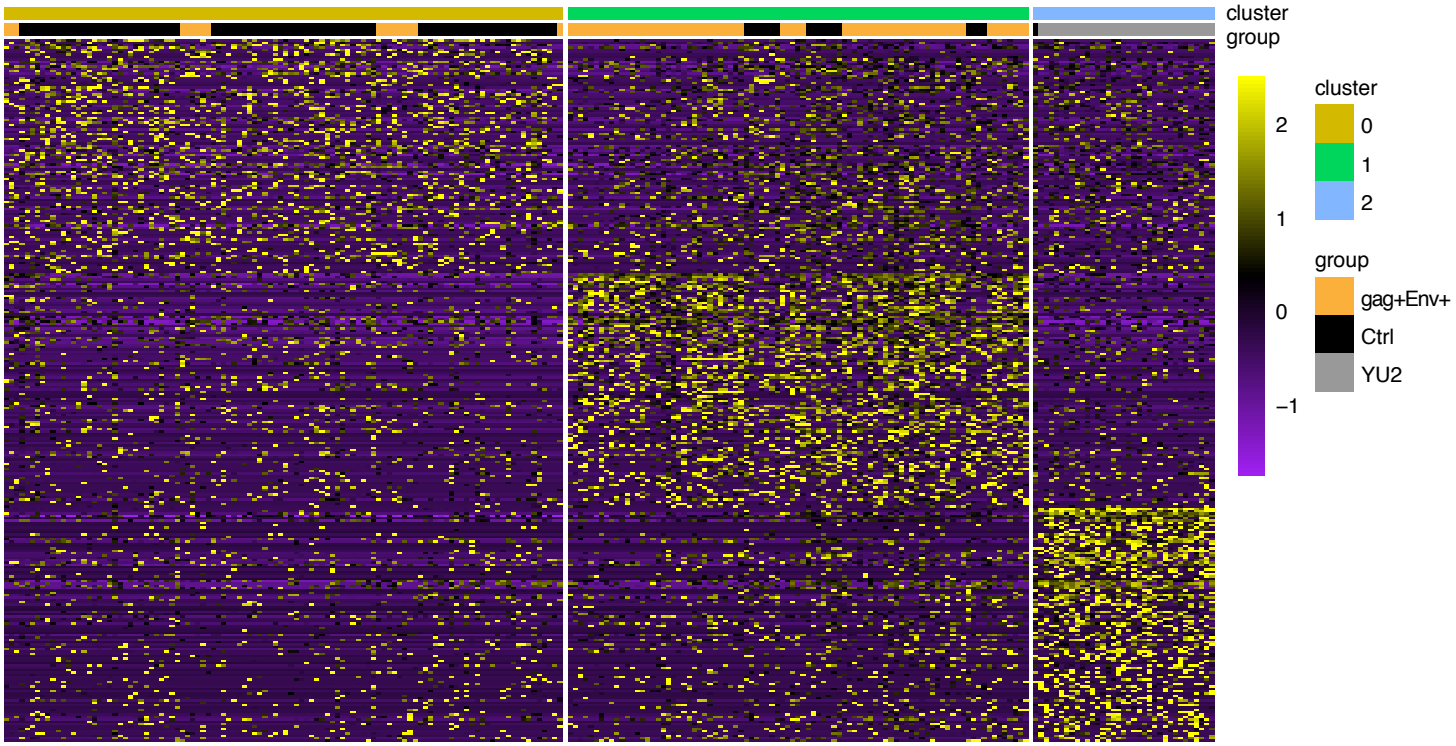
TCRBV7-8*01
TCRBD1*01
TCRBJ1-4*01

GATCGCTTCTTTGCAGAAAGGCCTGAGGGATCCGTCTCCACTCTGAAGATCCAGCGCA
CACAGCAGGAGGACTCCGCCGTGTATCTCTGTGCCAGCAGCTACCAGGGTACTGATG
AAAAACTGTTTTTTGGCAGTGGAACCCAGCTCTCTGTCTTGGAGGAC

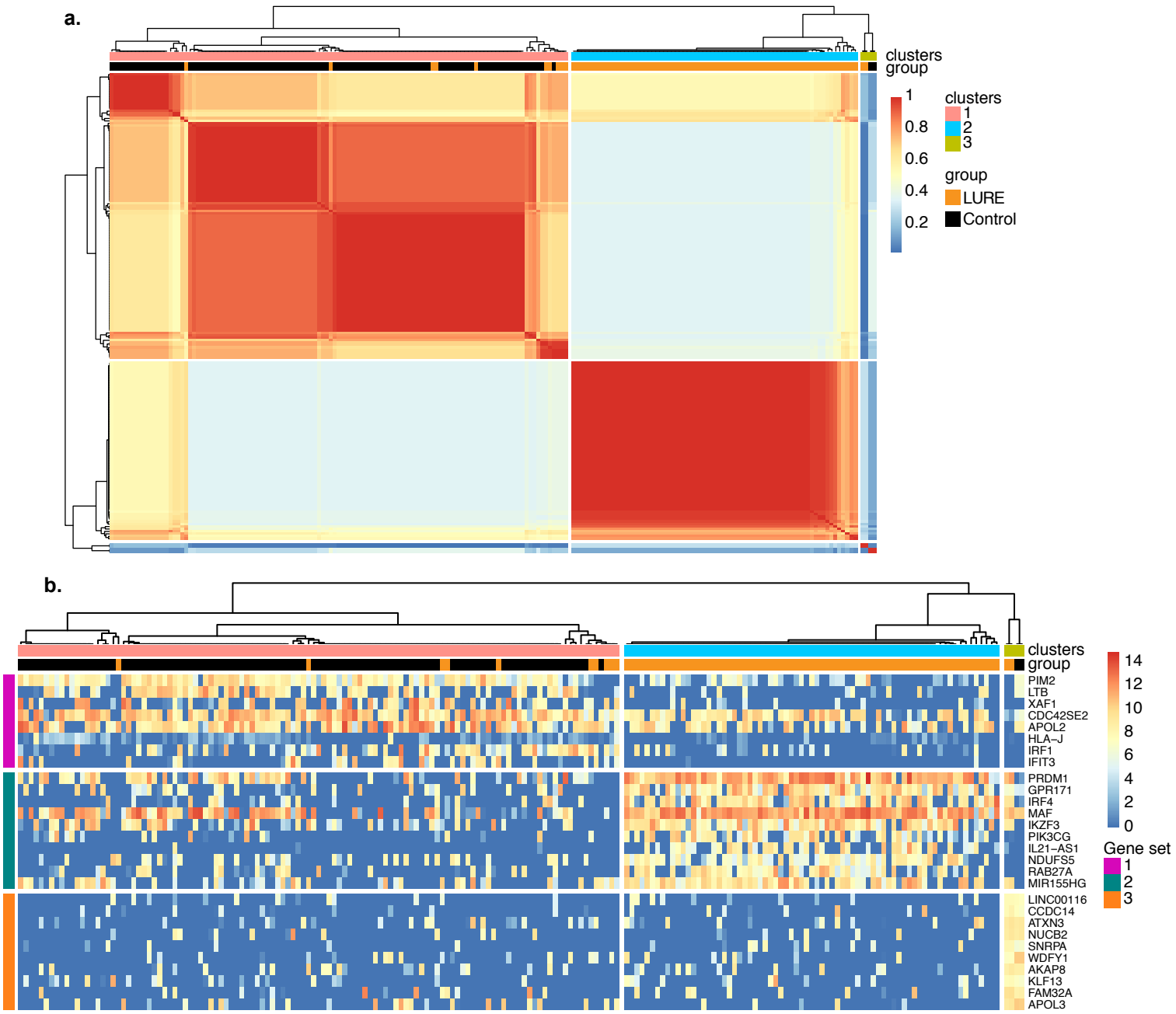
Supplemental Data Figure 6. TCR sequences from LURE cells. TCR sequences from LURE cells were recovered from RNASeq libraries or generated from specific TCR PCR sequences as needed. Shown here are the TCR Beta gene sequences and V, D, J assignments after IgBlast analysis.



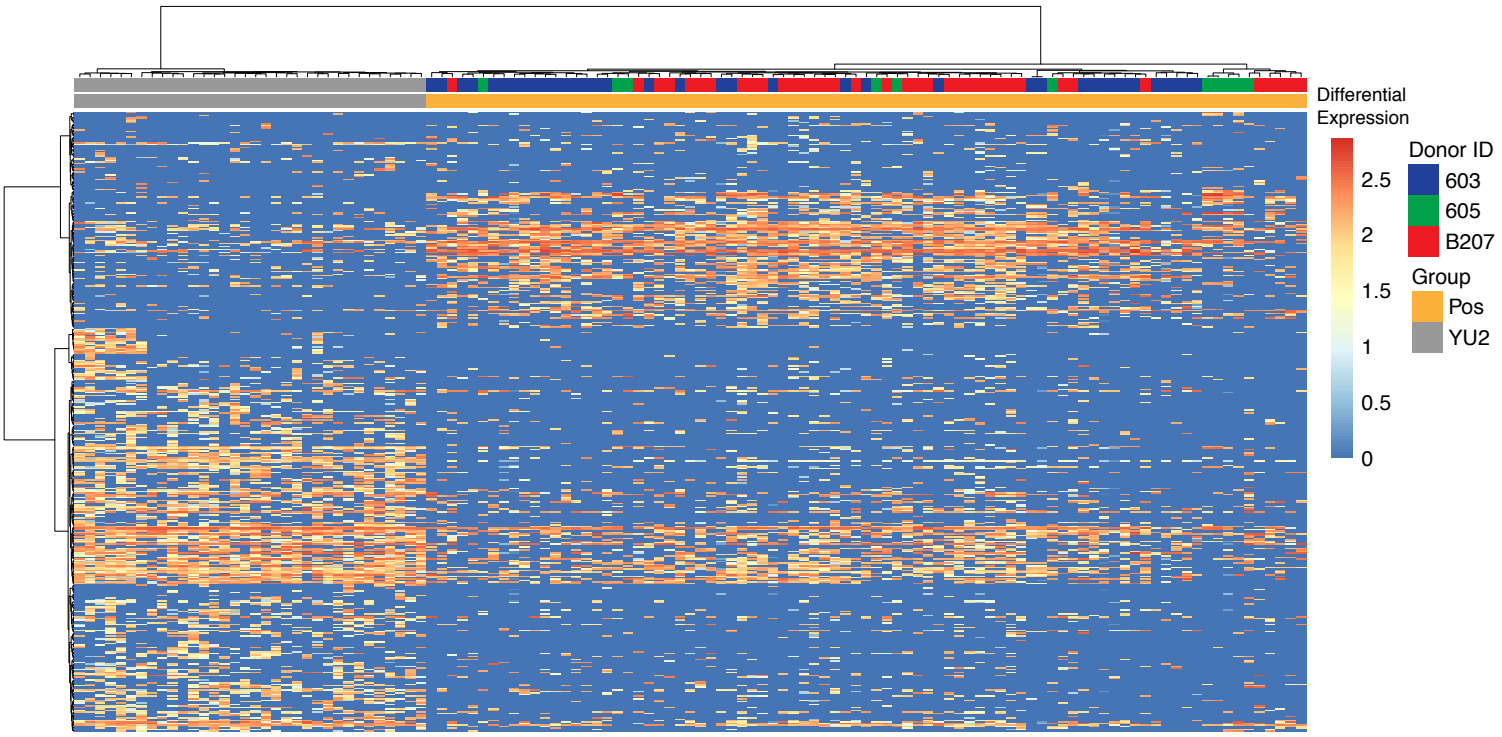
Supplemental Data Figure 7. No cell division in vitro within 36h in LURE cultures. LURE assay performed on CFSE labeled CD4 T cells. After 36 hours, cells were analyzed for CFSE dilution by FACS. Control and Envelope positive fraction show no cell division after 36h. Shown is one representative experiment of 6 independent samples from 3 separate patients.



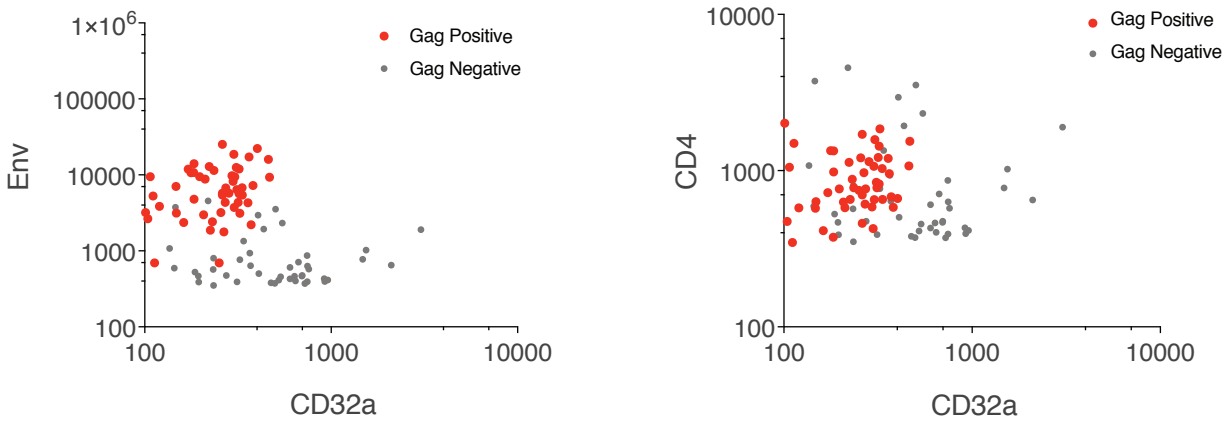
Supplemental Data Figure 8. Heatmap of genes identified by principal components analysis. Genes contributing to PCA are shown for each cell. Three groups, cluster 0 in mustard, 1 in green, and 2 in blue, are shown for expression of the top 100 genes involved in the PCA clustering. Cells contributing to each group are also shown (gag+Env+ in orange, control in black and YU2 in gray).



Supplemental Data Figure 9. Single-cell clustering segregates control from LURE Env+gag+ cells. a) Heatmap showing the similarity between cells based on clustering results. Single-cell consensus clustering (SC3) was used to cluster cells in an unsupervised manor. Color spectrum shows assignment of cells to different clusters, with blue indicating assignment to a different cluster and red indicating cells assigned to the same cluster. b) Heatmap showing the markers identified by SC3 to distinguish each cluster (blue: low expression, red: high expression of marker gene).



Supplemental Data Figure 10. Segregation of YU2 and LURE Env+gag+ cells. Heat-map shows unsupervised clustering of differentially expressed genes between gag+Env+ LURE purified cells (orange bars) and YU2 infected cells (gray bars).



Supplemental Data Figure 11. CD32a is not expressed by Env+gag+ cells. Single cell index sorting was performed on Env+ enriched cells from patient B207 followed by RNA isolation, cDNA synthesis and gag qPCR. Gag+ cells were then examined for CD32a, Env and CD4 expression by index sorting. Shown is the mean fluorescence in gag+ and gag- cells from the same experiment. Shown is one representative experiments of 3 independent experiments.

Patient ID	Age	Gender	Race	Year Dx	Years on ART (uninterrupted)	ART regimen	IUPM	Env+ bulk gag RNA enrichment, LURE	Single Cell LURE
B115	46	M	Black	1993	24	ATV/r/ABC/3TC	0.57	YES	NO
B155	61	M	Black	1993	17	RPV/FTC/TDF	1.07	YES	YES
B199	51	M	White	2009	6	RAL/FTC/TDF	0.61	ND	YES
B207	48	M	White/Hisp	2006	11	EFV/FTC/TDF	16	YES	YES
601	58	M	White	1994	20	LPV/r/ABC/3TC	0.49	YES	NO
603	45	M	White/Hisp	2003	12	EFV/FTC/TDF	3.17	YES	YES
605	38	M	White/Hisp	2001	15	RPV/FTC/TDF	0.71	YES	YES
610	31	M	White	2011	5	RPV/FTC/TDF	1.95	YES	YES
613	51	M	Multiple/Hisp	1997	19	ATV/r/FTC/TDF	1.23	YES	YES
9201	44	M	White/Hisp	2013	4	EGV/cobi/FTC/TDF	0.78	ND	YES
9204	23	M	White/Hisp	2012	5	EGV/cobi/FTC/TAF	0.79	YES	YES
9211	40	M	Black	2011	5	EGV/cobi/FTC/TDF	2.43	YES	YES

Supplemental Table 1. Patient demographics and LURE experiments. ART abbreviations, ATV: atazanavir, R: ritonavir, ABC: abacavir, 3TC: lamivudine, RPV: rilpivirine, FTC: emtricitabine, TDF: tenofovir disoproxil, RAL: raltegravir, EFV: efavirenz, LPV: lopinavir, EGV: elvitegravir, TAF: tenofovir alafenamide, cobi: cobicistat. Env+ bulk gag RNA enrichment, LURE: gag RNA enrichment performed on immunomagnetically isolated Env+ cellular fraction. YES: significant enrichment in Env+ fraction compared to controls. ND: not done. Single Cell LURE: single cell sort of Env+ enriched LURE cells. YES: gag+ cells identified by single cell qPCR. NO: no gag+ cells identified by single cell qPCR.

Supplemental Table 2. Genes that segregate cells into clusters by PCA. Genes identified using Seurat cluster cells into three distinct groups. Clusters are indicated in the table by 0, 1 or 2 and correspond to the data in Figure 4a. This table includes genes obtained using Seurat analysis with gene expression data from individuals 603, 605, B207 (109 control cells and 85 LURE cells) and HIV-1_{YU2} infected health donor cells (33 cells total). P-value was determined using the non-parameteric Wilcoxon rank sum test embedded in the Seurat software.

Supplemental Table 3. Differentially expressed gene list. Genes differently expressed by Env+ compared to control cells with $p<0.01$. Table includes p-value, false determination rate, and fold change. Positive fold-change values indicate higher expression in Env+ cells. Negative fold-change values indicate higher expression in control cells. If the mean expression in the Env+ group is equal to zero, the fold-change is indicated as "-Inf." Conversely, if the mean expression in control group is equal to zero, the fold-change is indicated as "+Inf." Shown is all data obtained from individuals 603, 605, B207 (109 control cells and 85 LURE cells). Significant differential expression was determined using the likelihood ratio test embedded in the MAST software.

Supplemental Table 4. Enriched biological processes and molecular functions using Gene Ontology database. Enrichment analysis was performed on 240 genes which overlapped between differential expression and PCA analyses using the Gene Ontology database. 282 significantly enriched biological processes and 7 significantly enriched molecular functions were identified. Shown are genes included in each gene ontology category, segregated by their differential expression in Env+ or control cells.

Supplemental Table 5. Genes included in gene ontology categories. The gene ontology database was cross referenced with the 240 genes which overlapped between differential expression and PCA analyses. Genes falling into any biological process or molecular function category are shown, segregated by their differential expression in Env+ or control cells.