Molecular Cell, Volume *82*

Supplemental information

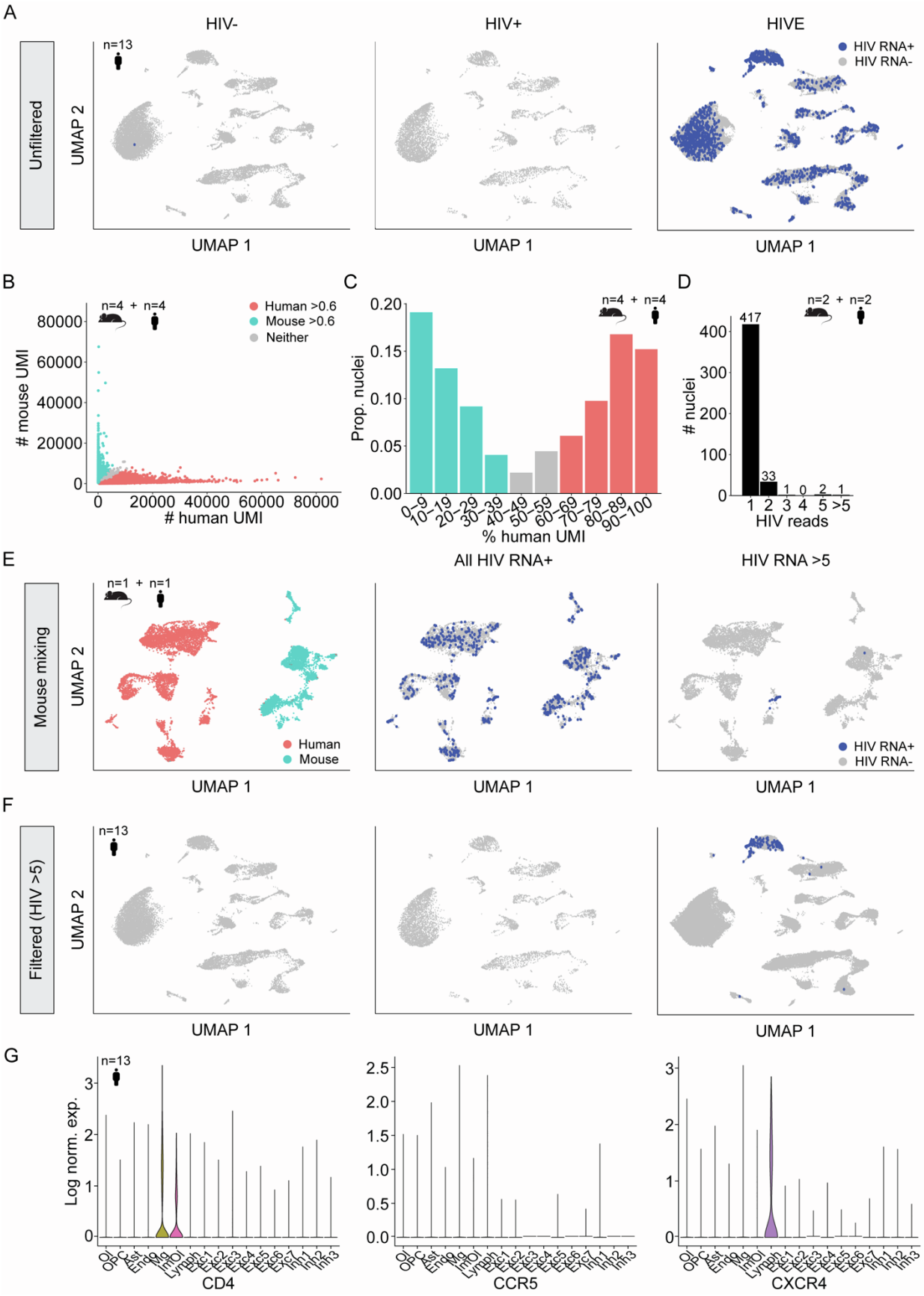
HIV integration in the human brain is linked to microglial activation and 3D genome remodeling

Amara L. Plaza-Jennings, Aditi Valada, Callan O'Shea, Marina Iskhakova, Benxia Hu, Behnam Javidfar, Gabriella Ben Hutta, Tova Y. Lambert, Jacinta Murray, Bibi Kassim, Sandhya Chandrasekaran, Benjamin K. Chen, Susan Morgello, Hyejung Won, and Schahram Akbarian

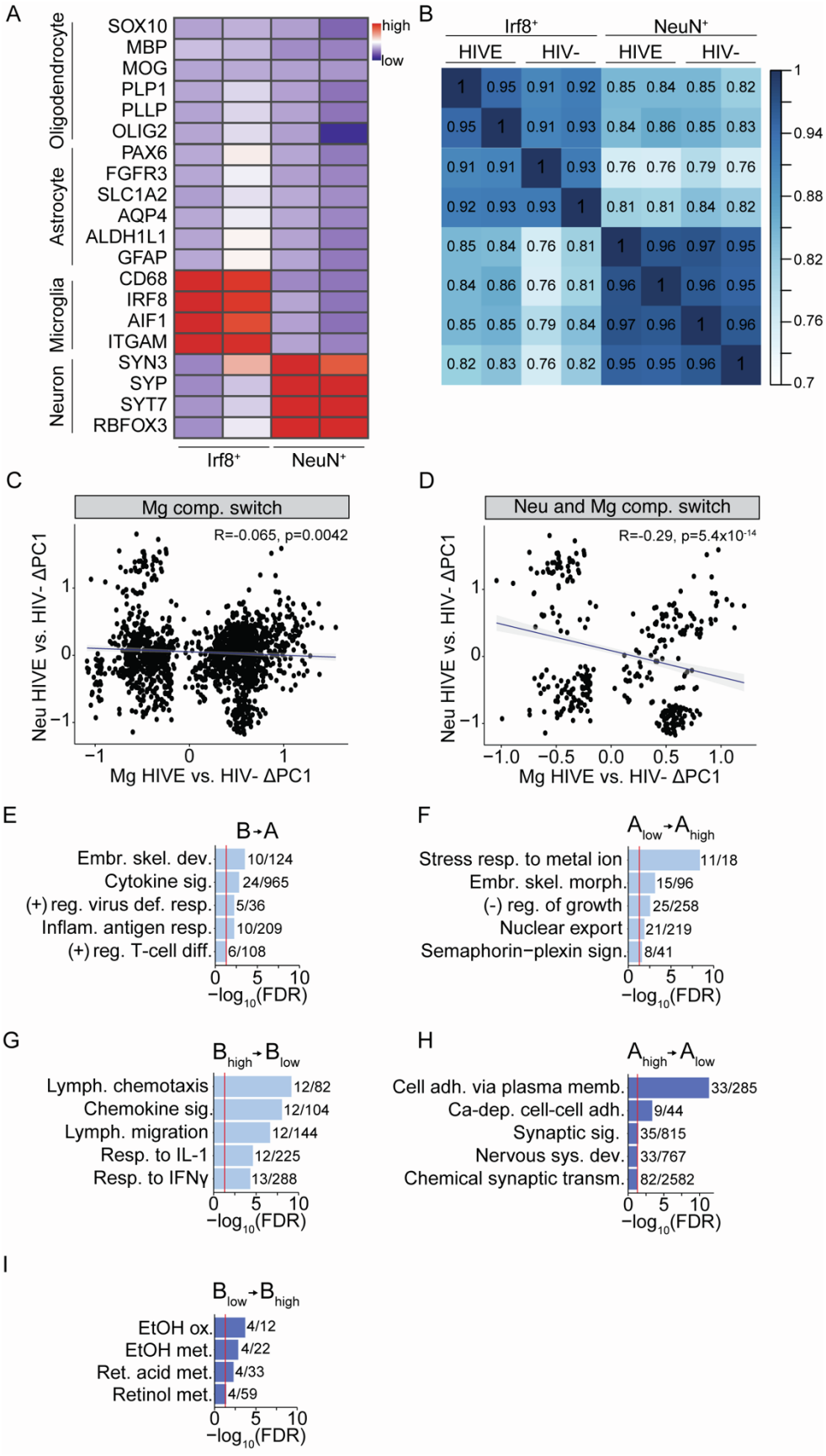
Supplemental information

HIV integration in the human brain is linked to microglial activation and 3D genome remodeling

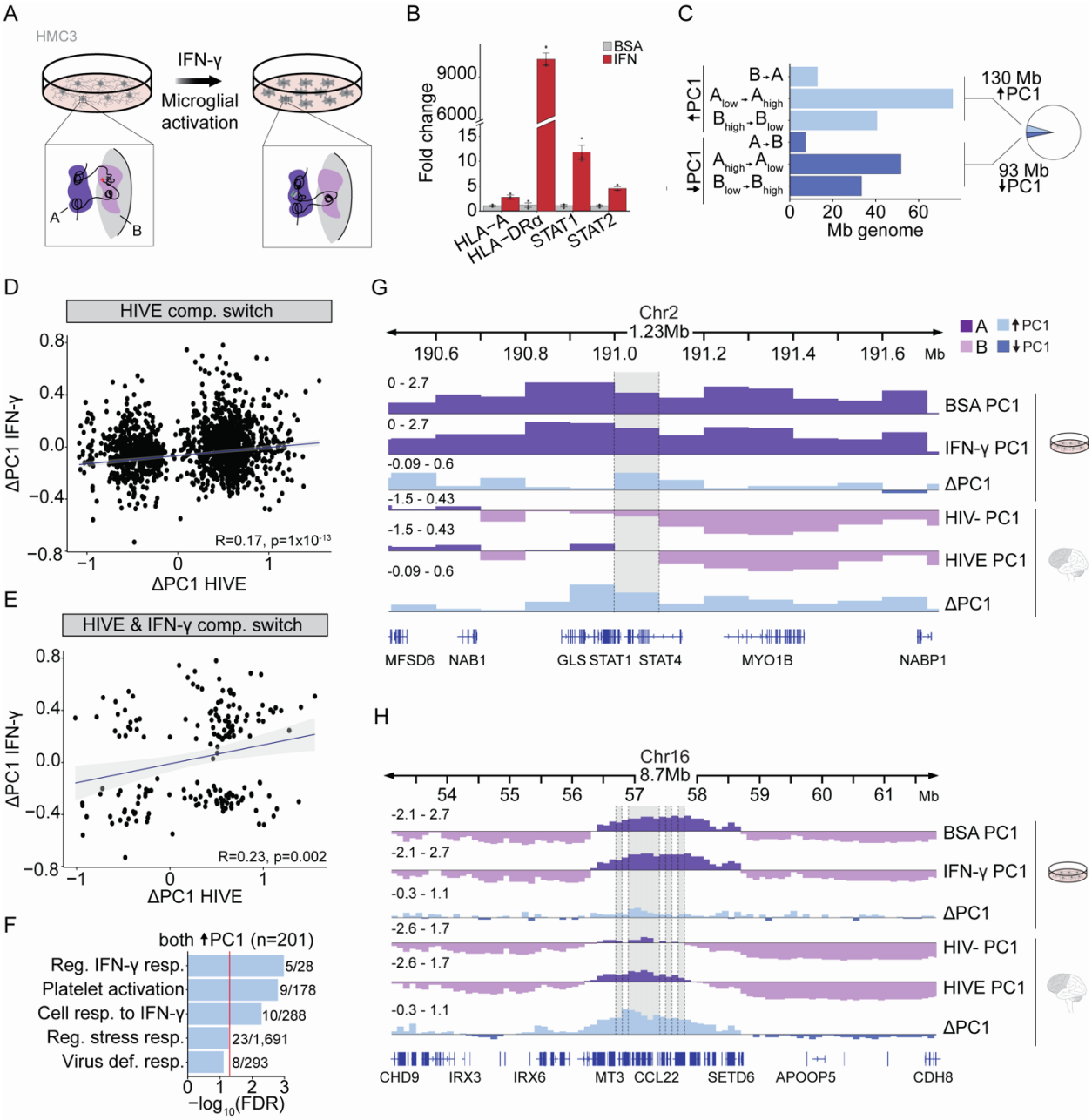
Amara L. Plaza-Jennings, Aditi Valada, Callan O'Shea, Marina Iskhakova, Benxia Hu, Behnam Javidfar, Gabriella Ben Hutta, Tova Y. Lambert, Jacinta Murray, Bibi Kassim, Sandhya Chandrasekaran, Benjamin K. Chen, Susan Morgello, Hyejung Won, and Schahram Akbarian



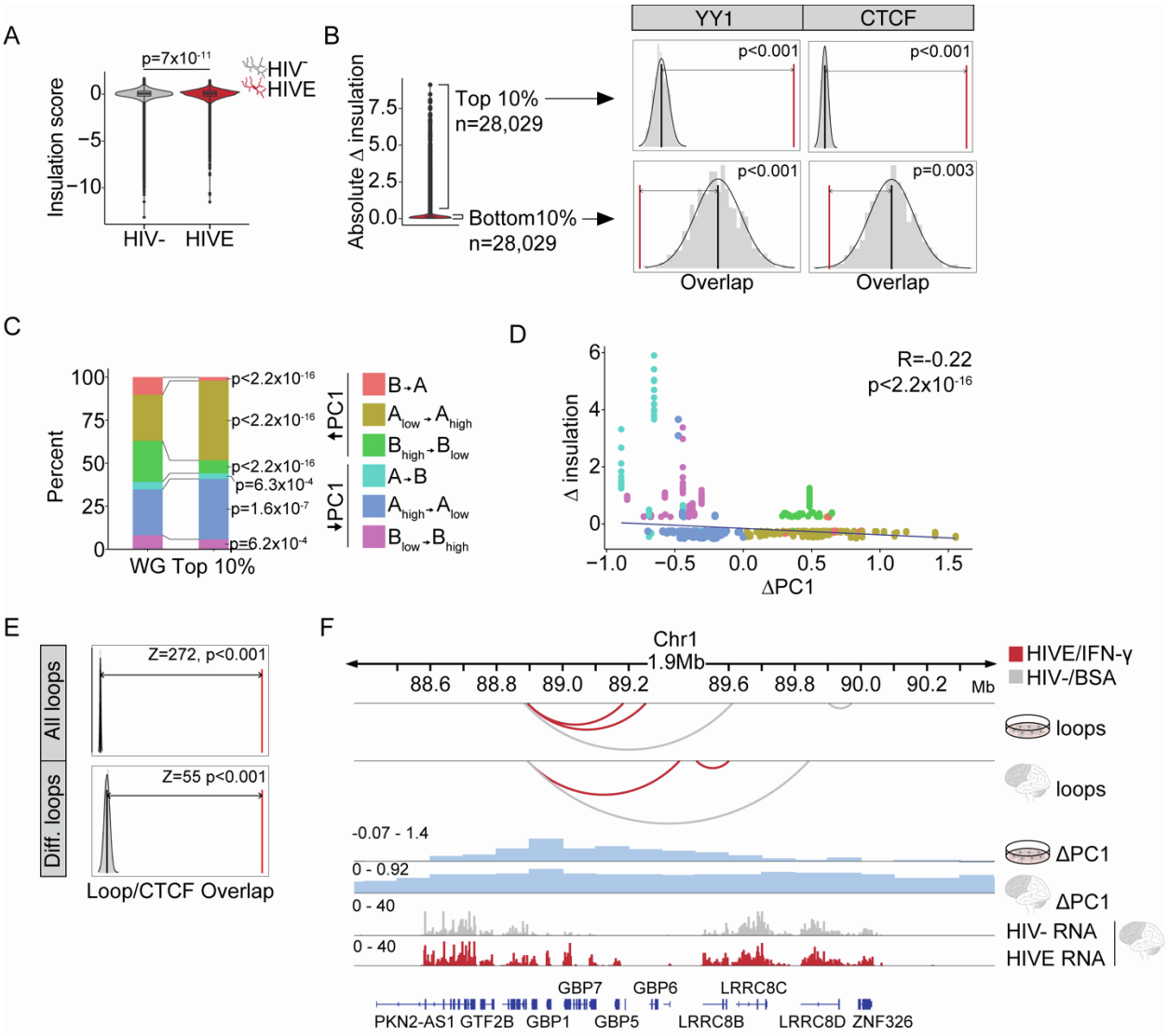
Supplementary Figure 1



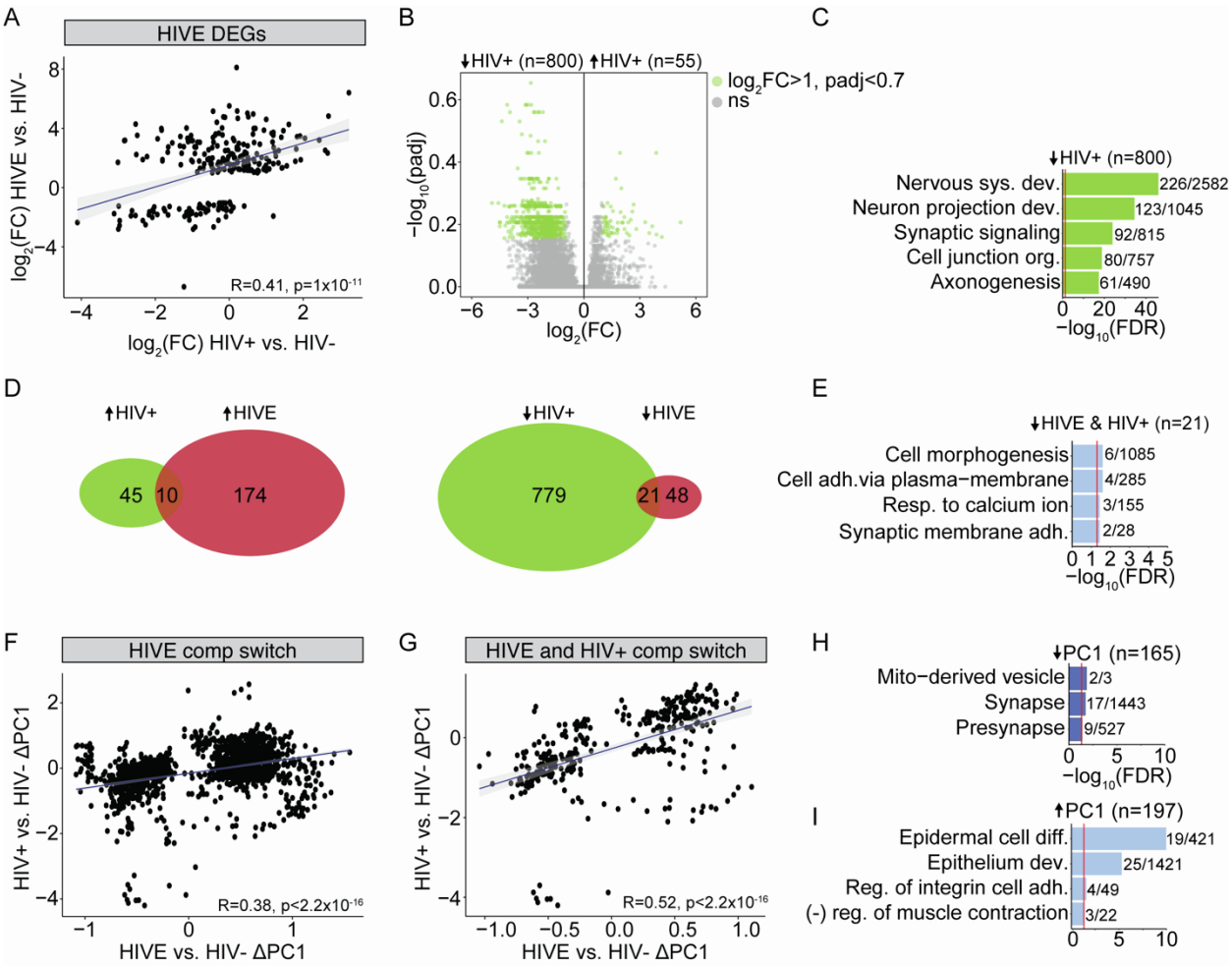
Supplementary Figure 2



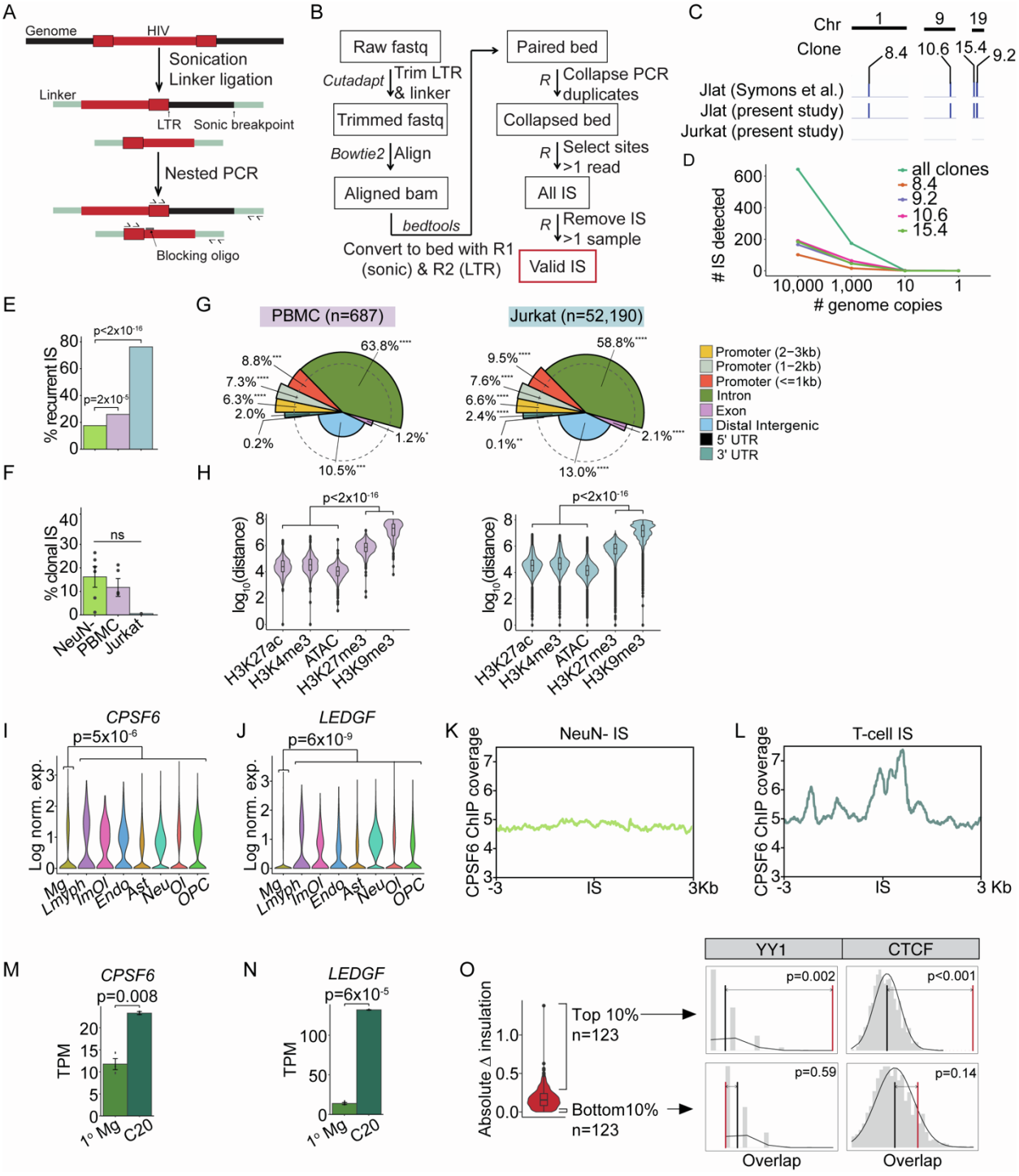
Supplementary Figure 3



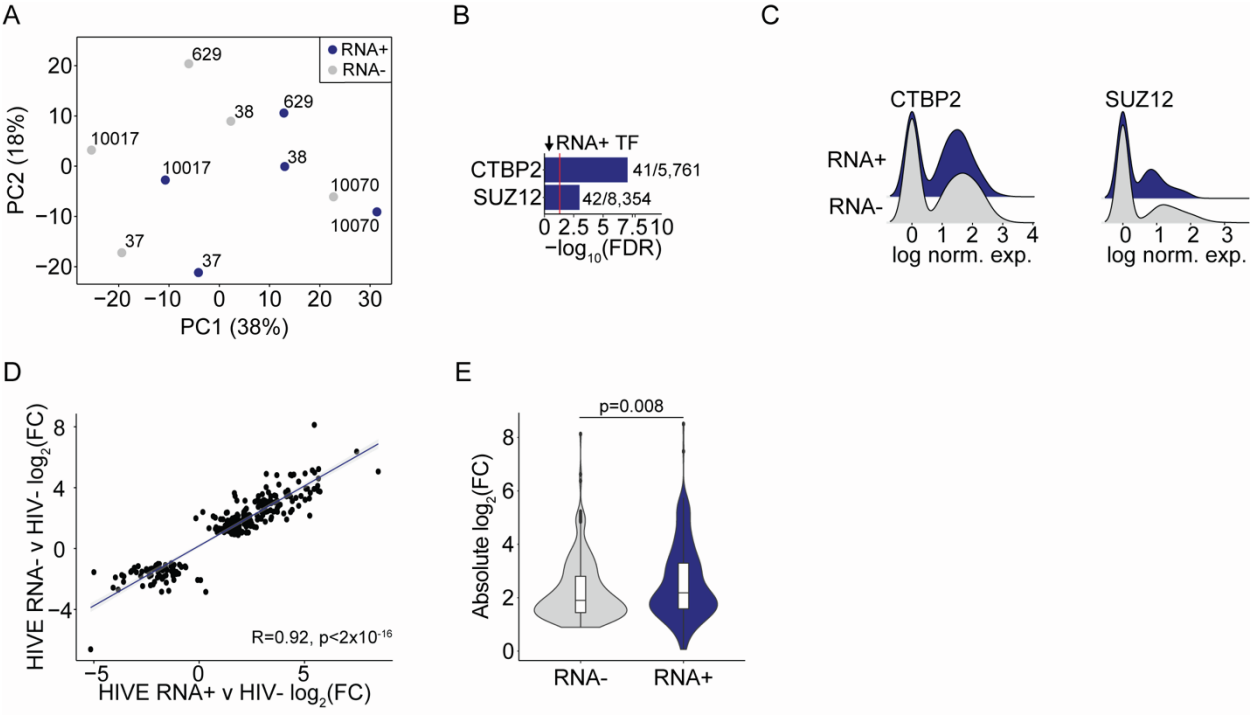
Supplementary Figure 4



Supplementary Figure 5



Supplementary Figure 6



Supplementary Figure 7

Donor characteristics											Assays performed §							
Donor ID	Sample type	Age	Sex	HIV diagnosis duration	Last CD4 count prior to death	Last plasma HIV load prior to death	Unusual clinical and/or pathological features	cART at death†	Clinical notes	HIV qPCR (copies /250,000 cells)‡	ISA	# NeuN+ IS lib ‖	# NeuN- IS lib ‖	NeuN+ RNA-seq	Irf8- RNA-seq	snRNA-seq	NeuN+ Hi-C	Irf8+ Hi-C
10002	HIV+	58	M	16	459	Und.	Elite controller	No	Elite controller	Und.	✓	1	1					
10011	HIV+	42	M	18	16	162642		No		Und.	✓	*	*					
10017	HIVE	43	M	9	7	389120		No		762888*	✓	2	2			✓	✓	✓
10052	HIV+	39	M	11	79	174175		No		Und.	✓		1					
10063	HIV+	51	M	13	136	65		Yes		Und.	✓	1	1					
10070	HIVE	37	M	4	1	>750000		No^		56336	✓	1	2			✓	✓	✓
10075	HIV+	67	M	24	69	81755		No		<100	✓	*	*					
10089	HIV+	53	M	16	78	Und.		Yes		Und.	✓	*	*					
10105	HIV+	48	M	12	63	5722		Yes		Und.	✓	*	*					
10116	HIV+	45	M	16	28	628		Yes		Und.	✓	*	1					
10149	HIV+	57	F	24	73	35723		No		<100	✓	2	1					
10162	HIV+	42	M	9	298	Und.	CD8 encephalitis	No	CD8 encephalitis	Und.*	✓	1	1					
10163	HIV+	68	M	19	527	Und.		No		n.d.	✓	1	*					
10186	HIV+	70	F	29	543	Und.		Yes		Und.	✓	*	1					
10201	HIV+	63	M	27	133	24		Yes		Und.	✓	*	*					
10267	HIV+	59	M	20	208	658		Yes		<100	✓	1	1			✓		
10297	HIV+	63	M	34	853	580		Yes		<100	✓	1	1			✓		
10304	HIV+	72	F	29	305	21		No		<10	✓	1	1					
10305	HIV+	62	M	24	172	Und.		Yes		Und.	✓	*	*					
10321	HIV+	74	M	29	429	202		Yes	CD8 encephalitis, compartmentalized	n.d.	✓	2	2					
20033	HIV+	49	M	16	103	Und.	CD8 encephalitis, CNS viral compartment documented	Yes		<100	✓	1	1					
20069	HIV+	60	M	7	38	Und.		Yes		Und.				✓	✓			
50014	HIV+	61	F	18	1167	Und.		Yes		Und.	✓	1	1	✓	✓			
50026	HIV+	53	F	24	493	43		Yes		<100	✓	1	1				✓	✓
50030	HIV+	63	F	20	516	Und.		No		<100	✓	1	1					
60003	HIV+	49	F	19	487	82		Yes		<100	✓	1	1			✓		
mhbb 037	HIVE	50	M	11	31	unknown	cART-naïve	No		n.d.	✓		2			✓		
mhbb 038	HIVE	43	M	4	unknown	unknown	cART-naïve	No		n.d.	✓	1	2			✓		
mhbb 059	HIVE	30	M	3	0	unknown	cART-naïve	No		n.d.						✓		
mhbb 078	HIVE	40	F	5	0	unknown	cART-naïve	No		n.d.	✓	*	2			✓		
mhbb 079	HIVE	45	M	6	40	unknown	cART-naïve	No		n.d.	✓	*	1			✓		
mhbb 127	HIV+	52	F	unknown	80	unknown		No^		n.d.	✓	*	1					
mhbb 153	HIV-	65	M	n.a.	n.a.	n.a.		n.a.		na	✓	*	1					
mhbb 154	HIV-	46	F	n.a.	n.a.	n.a.		n.a.		n.a.	✓		1			✓		
mhbb 155	HIV-	67	M	n.a.	n.a.	n.a.		n.a.		n.a.	✓	*	1			✓		
mhbb 629	HIVE	57	F	unknown	25	730085		Yes		<100*	✓	2	3			✓		
mhbb 667	HIV+	24	F	24	54	unknown		No^		n.d.	✓		1					
mhbb 694	HIV-	45	M	n.a.	n.a.	n.a.		n.a.		n.a.	✓	2	2			✓	✓	✓
mhbb 739	HIV+	81	F	21	1115	Und.		Yes		158	✓	2	2					
mhbb 742	HIV-	52	M	n.a.	n.a.	n.a.		n.a.		n.a.	✓	3	1			✓	✓	✓

Supplementary Table 1

Supplementary Figure Legends

Figure S1: **Single nuclei HIV expression thresholding and HIV receptor expression. Related to Figure 1. (A)** UMAP plot for the entire snRNA-seq dataset of 13 samples showing HIV RNA expression by sample type. **(B)** Graph showing the number of mouse UMIs vs. human UMIs for each nucleus of 4 mixed human and mouse samples. Nuclei with >60% of UMIs mapping to the human genome, >60% of UMIs mapping to the mouse genome, or neither are indicated. **(C)** Bar graph showing the proportion of all nuclei in the 4 mixed human and mouse samples that have the indicated percent of UMIs mapping to the human genome. **(D)** Bar graph showing the total count of nuclei with the indicated number of HIV-mapping reads for the 2 mixed human and mouse samples that had any HIV-mapping reads. **(E)** UMAP plot for a representative example of snRNA-seq of mixed mouse and human HIVE tissue showing all HIV RNA+ nuclei and those with >5 HIV reads. **(F)** UMAP plot for the entire snRNA-seq dataset of 13 samples showing HIV RNA expression only for those nuclei with >5 HIV reads. **(G)** Violin plots showing expression of CD4, CCR5, and CXCR4 across each cell cluster for the entire snRNA-seq dataset of 13 FC/WM samples (including HIV-, HIV+, and HIVE).

Figure S2: Hi-C library correlation and differential compartment analysis. Related to Figure 2. **(A)** RNA-seq from ensembles of sorted NeuN+ and Irf8+ nuclei showing specific expression of neuronal and microglial marker genes in NeuN+ and Irf8+ nuclei fractions respectively. Heatmap displays scaled transcript per million (TPM) values. **(B)** Sample-by-sample correlations across Hi-C libraries computed using HiCRep. **(C-D)** Pearson correlation between the ΔPC1 value for HIVE microglia and HIVE neurons for all compartments that had a significant ΔPC1 in HIVE microglia (C) and compartments with a significant ΔPC1 in both HIVE microglia and HIVE neurons (D). **E-I**) Gene ontology enrichment for genes in linear genome 10kb bins that had a significant increase (E-G) or

decrease (H-I) in PC1 eigenvector value in HIVE microglia compared to HIV-. Specific compartment switches as indicated. Red lines mark significance threshold at FDR=0.05. Numbers indicate the number of genes in the dataset/the number of genes in the corresponding GO category. GO categories: embr. skel. dev., embryonic skeletal development; cytokine sig., cytokine signaling; (+) reg. virus def. resp., positive regulation of virus defense response; inflam. antigen resp.; inflammatory antigen response; cell adh. via plasma memb., cell adhesion via plasma membrane; synaptic sig., synaptic signaling; nervous sys. dev., nervous system development; chemical synaptic transm., chemical synaptic transmission. The group of A→B genes (N=145) had no significant GO enrichment and is thus not pictured.

Figure S3: 3D genomic alterations in IFN-γ stimulated HMC3 cells. Related to Figure 2. **(A)** HMC3 microglial cells were exposed to IFN-g or BSA for 24hr and then *in situ* Hi-C was performed. **(B)** qPCR of IFN response genes in IFN-g and BSA treated HMC3 cells. Fold change of ΔΔct values are shown, points represent biological replicates (n=3 each). **(C)** Bar graph summarizing genome-wide compartment remodeling in IFN-γ stimulated HMC3 cells microglia compared to BSA-treated cells determined using dcHiC. X-axis shows the number of Mb of the genome undergoing the compartment switch, Y-axis shows the different compartment switches from HIV- to HIVE with A-compartmentalization events in light blue (DPC1 HIVE-HIV- >0, FDR<0.05) and B-compartmentalization events in dark blue (DPC1 HIVE-HIV- <0, FDR<0.05). Inset pie chart shows the total amount of significant A- and B- compartmentalization as a proportion of the genome. **(D-E)** Pearson correlation between the ΔPC1value for HIVE microglia and IFN-γ stimulated HMC3 cells for all compartments that had a significant DPC1 in HIVE microglia (D) and compartments with a significant ΔPC1in both HIVE microglia and IFN-γ stimulated HMC3 cells (E). **(F)** Biological process

GO enrichment for genes in regions that had a significant ΔPC1in both HIVE microglia and IFN-γ stimulated HMC3 cells. Red line marks significance for FDR=0.05, numbers next to each bar show the n genes differentially expression/the total n genes in that GO category. Reg. IFN-γ resp., regulation of the IFN--γ response; cell resp. to IFN-γ, cellular response to IFN-γ; reg. stress resp., regulation of the stress response; virus def. resp., cellular defense response to virus. **(G-H)** Browser shots showing regions that have significantly increased PC1 values in both IFN-γ treated HMC3 cells and HIVE microglia (gray highlighted regions). PC1 values for each condition and ΔPC1values between conditions are plotted.

Figure S4: Hi-C insulation score and loop analysis. Related to Figure 3. (A) Violin plot showing insulation score across the whole genome (WG) in HIV- and HIVE microglia. Wilcoxon rank sum test. **(B)** Absolute value of the change in insulation score between HIVE and HIV- microglia for each 10kb bin of the genome. An overlap permutation (n=1000) test was performed between bins in either top or bottom 10% change in insulation score of the whole genome and ChIP-seq peaks from ENCODE for YY1 from GM12878 cells and CTCF from CD14+ monocytes. Red bar shows the observed number of overlaps between YY1/CTCF ChIP peaks and insulation score regions, black line shows the overlap expected by chance given the whole genome as a background. **(C)** Bar plot showing the percentage breakdown of significant compartment switch 100kb bins across the whole genome (WG, n=1,938) and for regions that also had at least one change in insulation score in the top 10% (Top 10%, n=279). P-values from Chi-square test are shown for significant changes. **(D)** Scatter plot of the change in insulation score (10kb bins, n=2,691) vs. the change in PC1 value (100kb bins, n=279) expressed as HIVE – HIV- for those regions that have a significant change in insulation score in the top 10% and a significant compartment switch. Pearson correlation line and statistics are indicated. **(E)** Results of

overlap permutation (n=1000) test between all HIVE and HIV- loop anchors (top) or differential HIVE and HIV- loop anchors (bottom) and monocyte CTCF ChIP-seq peaks from ENCODE. Red bar shows the observed number of overlaps between CTCF ChIP peaks and insulation score regions, black line shows the overlap expected by chance given the whole genome as a background. Z-score and p-value are indicated. **(F)** Browser shot showing shared looping changes in HIVE and IFN-γ treated cells. Tracks from top to bottom: loops in HMC3 cells with IFN-γ gained loops in red and IFN-γ lost loops in gray, loops in human brain microglia with HIVE gained loops in red and HIVE lost loops in gray, DPC1 values in HMC3 cells, ΔPC1 in human brain microglia, microglial gene expression from snRNA-seq of HIV- brain, microglial genes expression from snRNA-seq of HIVE brain.

Figure S5: Shared transcriptional and 3D genomic changes in HIV+ and HIVE microglia. **Related to Figure 2.** (A) Pearson correlation between the $\log_2$(fold change) in expression of DEGs (shown in Figure 2I) of HIVE vs. HIV- microglia (Y-axis) and HIV+ vs. HIV- microglia (X-axis). **(B)** Volcano plot showing -$\log_{10}$(adjust p-value) (Y-axis) and $\log_2$(Fold Change) (X-axis) for snRNA-seq differential expression of the microglial cluster (Figure 1C, Mg) from n=3 HIV- and n=3 HIV+ brains. Points in green are genes with a $\log_2$(FC)>1 and padj<0.7. **(C)** Biological process GO enrichment for downregulated genes from (B). Red line marks significance for FDR=0.05, numbers next to each bar show the n genes differentially expression/the total n genes in that GO category. **(D)** Venn diagrams showing the overlap of HIVE and HIV+ DEGs. Left, upregulated genes; right, downregulated genes. **(E)** Biological process GO enrichment for downregulated genes shared by HIVE and HIV+. Red line marks significance for FDR=0.05, numbers next to each bar show the n genes differentially expression/the total n genes in that GO category. **(F)** Pearson correlation between the ΔPC1 of HIVE and HIV+ microglia for those 100kb bins with a significant compartment switch in HIVE (N=1,938).

(G) Pearson correlation between the ΔPC1 of HIVE and HIV+ microglia for those 100kb bins with a significant compartment switch in both HIVE and HIV+ (n=339). **(H-I)** Cell component (H) or biological process (I) GO enrichment for genes in regions with significant compartment switching in both HIVE and HIV+ microglia. Shown for genes with significant decrease in PC1 (H) or genes with significant increase in PC1 (I). Red line marks significance for FDR=0.05, numbers next to each bar show the n genes differentially expression/the total n genes in that GO category.

Figure S6: HIV IS-seq validation and analysis. Related to Figures 4-6. (A) Schematic of integration site library preparation. Sonication and linker ligation was following by nested PCR employing a blocking oligonucleotide to prevent amplification from the 5'LTR. **(B)** Bioinformatic pipeline used for identification of integration sites. **(C)** Browser shots showing results of Jlat integration site sequencing compared to previously published Jlat sequencing by Symons et al.[1] Uninfected Jurkat DNA was used as a negative control. **(D)** Graph showing titration of Jlat clone integration site sequencing. The number of Jlat genomes input into the library prep is graphed against the number of IS detected. **(E-F)** Percent of IS that are recurrent (E) or clonal (F) for all NeuN- IS, *in vivo* PBMC IS from publicly available IS-seq data[2] here analyzed with our pipeline, and integration sites from *in vitro* infected Jurkat cells. Significance by chi-square (E) or two-way ANOVA (F). **(G)** Spie charts comparing the distribution of IS across genomic features to the distribution of those features in the genome. The angle of the pie slice represents the proportion of genome occupied by that feature, while the height of the pie slice represents the relative proportion of that feature in the IS dataset as compared to the genome, dashed circle represents equal proportions. Percentages label the percent of IS found in each genomic feature. Significance using Chi-square test. * $0.01 < p < 0.05$, ** $0.001 < p < 0.01$, *** $0.0001 < p < 0.001$, **** $p<0.0001$. **(H)** Violin plots showing the distance between IS and the nearest ChIP-seq peak.

NeuN- IS are compared to NeuN- ChIP-seq and PBMC and Jurkat datasets are compared to Jurkat ChIP-seq. Significance from Kruskall Wallis rank sum test. **(I-J)** Violin plots showing expression of *CPSF6* (I) and *LEDGF* (J) from snRNA-seq data for each cell type. For cell types composed of more than 1 subcluster, values were averaged across all cells. P-values reflect DESeq2 differential expression of Mg vs. all other cell types combined. **(K-L)** Plots showing publicly available CPSF6 ChIP-seq read coverage [3] centered on NeuN- (K) or T-cell (L) IS. **(M-N)** Bar plots showing transcript per million (TPM) expression of *CPSF6* (M) and *LEDGF* (N) in human adult primary microglia and the C20 human immortalized microglial cell line from a previously published study [4]. Two-tailed Student's t-test. **(O)** Absolute value of the change in insulation score between HIVE and HIV- microglia for each 10kb bin of the genome that contained an IS. An overlap permutation (n=1000) test was performed between bins in either top or bottom 10% and ChIP-seq peaks from ENCODE for YY1 from GM12878 cells and CTCF from CD14+ monocytes. Red bar shows the observed number of overlaps between YY1/CTCF ChIP peaks and insulation score regions, black line shows the overlap expected by chance given the whole genome as a background.

Figure S7: HIV RNA+ microglia are more activated than HIV RNA- microglia. Related to Figure 7. (A) PCA of RNA+ and RNA- cells from each HIVE brain with HIV mapping reads present. Donor numbers are indicated. **(B)** Enrichment for ENCODE transcription factor (TF) binding to downregulated genes in HIV RNA+ microglia (shown in Figure 7B). Red line marks significance for FDR=0.05, numbers show the number of differential genes bound by the TF/the total number of genes bound by the TF. **(C)** Ridge plots showing log normalized expression of CTBP2 and SUZ12 in HIV RNA+ and RNA- HIVE microglia. **(D)** Pearson correlation of the $\log_2$(Fold Change) in expression between HIV RNA- HIVE Mg vs. HIV- Mg (Y-axis) and the $\log_2$(Fold Change) in expression between

HIV RNA+ HIVE Mg vs. HIV- Mg (X-axis). Only HIVE vs. HIV- DEGs are shown (from Figure 2I). **(E)** Violin plots showing the absolute $\log_2$(Fold Change) for HIV RNA- (gray) or HIV RNA+ (blue) HIVE Mg. Wilcoxon rank sum test.

Supplementary Table 1: Donor characteristics and assays performed. Related to Figures 1-7.

Abbreviations: Und., undetectable; n.a. not applicable; n.d. assay not performed

Threshold for undetectable plasma viral loads: 20 copies/ml

† cART, whether on cART within 2 weeks of death; unless noted as being cART-naïve, these individuals had prior cART exposures during their lifetimes. Individuals off cART at death with "^" following "No" value: historical cART exposure could not be confirmed.

‡ Results of qPCR for HIV DNA in DNA extracted from bulk tissue performed by the Manhattan HIV Brain Bank; values followed by asterisk expressed as copies per ug of tissue.

§ Assays performed for each brain. Boxes checked mean that assay was performed for that donor. ISA, integration site analysis; RNA-seq, bulk RNA-seq from FANS-isolated nuclei; snRNA-seq, single nucleus RNA-seq

‖ The number of times integration site libraries were prepared for either NeuN+ or NeuN- nuclei. Stars indicate that a library was prepared but the concentration was too low for sequencing.

References

1. Symons, J., Chopra, A., Malatinkova, E., De Spiegelaere, W., Leary, S., Cooper, D., Abana, C.O., Rhodes, A., Rezaei, S.D., Vandekerckhove, L., et al. (2017). HIV integration sites in latently infected cell lines: evidence of ongoing replication. Retrovirology *14*, 2. 10.1186/s12977-016-0325-2.
2. Maldarelli, F., Wu, X., Su, L., Simonetti, F.R., Shao, W., Hill, S., Spindler, J., Ferris, A.L., Mellors, J.W., Kearney, M.F., et al. (2014). HIV latency. Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. Science *345*, 179-183. 10.1126/science.1254194.
3. Jalihal, A.P., Pitchiaya, S., Xiao, L., Bawa, P., Jiang, X., Bedi, K., Parolia, A., Cieslik, M., Ljungman, M., Chinnaiyan, A.M., and Walter, N.G. (2020). Multivalent Proteins Rapidly and Reversibly Phase-Separate upon Osmotic Cell Volume Change. Mol Cell *79*, 978-990.e975. 10.1016/j.molcel.2020.08.004.
4. Rai, M.A., Hammonds, J., Pujato, M., Mayhew, C., Roskin, K., and Spearman, P. (2020). Comparative analysis of human microglial models for studies of HIV replication and pathogenesis. Retrovirology *17*, 35. 10.1186/s12977-020-00544-y.