JCI insight

Single-cell RNA sequencing reveals microglia-like cells in cerebrospinal fluid during virologically suppressed HIV

Shelli F. Farhadian,[1,2] Sameet S. Mehta,[3] Chrysoula Zografou,[2,4] Kevin Robertson,[5] Richard W. Price,[6] Jenna Pappalardo,[3,4] Jennifer Chiarella,[2] David A. Hafler,[2,4] and Serena S. Spudich[2]

[1]Department of Medicine, Section of Infectious Diseases, [2]Department of Neurology, [3]Yale Center for Genome Analysis, and [4]Department of Immunobiology, Yale School of Medicine, New Haven, Connecticut, USA. [5]Department of Neurology, University of North Carolina, Chapel Hill, North Carolina, USA. [6]Department of Neurology, University of California San Francisco, San Francisco, California, USA.

Central nervous system (CNS) immune activation is an important driver of neuronal injury during several neurodegenerative and neuroinflammatory diseases. During HIV infection, CNS immune activation is associated with high rates of neurocognitive impairment, even during sustained long-term suppressive antiretroviral therapy (ART). However, the cellular subsets that drive immune activation and neuronal damage in the CNS during HIV infection and other neurological conditions remain unknown, in part because CNS cells are difficult to access in living humans. Using single-cell RNA sequencing (scRNA-seq) on cerebrospinal fluid (CSF) and blood from adults with and without HIV, we identified a rare (<5% of cells) subset of myeloid cells that are found only in CSF and that present a gene expression signature that overlaps significantly with neurodegenerative disease–associated microglia. This highlights the power of scRNA-seq of CSF to identify rare CNS immune cell subsets that may perpetuate neuronal injury during HIV infection and other conditions.

Introduction

Neuronal injury during infections and other inflammatory processes may be due to both the direct effect of a toxic pathogen as well as bystander effect of nearby, activated immune cells. During HIV infection, central nervous system (CNS) immune activation has been implicated in the stubbornly high rates of neurological impairment seen in adults with HIV, a phenomenon that persists even during virological suppression with antiretroviral therapy (ART) (1–3). However, the specific cellular subsets and genes that drive CNS immune activation during HIV and other neuroinflammatory conditions remain incompletely understood. This is in part because brain tissue is largely inaccessible for routine studies of neurological disease. New tools are thus needed to interrogate CNS cells in living humans.

Cerebrospinal fluid is (CSF) produced within the brain in the choroid plexus and we have previously shown in other neuroinflammatory diseases that CSF immune cells reflect infiltrating brain parenchymal immune cells (4, 5). Thus, CSF can provide unique diagnostic information in infectious, inflammatory, and neurodegenerative CNS disorders (4–8). In adults with HIV infection, while rare studies have analyzed brain tissue obtained at autopsy of affected patients to understand the cellular basis for persistent CNS immune activation during HIV infection, a majority of studies have used CSF analysis, though primarily focusing on soluble biomarkers that reflect nonspecific immune activation, or on flow cytometry analysis of CSF immune cells (9–12). However, a major limitation of flow cytometry is that cell populations are identified based only on a small set of surface markers, thus potentially missing rare and important cells that are unique to the CNS and which may be transcriptionally distinct but may not display different surface markers.

Myeloid-lineage cells in the CNS, including microglia and circulating monocytes, are particularly challenging to study using conventional approaches. These cells have been proposed to play a role in causing or exacerbating neuronal injury during HIV infection. During acute infection, monocytes traffic to the CNS where they have been proposed to contribute to neuroinflammation and may contribute to the establishment of latent CNS infection (13). Microglia, the resident tissue macrophages of the CNS, are a potential site for persistent latent infection or low-level viral replication in the CNS, as well as a potential

driver of neuronal injury, even during ART (14, 15). CSF studies in adults with virologically suppressed HIV have revealed elevated levels of neopterin and soluble CD163, biomarkers that are associated with monocyte and microglial activation, further supporting a role for monocytes and microglia in perpetuating neuronal injury during HIV infection (9, 16). However, since myeloid cells comprise a minor population of CSF cells (<10%), they are challenging to detect and fully characterize through cytometry-based studies (9). Moreover, microglia have thus far only been characterized in humans through autopsy brain tissue or through in vitro cell culture–based studies.

Here, we used massively parallel single-cell RNA-sequencing (scRNA-seq) to characterize the immune cell landscape of CSF in HIV-infected individuals with virologic suppression. Recent advances in scRNA-seq now allow for simultaneous examination of more than 10,000 single-cell transcriptomes and have led to the characterization of new immune cell subsets when applied to whole blood as well as various human tissues in both healthy and diseased states (17–19). We performed scRNA-seq on large-volume CSF and paired blood samples from individuals with virologically suppressed HIV to discover cell populations that are enriched in the CSF, including myeloid cell subsets. By studying CSF immune cells in an unbiased, surface-marker-free approach, we sought to identify de novo cell populations that are associated with CNS immune activation.

Results

Unbiased discovery of CSF-specific myeloid cells. For the initial identification of CSF-specific cellular subsets, we performed scRNA-seq on blood and CSF derived from 2 HIV+ participants (HIV1 and HIV2), with clinical characteristics as in Supplemental Table 1 (supplemental material available online with this article; https://doi.org/10.1172/jci.insight.121718DS1). High-quality cDNA was produced from single-cell whole transcriptomes derived from CSF and blood that were collected in parallel (Figure 1A). We detected a mean of 775 genes per cell in CSF and 758 genes per cell in blood. CSF and peripheral blood mononuclear cell (PBMC) single-cell transcriptomes contained similar proportions of mitochondrial genes (mean 10%). We pooled blood and CSF cells for the initial analysis, since prior scRNA-seq studies suggest that including a large number of cells (>10,000) in the analysis increases the likelihood of detecting rare cells (20–23). A total of 8,774 single-cell transcriptomes derived from CSF (3,160 cells) and PBMCs (5,580 cells) were pooled for initial unsupervised cluster analysis that did not rely on known markers of cell types, thus allowing for the identification of potentially novel cell populations. Fourteen clusters representing distinct immune cell subsets were identified (Figure 1B). Identities were assigned to cell clusters based on expression levels of canonical marker genes and of differentially expressed genes (Supplemental Figure 1). We then overlaid the tissue of origin of each cell (i.e., blood or CSF), and in this way identified differences in cellular distributions between blood and CSF (Figure 1C). We found that a majority of cells in CSF were T lymphocytes (89%), with smaller populations of myeloid cells (7%), NK cells (<2%), and B cells and plasma cells (<1%). When compared with blood, CSF contained higher proportions of CD8+ T cells, and 2 myeloid cellular subsets (Figure 1D).

We next performed subanalysis of the myeloid cells identified from the larger, combined sample of blood and CSF cells (Figure 2A). We found 5 distinct myeloid cellular subsets, 2 of which consisted of cells found predominantly (>50% of cells) in CSF (Myeloid-2 and Myeloid-5). Differential gene expression analysis revealed that Myeloid-5 consists of cells with high levels of expression of genes found in myeloid-derived dendritic cells, including *CD1C* and *FCER1A*.

Myeloid-2 is a myeloid subset composed almost exclusively of CSF cells and is characterized by high expression of 60 distinct, nonribosomal genes when compared with all other myeloid subsets (FDR < 0.01) (Supplemental Table 2). Several of these genes are produced in CNS almost exclusively by microglia, including *C1QA–C* and *TREM2* (24, 25). We therefore compared the genes that characterize Myeloid-2 to recently published transcriptomic studies that examined microglia derived from mouse models of neurodegenerative diseases (26, 27) (Figure 2B). We found significant overlap between genes that are overexpressed in Myeloid-2 and genes that are enriched in neurodegenerative disease–associated microglia, including *APOE*, *AXL*, and *TREM2* ($P < 10^{-13}$ and $P < 10^{-4}$ for comparisons to refs. 26 and 27, respectively). Furthermore, we found that, compared with the 4 other myeloid subsets we identified, Myeloid-2 expressed higher levels of *CTSB*, *APOC1*, and *MSR1* (CD204), molecules that are also associated with the microglial neurodegenerative phenotype (22). We therefore reason that Myeloid-2 consists of CSF neurodegenerative disease–associated microglia-like cells.
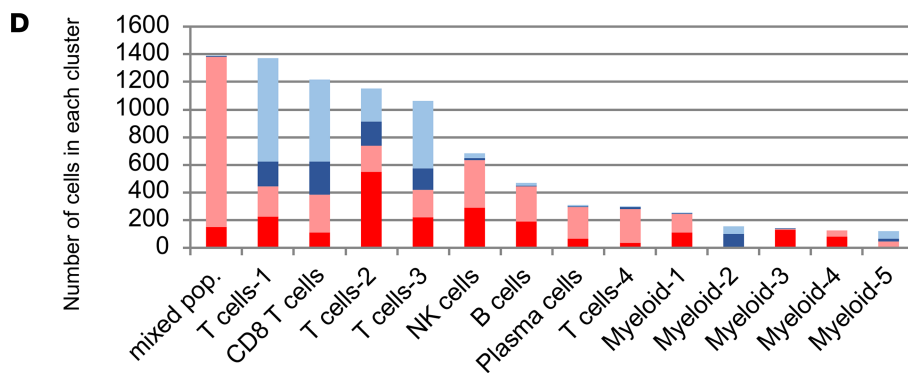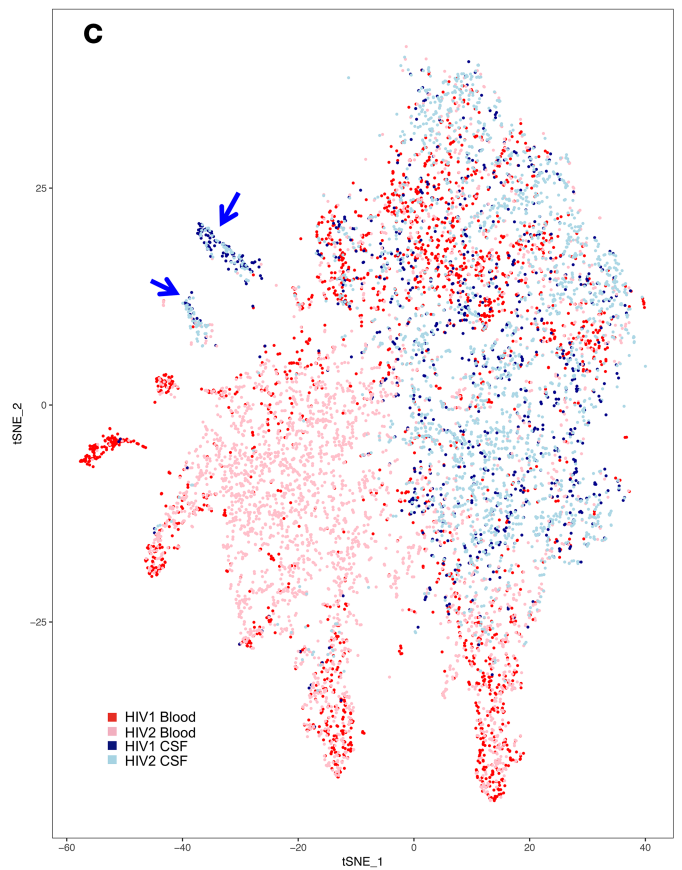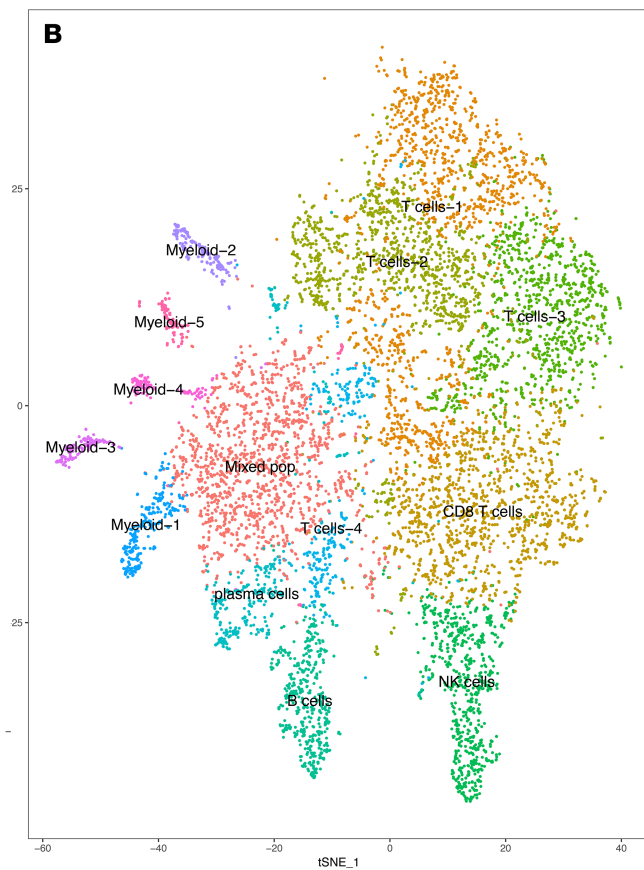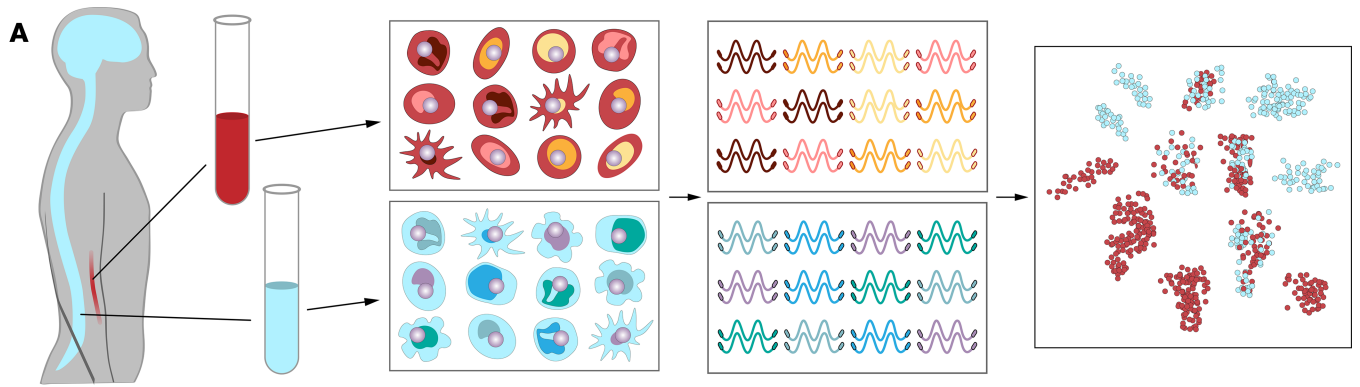
Figure 1. Single-cell RNA sequencing of paired blood and CSF samples. (**A**) CSF and blood were collected and processed in parallel. Single PBMCs and CSF cells were applied to separate arrays, with 10,000 cells applied to each SeqWell array. Single-cell whole transcriptomes were tagged to retain information about the cell, and then sequenced. Sequencing data from single cells was pooled for cluster analysis based on differential gene expression, revealing clusters that are composed predominantly of CSF cells. (**B**) t-Distributed Stochastic Neighbor Embedding (tSNE) visualization of cell populations identified through unsupervised cluster analysis of blood and CSF derived from 2 HIV+ individuals (HIV1 and HIV2). Clusters are labeled based on expression of canonical marker genes (see also Supplemental Figure 1). (**C**) tSNE visualization of cells, labeled by sample (HIV1 or HIV2) and tissue (CSF or blood) of origin, revealing 2 myeloid clusters that are predominantly composed of CSF cells (blue arrows). (**D**) Bar plots quantifying the number of cells from blood and CSF from each sample in each cluster.

Validation of CSF-specific microglia-like cells in HIV+ and uninfected subjects. To validate the presence of these CSF-specific cells, and to determine if they were more frequent in HIV+ individuals than in uninfected controls, we asked whether Myeloid-2–like cells could be detected in CSF in an independent sample derived from a third HIV+ participant (HIV3). We found a group of 93 cells with high aggregate expression of the set of genes that defined the Myeloid-2 population of CSF-associated microglia-like cells (Figure 2C). This represents 3.3% of all cells in the independent CSF sample. This indicates that the rare, CSF-associated microglia-like cells we identified earlier are also present in CSF from an independent HIV+ participant. We next examined CSF from 2 HIV-uninfected research participants (Supplemental Table 1; Uninfected 1 and Uninfected 2) for the presence of Myeloid-2–like cells (Figure 2D). Myeloid-2 transcripts were detected in CSF cells in the uninfected participants, but at a significantly lower frequency when compared with CSF cells derived from HIV+ participants ($P < 10^{-7}$). Both groups of CSF cells (HIV+ and uninfected) had significantly higher rates of Myeloid-2 transcripts when compared with blood cells ($P < 10^{-16}$ and $P < 0.07$, respectively).

Discussion

Using an unbiased, surface-marker-free approach to characterize CNS immune cell populations during virologically suppressed HIV infection, we identified what we believe is a novel myeloid subset in CSF with gene expression characteristics of neurodegenerative disease–associated microglia. To our knowledge, this is the first study to characterize the full immune cell landscape in CSF at the transcriptional level, in any disease, and the first to identify circulating microglia-like cells in CSF. These CSF-associated microglia-like cells are rare, representing less than 5% of all cells we analyzed in CSF, and thus would likely not be reliably detected using traditional flow cytometry–based studies of CSF. We analyzed CSF cells from 3 HIV+ and 2 uninfected participants and found a significant trend toward higher frequencies of these microglia-like cells in the CSF of HIV+ individuals, suggesting a potential link between the presence of these cells and chronic immune activation in the CNS during HIV infection.

CNS myeloid cells, and microglia in particular, have been proposed to play several important roles in HIV infection, including acting as a potential site of CNS viral replication. In their normal, homeostatic state, microglia play an important role in maintaining neuronal integrity, by promoting immune responses to infection and through clearance of debris and plaque. During chronic neurological diseases, including Alzheimer's disease and ALS, microglia may become dysfunctional and promote neuronal injury. This switch, from homeostatic to neurodegenerative microglia, has been linked to an APOE/TREM2 signaling pathway that disrupts the normal microglial response to damaged neurons and other debris. Likewise, microglial dysregulation has been proposed as a possible mechanism for neuronal toxicity during HIV infection. Our analysis of publically available transcriptome data from mouse models of neurodegenerative disease and our current study of chronic HIV infection suggests a potential common role for the APOE/TREM2 pathway in mediating neurodegeneration during these conditions. Further studies in human CNS tissue will clarify the extent to which neuropathology during HIV infection overlaps with neuropathology during other neuroinflammatory and neurodegenerative disease states.

An important question is whether compartmentalized, low-level viral replication in the CNS drives persistent microglia and macrophage activity even during HIV treatment. Two of the 3 HIV+ participants in this study had evidence for CSF escape or detectable HIV RNA in the CNS, despite effective plasma viral suppression; thus, we cannot draw definitive conclusions regarding an association between CNS viral replication and the frequency of neurodegenerative microglia-like cells in CSF. Our study is also limited by the small number of research participants, limiting the ability to draw associations between HIV-specific CSF immune cells and markers of neuroinflammation and
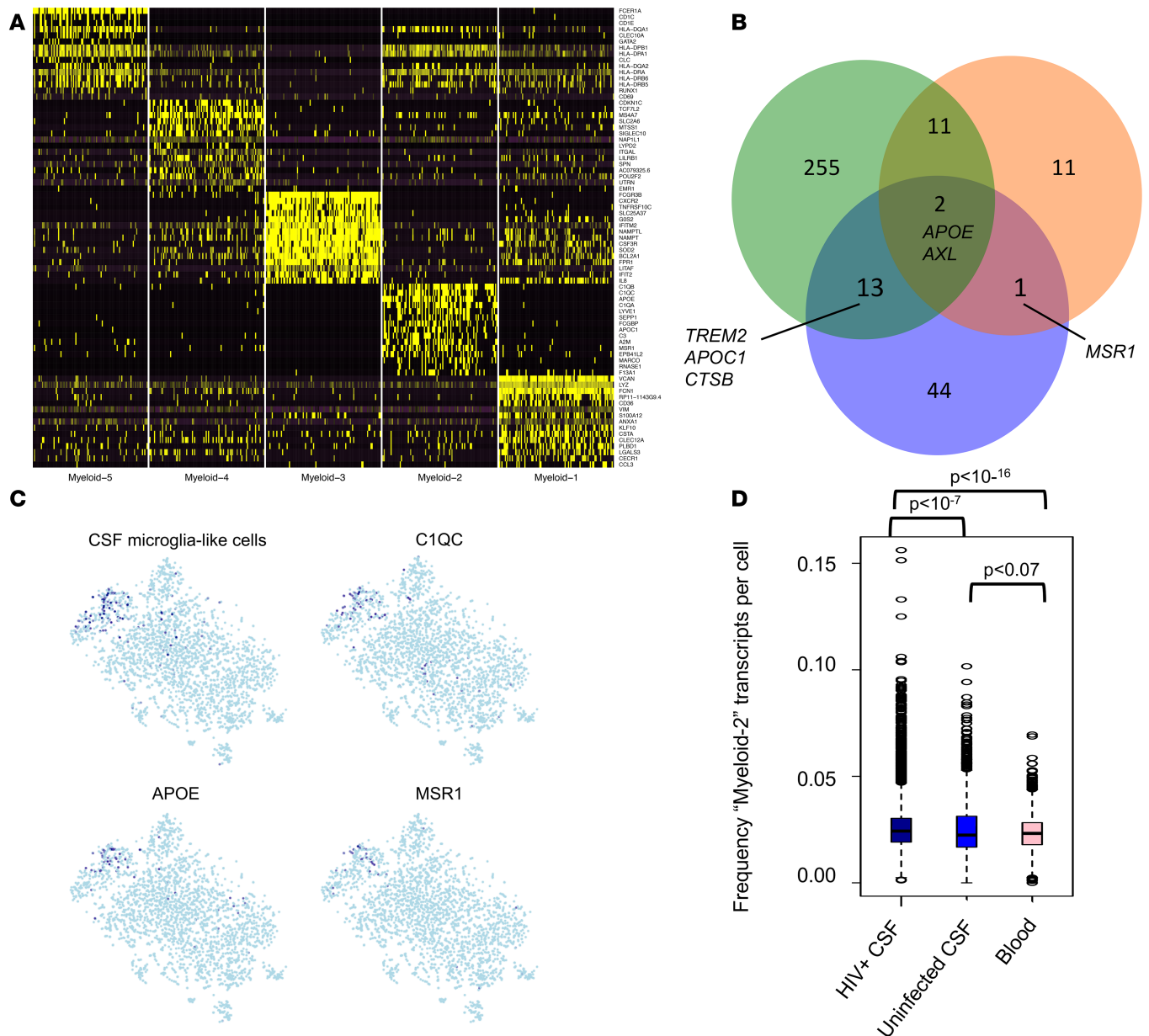
Figure 2. Single-cell RNA sequencing reveals a disease-associated microglia-like cellular subset in cerebrospinal fluid. (**A**) Heatmap showing the top 15 genes whose expression levels are highest and most differentiating in each of the 5 myeloid subsets. (**B**) Venn diagram shows genes that were found in Myeloid-2 (blue) and across 2 independent transcriptomic studies of neurodegenerative disease–associated microglial cells derived from mouse brain (green [ref. 16] and orange [ref. 27]), with $P < 10^{-13}$ and $P < 10^{-4}$ for the probability of finding overlapping genes when comparing the Myeloid-2 subset to refs. 26 and 27, respectively). (**C**) Examination of an independent CSF sample from an HIV+ participant (HIV3) reveals a group of cells with high expression of the genes that characterize the CSF microglia-like myeloid subset (Myeloid-2). (**D**) Box plots showing the percentage of Myeloid-2 transcripts per cell in HIV+ CSF (n = 5,919 pooled cells), uninfected CSF (n = 1,770 pooled), and blood (n = 5,581 pooled cells), with P values comparing the distribution of Myeloid-2 transcripts between samples (Wilcoxon's rank-sum test). Shown are the median (black line) and 25th and 75th percentiles (box outlines).

cognitive impairment. Addressing these questions will require larger studies to specifically assess the relationship between detection of HIV in the CNS, markers of neuroinflammation, and the presence of specific cellular subsets in CSF.

In conclusion, our study identified microglia-like cells in human CSF with a transcriptional profile that is similar to disease-associated microglia in animal models of neurodegenerative disease. This demonstrates a potential mechanistic link between pathways of neuronal injury in HIV and other neurodegenerative conditions. Furthermore, our study illustrates the potential for genomic-based studies of CSF to uncover rare immune cell populations that may drive a wide range of CNS disease.

Methods

Study participants and procedures. Research participants were seen for study visits at the Yale School of Medicine. Participants were HIV+ and on stable ART with plasma HIV RNA levels below 20 copies/ml for more than 1 year. None had active neurological disease or other infection. Blood and CSF were processed separately for scRNA-seq using SeqWell (19). Two HIV-uninfected participants were recruited through word-of-mouth referrals from HIV+ participants.

CSF and blood processing. Fresh CSF and blood were processed within 1 hour of collection. CSF was centrifuged at 400 g for 10 minutes, and the isolated cells were immediately applied to SeqWell arrays, which were provided by the Love Lab at MIT/Koch Institute (Cambridge, Massachusetts, USA). PBMCs were isolated via Ficoll gradient. Approximately 10,000 cells were loaded onto each SeqWell array, with 1 array per CSF or PBMC sample. Cells were processed as in Geirahn et al. (19). Briefly, cells were added to custom microwell arrays that were preloaded with Drop-Seq beads (Chemgenes CSO-2011). These beads contain oligonucleotides comprising a unique cellular barcode, poly-T for mRNA capture, a unique molecular identifier, and a PCR handle (18). Each microwell thus ideally contained 1 barcoded bead and 1 cell. A semipermeable membrane was applied to each array, cells were lysed, and mRNA was hybridized to the barcoded beads. Membranes were then removed, microwell contents were pooled, cDNA was generated, and Nextera sequencing libraries were prepared. Sequencing was performed on the Illumina HiSeq 4000 platform at approximately 50,000 reads per cell. scRNA-seq data were deposited in NCBI's Gene Expression Omnibus database (GEO GSE117397).

Single-cell transcriptome analysis. Low-quality cells containing less than 500 or more than 2,500 genes detected were removed, as well as the 10% of cells with the highest mitochondrial content. Genes that were present in less than 3 cells were excluded from analysis. Gene expression values were then normalized, scaled, and log transformed.

The detection of highly variable genes and unbiased identification of cell clusters was performed with Seurat (28). Single-cell transcriptomes from CSF and blood from the first 2 HIV+ participants were pooled prior to unsupervised cluster analysis. Four clusters that were characterized primarily by mitochondrial and ribosomal gene transcripts (indicating low-quality cells) were removed. Histograms were generated to determine the relative contributions of CSF and of blood cells to each cluster, thus permitting the identification of clusters that were composed predominantly of CSF cells. For analysis of the independent HIV+ CSF sample and of 2 HIV-uninfected CSF samples (Figure 2, C and D), we performed supervised analysis to look for cells that contained the transcripts that were previously identified as marker genes for Myeloid-2, the microglia-like subset identified during the unsupervised analysis.

Neuropsychological testing. Neuropsychological testing was conducted across the following domains: language/fluency, executive function, speed of information processing, attention/working memory, verbal and visual learning, verbal and visual memory, and motor skills (29). All measures, except timed gait, were normalized according to age, education, gender, and ethnicity. Measures were evaluated as domain scores (average of test Z scored for each domain) and summarized as the total Z score for all tests.

Statistics. The probability of finding overlapping genes between gene sets was calculated using the hypergeometric probability formula: $C(D, x) \times C(N - D, n - x)/C(N,n)$, available at www.nemates.org. Cluster-defining genes were deemed significant if the FDR for differentially expressed genes, as reported by the FindAllMarkers function in Seurat, was less than 0.01. To compare the distribution of the frequency of Myeloid-2 transcripts in single cells across the three conditions (HIV+ CSF, uninfected CSF, and blood), Wilcoxon's rank-sum test was used.

Study approval. This study was approved by the Yale University Human Investigations Committee. Informed consent was obtained using protocols approved by the Yale Human Research Protection Program.

Author contributions

SFF and SSS designed the study. SFF and CZ conducted SeqWell experiments. SSM and SFF performed data analysis. JP contributed to data analysis. JC contributed to participant recruitment. DAH provided reagents and contributed to study design. KR performed neuropsychological testing analysis. RWP contributed to the analysis of CSF escape. SFF and SSS wrote the manuscript.

1. Tozzi V, et al. Persistence of neuropsychologic deficits despite long-term highly active antiretroviral therapy in patients with HIV-related neurocognitive impairment: prevalence and risk factors. *J Acquir Immune Defic Syndr*. 2007;45(2):174–182.
2. Heaton RK, et al. Neurocognitive change in the era of HIV combination antiretroviral therapy: the longitudinal CHARTER study. *Clin Infect Dis*. 2015;60(3):473–480.
3. Grauer OM, et al. Neurocognitive decline in HIV patients is associated with ongoing T-cell activation in the cerebrospinal fluid. *Ann Clin Transl Neurol*. 2015;2(9):906–919.
4. Lovato L, et al. Related B cell clones populate the meninges and parenchyma of patients with multiple sclerosis. *Brain*. 2011;134(Pt 2):534–541.
5. Stern JN, et al. B cells populating the multiple sclerosis brain mature in the draining cervical lymph nodes. *Sci Transl Med*. 2014;6(248):248ra107.
6. Iorio R, Lennon VA. Neural antigen-specific autoimmune disorders. *Immunol Rev*. 2012;248(1):104–121.
7. Damkier HH, Brown PD, Praetorius J. Cerebrospinal fluid secretion by the choroid plexus. *Physiol Rev*. 2013;93(4):1847–1892.
8. Tyler KL. Herpes simplex virus infections of the central nervous system: encephalitis and meningitis, including Mollaret's. *Herpes*. 2004;11 Suppl 2:57A–64A.
9. Ho EL, Ronquillo R, Altmeppen H, Spudich SS, Price RW, Sinclair E. Cellular composition of cerebrospinal fluid in HIV-1 infected and uninfected subjects. *PLoS ONE*. 2013;8(6):e66188.
10. Yilmaz A, et al. Cerebrospinal fluid neopterin decay characteristics after initiation of antiretroviral therapy. *J Neuroinflammation*. 2013;10:62.
11. Edén A, et al. Increased intrathecal immune activation in virally suppressed HIV-1 infected patients with neurocognitive impairment. *PLoS ONE*. 2016;11(6):e0157160.
12. Burdo TH, Weiffenbach A, Woods SP, Letendre S, Ellis RJ, Williams KC. Elevated sCD163 in plasma but not cerebrospinal fluid is a marker of neurocognitive impairment in HIV infection. *AIDS*. 2013;27(9):1387–1395.
13. Williams DW, Veenstra M, Gaskill PJ, Morgello S, Calderon TM, Berman JW. Monocytes mediate HIV neuropathogenesis: mechanisms that contribute to HIV associated neurocognitive disorders. *Curr HIV Res*. 2014;12(2):85–96.
14. Williams K, Burdo TH. Monocyte mobilization, activation markers, and unique macrophage populations in the brain: observations from SIV infected monkeys are informative with regard to pathogenic mechanisms of HIV infection in humans. *J Neuroimmune Pharmacol*. 2012;7(2):363–371.
15. Honeycutt JB, et al. Macrophages sustain HIV replication in vivo independently of T cells. *J Clin Invest*. 2016;126(4):1353–1366.
16. Neuenburg JK, et al. T-cell activation and memory phenotypes in cerebrospinal fluid during HIV infection. *J Acquir Immune Defic Syndr*. 2005;39(1):16–22.
17. Gierahn TM, et al. Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. *Nat Methods*. 2017;14(4):395–398.
18. Macosko EZ, et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell*. 2015;161(5):1202–1214.
19. Villani AC, et al. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science*. 2017;356(6335):eaah4573.
20. Torre E, et al. Rare cell detection by single-cell RNA sequencing as guided by single-molecule RNA FISH. *Cell Syst*. 2018;6(2):171–179.e5.
21. Campbell JN, et al. A molecular census of arcuate hypothalamus and median eminence cell types. *Nat Neurosci*. 2017;20(3):484–496.
22. Grün D, et al. Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature*. 2015;525(7568):251–255.
23. Segerstolpe Å, et al. Single-cell transcriptome profiling of human pancreatic islets in health and type 2 diabetes. *Cell Metab*. 2016;24(4):593–607.
24. Fonseca MI, et al. Cell-specific deletion of C1qa identifies microglia as the dominant source of C1q in mouse brain. *J Neuroinflammation*. 2017;14(1):48.
25. Butovsky O, et al. Identification of a unique TGF-β–dependent molecular and functional signature in microglia. *Nat Neurosci*. 2014;17(1):131–143.
26. Keren-Shaul H, et al. A unique microglia type associated with restricting development of Alzheimer's disease. *Cell*. 2017;169(7):1276–1290.e17.
27. Krasemann S, et al. The TREM2-APOE pathway drives the transcriptional phenotype of dysfunctional microglia in neurodegenerative diseases. *Immunity*. 2017;47(3):566–581.e9.
28. Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol*. 2015;33(5):495–502.
29. Robertson K, Yosief S. Neurocognitive assessment in the diagnosis of HIV-associated neurocognitive disorders. *Semin Neurol*. 2014;34(1):21–26.