Methods to investigate intrathecal adaptive immunity in neurodegeneration

Hamilton Oh
Stanford University
Olivia Leventhal
Stanford University
Divya Channappa
Stanford University
Victor W. Henderson
Stanford University
Tony Wyss-Coray
Stanford University
Benoit Lehallier
Stanford University
David Gate (dgate@stanford.edu)
Stanford University https://orcid.org/0000-0003-0481-9657

Methodology

Keywords: cerebrospinal fluid cells, CSF, intrathecal cells, neurodegeneration, adaptive immunity, T cells, T cell receptor (TCR), antigen

Posted Date: January 4th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-51188/v2

Version of Record: A version of this preprint was published on January 22nd, 2021. See the published version at https://doi.org/10.1186/s13024-021-00423-w.

Abstract

Background: Cerebrospinal fluid (CSF) provides basic mechanical and immunological protection to the brain. Historically, analysis of CSF has focused on protein changes, yet recent studies have shed light on cellular alterations. Evidence now exists for involvement of intrathecal T cells in the pathobiology of neurodegenerative diseases. However, a standardized method for long-term preservation of CSF immune cells is lacking. Further, the functional role of CSF T cells and their cognate antigens in neurodegenerative diseases are largely unknown.

Results: We present a method for long-term cryopreservation of CSF immune cells for downstream single cell RNA and T cell receptor sequencing (scRNA-TCRseq) analysis. We observe preservation of CSF immune cells, consisting primarily of memory $CD4^+$ and $CD8^+$ T cells. We then utilize unbiased bioinformatics approaches to quantify and visualize TCR sequence similarity within and between disease groups. By this method, we identify clusters of disease-associated, antigen-specific TCRs from clonally expanded CSF T cells of patients with neurodegenerative diseases.

Conclusions: Here, we provide a standardized approach for long-term storage of CSF immune cells. Additionally, we present unbiased bioinformatic approaches that will facilitate the discovery of target antigens of clonally expanded T cells in neurodegenerative diseases. These novel methods will help improve our understanding of adaptive immunity in the central nervous system.

Background

The cerebrospinal fluid (CSF) provides insight into brain physiology of living individuals. Biochemical analysis of CSF is routinely utilized as a diagnostic tool in neurodegeneration (1-3). For example, in Alzheimer's disease (AD), changes in protein levels of tau and amyloid-β are indicative of disease pathology. However, while CSF protein biomarkers guide diagnosis of patients with neurodegenerative diseases, the cells patrolling the interstitial fluid are often centrifuged and discarded. Only in cases of extreme central nervous system inflammation, such as meningitis or encephalitis, are CSF cells utilized as a diagnostic. Blood cells, on the other hand, are routinely used to assess health and disease. We surveyed the literature to gain insight into CSF cell composition and summarized the main studies (Table 1). The CSF is primarily composed of antigen-experienced memory $CD4^+$ and $CD8^+$ T cells, plasmacytoid dendritic cells, and $CD56^{high}$ natural killer (NK) cells. Naïve T cells, monocytes, granulocytes, myeloid dendritic cells, basophils, and B cells are not as abundant in the CSF as they are in peripheral blood (2). Moreover, the cellular density of CSF (1,000-3000 cells per mL) is drastically diluted compared to blood (millions of cells per mL) (1, 4, 5).

To our knowledge, only the multiple sclerosis field has rigorously evaluated live CSF cells to assess disease (1, 5-14). However, recent studies provide evidence of CSF T cells in the pathobiology of other neurodegenerative diseases. These include the discovery of hypocretin-specific $CD8^+$ T cells in narcolepsy (15) and T cell clones shared between Rasmussen's encephalitis patients (16). Moreover, a

recent report identified α-synuclein-specific T cells in peripheral blood of Parkinson's disease (PD) patients (17), but whether these cells enter the CSF is unknown. Finally, our group reported the presence of clonally expanded T cells in the CSF of AD patients and granzyme-A$^+$ CD8$^+$ T cells in the parenchyma of AD brains (18). Cumulatively, these advances provide evidence that T cells may be involved in the pathobiology of several neurodegenerative diseases and emphasize the value of CSF cells to study the role of adaptive immunity in neurodegenerative disease.

Most of the aforementioned studies have relied on freshly isolated CSF cells (1, 5, 6, 9-20) or extraction of genomic DNA or RNA from frozen cells (7, 8). Analyzing fresh CSF cells provides the highest number of viable cells and the closest approximation of their endogenous physiology, but introduces batch effects which are especially important to avoid in high-throughput sequencing experiments. Conversely, cryopreservation allows for parallel analysis of multiple samples acquired longitudinally, minimizing batch effects and reducing total time spent. Increased evidence of CSF cell involvement in neurodegenerative disease warrants a standardized approach for the long-term preservation of CSF immune cells and methods to assess the role of CSF immune cells in pathobiology. Here, we report a method for the long-term storage and subsequent analysis of CSF cells by scRNA-TCRseq. Moreover, we present unbiased bioinformatics approaches to identify and visualize disease-associated TCRs. These methods will enable researchers to discover novel disease mechanisms in neurodegenerative diseases, particularly in the context of adaptive immunity.

Results

We developed a standardized workflow for the isolation and cryopreservation of CSF immune cells for scRNA-TCRseq (Figure 1a). Following extraction of CSF by lumbar puncture, cells were pelleted by centrifugation, checked for blood contamination (Supplementary Figure 1a), and cryopreserved. Samples without blood contamination were thawed at 37 °C and live cells were sorted by fluorescence activated cell sorting (FACS; Figure 1b), resulting in an average of 4,961 live cells per sample (Figure 1c) and 71% viability of singlets (Figure 1d). Length of storage had no significant impact on number of cells acquired after thaw (Figure 1e).

Here, we sequenced CSF cells from a total of 24 subjects – 8 healthy controls, 4 patients with clinical AD, 5 with mild cognitive impairment (MCI), and 7 with PD (Table 2). After thawing and sorting live cells, we prepared single cell libraries then amplified the global transcriptome and TCRab genes from each sample. This resulted in an average of 1,323 sequenced cells per sample (Figure 1f). After quality control (Supplementary Figure 2a) and batch correction, dimensionality reduction and clustering revealed separation of CD8$^+$ and CD4$^+$ T cells, NK cells, monocytes, dendritic cells (DCs), B cells, and plasma cells (Figure 2a), based on marker gene expression analysis (Figure 2b). Cells with sequenced TCRs (Figure 2c) overlapped with cells that expressed the pan-T cell marker CD3D (Figure 2d), confirming T cell identity and productive TCR sequencing. Importantly, we did not detect platelet genes, such as *PPBP*, among our CSF cells, confirming that our samples were not contaminated with blood. Additionally, clusters were not enriched for particular subjects, diagnosis, processing batch or sequencing batch (Supplementary Figure

2b-e). After clustering and annotating cells, we calculated the average proportion of each cell type in our CSF samples (Figure 2e). $CD4^+$ T cells were the most abundant cell type (64%) followed by $CD8^+$ T Cells (30%), myeloid cells (3.5%), NK cells (2%), plasma cells (0.5%) and B cells (0.3%) (Figure 2e). There was no significant difference in the cell type composition between disease groups (Supplementary Figure 2f). These results indicate that our method of long-term preservation of CSF cells retains several immune cell populations and is suitable for downstream scRNA-TCRseq analysis.

We next sought to develop methods to facilitate the discovery of TCR-antigen specificities relevant to neurodegenerative diseases. We first assessed whether clonally expanded T cells were present in the CSF, as determined by identification of cells with identical TCR sequences. Indeed, we identified numerous clonally expanded T cells (Figure 2f) and observed that larger clones (of size > 5) were enriched in $CD8^+$ T cells compared to $CD4^+$ T cells (Figure 2g-i). These larger $CD8^+$ T cell clones showed significant increased expression of genes related to cytotoxicity (*GZMA*, *GZMH*, *GZMM*, *NKG7*), cytokine signaling (*CCL4*, *CCL5*) and antigen presentation (*HLA-B*, *HLA-C*, *HLA-DP*) as well as decreased expression of genes associated with naïve T cells (*SELL*) and $CD4^+$ T cells (*IL7R*) (Figure 2g). Interestingly the expression level of these genes appeared to follow a gradient based on clone size, showing that our method of CSF preservation retains subtle gene expression patterns within cell types.

Given the abundance of clonally expanded T cells in the CSF, we reasoned that examining TCR similarity within and between disease groups could provide insight into disease-relevant TCR-antigen specificities. We first used a Levenshtein similarity (L-sim) score to quantify TCR similarity. This score is based on the Levenshtein distance algorithm, which calculates the number of edits, deletions, or insertions required to make two strings identical. L-sim includes a string length normalization and transformation of the final value to be between 0 and 1, with 1 representing identical TCR sequences. To examine TCR similarity within and between disease groups, we calculated the L-sim score for each possible pair of TCRs in our dataset, including healthy controls. To increase the confidence of this analysis and to focus on disease-relevant TCRs, we filtered our dataset to only include complete, unambiguous TCRab sequences from clonally expanded T cells (Figure 3a). Intriguingly, L-sim scores followed a normal distribution with very few highly similar TCRs in the dataset (Supplementary Figure 3a). By selecting TCRs with L-sim > 0.8, we identified several clusters of similar TCRab sequences (Figure 3b). Interestingly, the largest clusters contained TCR sequences exclusively from patients with neurodegenerative diseases, even though healthy control TCRs were also included in the analysis (Figure 3c), and were expressed by $CD8^+$ T cells (Figure 3d). Moreover, three of the TCRb sequences in Cluster 1 were identical (Supplementary Figure 3b) (CASSLGQAYEQYF) and have been shown to be specific for Epstein-Barr virus (EBV) epitope EBNA3A (18).

To more broadly understand TCR similarity within and between disease groups, we developed a TCR node network visualization that displays connections between similar TCRs (L-sim > 0.8) organized by unique samples and disease groups. This analysis revealed that AD and MCI patients have numerous similar TCR sequences, though other disease groups and healthy controls also share some similar TCR

sequences (Figure 3e). Visualizing TCR similarity using a stricter L-sim cutoff of 0.9 showed that highly similar TCRs in our dataset were almost exclusively shared within AD/MCI patients, and no similar TCRs were shared with healthy controls (Figure 3f). This suggests that AD and MCI patients may have similar T cell antigen specificities in the CSF, perhaps indicative of a disease-specific adaptive immune response. To validate our node network method, we utilized a ground-truth dataset to identify candidate TCRs involved in a separate neurodegenerative disease. We thus generated TCR networks from publicly available CSF TCRb sequences from narcolepsy patients. This dataset contains experimentally determined TCRb sequences that target hypocretin (HCRT) and Tribbles Pseudokinase 2 (TRIB2) antigens (15). We identified several identical TCRb sequences shared between unique narcolepsy CSF $CD4^+$ T cell samples. Interestingly, we also found a pair of identical TCRb sequences (one being derived from CSF and another from blood) that was previously determined to be HCRT-specific (15) (Supplementary Figure 4a). Moreover, we identified a candidate HCRT-specific TCRb sequence shared between three unique narcolepsy CSF $CD8^+$ T cell clones (Supplementary Figure 4b). Interestingly, TCRs from narcolepsy did not show overlap with TCRs from healthy controls, suggesting that these TCRs are relevant and specific to narcolepsy. Cumulatively, using an independent TCR dataset, we validate L-sim and TCR network visualization as tools to reveal disease-relevant TCRs.

In addition to L-sim, we show that motif analysis can be used to identify TCR sequences that share antigen specificity within and between disease groups. Using a sliding window approach, we quantified the frequency of every motif of every size from our pool of TCRb sequences (Figure 3g). We then inspected antigen specificity of large, frequent motifs by searching the motifs in the McPAS-TCR specificity public database (21). This search uncovered motifs which were highly correlated with specific antigens, including the SSLGQAYEQ motif which was found only in patients with AD or MCI (Figure 3h). This motif has been reported to be expressed primarily by $CD8^+$ T cells with specificity against EBV (Figure 3i). In summary, we show that TCR motif analysis can be used to identify disease-associated TCRs as well as their established antigen targets.

Discussion

Most studies performed on human cells rely on peripheral blood mononuclear cells. However, utilizing peripheral cells as a read-out of immune changes in neurodegeneration limits our ability to understand central immunity. CSF cells, on the other hand, provide a way to directly study immunology in the central nervous system. Indeed, recent studies of intrathecal immunity by us and others have shown this understudied immune compartment to be relevant to the pathobiology of neurodegenerative diseases (1, 5-20, 22). Yet, CSF cells have been widely understudied because of the invasive method of CSF extraction (typically lumbar puncture), and because cells are often discarded for proteomic analysis. Cells not discarded are typically analyzed fresh, given the delicate nature of CSF cells (23). However, independent processing of samples can lead to major batch effects, which are especially of concern with high-throughput sequencing experiments. Therefore, to allow for multiple samples to be processed simultaneously, we developed a method for long-term preservation of CSF cells.

Our simple and quick method allows for the preservation of an average 4,961 live cells after thawing, with no significant effect of storage length on the number of cells acquired post thaw. We find 64% of intrathecal cells are $CD4^+$ T cells and 30% are $CD8^+$ T cells, compared to 60-83% and 11-20%, respectively, reported by others (1, 20). We find that monocytes, dendritic cells, and NK cells combined make up 6% of CSF cells, whereas other studies report that monocytes make up 5-12%, dendritic cells less than 4%, and NK cells around 5% of fresh CSF cells (1). We also detected small proportions of plasma cells (0.3%) and B cells (0.5%), which aligns with the less than 1% of B cells previously reported in fresh CSF (22). The differences in CSF composition between this study and previous studies may be explained by the increased susceptibility to cell loss of myeloid cells compared to lymphocytes from freshly isolated CSF samples, (23) combined with the added stress of the freeze and thaw process. Overall, while we find a slight reduction in the number of recovered myeloid cells and NK cells, our method preserves T cells, the most abundant cell type in fresh CSF. Preservation of these cells enables the molecular study of adaptive immunity in neurodegenerative disease by scRNA-TCRseq while minimizing batch effects and time restrictions.

Beyond introducing a method to store human CSF cells, we show that scRNA-TCRseq analysis can provide insight into the TCR repertoire and antigen specificities of clonally expanded T cells within and between neurodegenerative disease groups. We provide an easy-to-use and readily modifiable R script that can act as a base template for future studies. By quantifying the similarity of each TCR to every other TCR using the L-sim algorithm and displaying similar TCRs in a patient node network, we show that TCRs from patients with neurodegenerative diseases, namely AD and MCI, cluster together. We also validate L-sim and TCR network visualization using an experimentally determined HCRT-specific TCRb sequence in narcolepsy CSF. This approach also identified several candidate HCRT-specific $CD4^+$ and $CD8^+$ T cell TCRb sequences. Additionally, we show that motif analysis can be used to identify TCRs that may share antigen specificity. Both similarity analyses independently revealed clusters of TCRs in patients with neurodegenerative diseases that were specific to EBV. However, these clusters represent only a small fraction of the TCR repertoire in the CSF. Moreover, our current knowledge of TCR antigen-specificity is largely limited to studies of infectious disease, biasing results to viral epitopes.

It will be important to continue to expand our knowledge of TCR antigen specificity in neurodegenerative diseases, especially in diseases where infection is not thought to be the primary cause. This is especially important given the growing evidence that T cells may negatively contribute to the pathobiology of several neurodegenerative diseases, including AD and PD (17, 18). Utilizing these methods to identify candidate disease-associated TCRs will aid in the identification of self/non-self antigen targets of clonally expanded T cells in neurodegenerative diseases.

Conclusions

In conclusion, we provide a detailed method for long-term preservation of CSF cells and subsequent analysis by scRNA-TCRseq to study clonally expanded T cells in neurodegenerative disease. Importantly, we uncover several disease-associated TCR clusters of unknown specificity. Although these TCRs are

highly similar, additional studies will need to confirm their shared specificity for antigens. Future studies utilizing high throughput peptide screens, such as yeast surface display platforms, could be employed to identify novel antigens. Theoretically, these novel antigens could serve as novel therapeutic targets or biomarkers for neurodegenerative diseases. With increasing innovation and throughput in technologies that aid in the identification of TCR-antigen specificities, the mapping of clonally expanded TCRs will improve our understanding of the role of T cells in the pathobiology of neurodegenerative diseases. Ultimately, we anticipate that these methods will aid in the development of novel adaptive immune-focused treatments for neurodegenerative disease.

Methods

Study participants

Samples were acquired through the Stanford Brain Rejuvenation Program, the NIA funded Stanford Alzheimer's Disease Research Center (ADRC), the University of California at San Francisco ADRC and the University of California at San Diego ADRC. Collection of CSF was approved by the Institutional Review Board of each university; written consent was obtained from all subjects. A total of 34 living subjects were used in this study, 24 of which were used for scRNA-TCRseq. The 24 subjects included 8 healthy controls, 4 patients with AD, 5 with MCI, and 7 with PD.

Cryopreservation of CSF cells

CSF was collected by lumbar puncture, then centrifuged at 300 *rcf* for 10 minutes at 4 °C to pellet immune cells. Importantly, CSF samples were checked for blood contamination by examining the pellet for the presence of red blood cells by eye. An example of a CSF sample contaminated with blood is shown in Supplemental Figure 1a. Note that cells should remain at 4 °C until they are further processed, but it is best to freeze the cells as quickly as possible to limit cell death. The supernatant (cell free CSF) was aliquoted, carefully leaving behind 100 µl of CSF with the pelleted cells. 100 µl of CSF was left so that cells were concentrated enough for counting and viability measurements. The pelleted cells were then gently resuspended in the 100 µl CSF and 10 µl of resuspended cells were then removed for counting. Importantly, cells were gently resuspended by first tapping the bottom of the tube and then gently triturating 10 times, making sure not to touch the pipette tip to the edge of the tube. Then 10 µl CSF was removed and mixed with 10 µl trypan blue to assess red blood cell content and viability. Cells were then visualized on a TC20 automated cell counter (BioRad) and cell number, viability and the presence or absence of red blood cells was recorded. CSF samples contaminated with blood were discarded. The resuspended cells were then mixed with 900 µl Recovery Cell Culture Freezing Medium (Thermo Fisher Scientific). This medium is an optimized version of the typical freezing medium, containing high-glucose Dulbecco's Modified Eagle Medium with 10% serum and 10% dimethyl sulfoxide. We utilized this medium because it is quality tested for pH, osmolality, sterility, and endotoxin and each lot is quality tested on CHO-K1 cells. The freezing medium was first thawed at 37 °C, aliquoted, and stored at -20 °C. Before use, the medium was thawed at 37 °C and kept on ice. After each aliquot was thawed, the freezing medium

was stored at 4 °C for up to one month. All samples were frozen overnight at −80 °C in a Mr. Frosty freezing container (Thermo Fisher Scientific) and transferred the following day to liquid nitrogen for storage. CSF cells were stored in liquid nitrogen on average 266 days.

Preparation of frozen CSF cells for analysis

CSF cells were thawed at 37 °C in a water bath with the media submerged and the top of the tube out of the water. Cells were kept in the water bath for as little as possible and removed when the media was nearly completely thawed. The cells were then removed and gently pipetted into a 5 mL flow cytometry tube containing 3 mL pre-warmed (37 °C) sorting buffer (PBS with 0.04% bovine serum albumin (BSA)). The tube was then rinsed once with the sorting buffer and placed into the same flow cytometry tube. Cells were then centrifuged at 350 *rcf* for ten minutes. The supernatant was removed, and cells were resuspended in 500 µl sorting buffer. ½ µl of Sytox Red (Thermo Fisher Scientific) was added to the sample immediately before sorting by flow cytometry. Live cells were sorted into 1.5 mL Eppendorf tubes containing 750 uL sorting buffer. Once all the samples were sorted, cells were spun at 350 *rcf* at 4 °C in a spinning bucket centrifuge for 7 minutes. The supernatant was then removed, leaving behind 10 µl. 10 µl was left behind to resuspend the CSF cells and load the entire volume for droplet scRNA-seq.

Drop-seq of CSF cells

Chromium Single Cell 5’ Library & Gel Bead kit, Chromium Single Cell 5’ Library construction kit, Chromium Single Cell A Chip Kit, and Chromium i7 Multiplex kit (10X Genomics) were used for scRNAseq of CSF cells. We followed 10x Genomic’s User Guide for library construction. The only change we made to their protocol was in Step 1, GEM Generation & Barcoding. The user guide recommends loading a certain volume of cell suspension stock depending on the concentration of the cell suspension and the user’s desired cell recovery. However, because CSF contains such low cell numbers, we loaded all the cells that were resuspended in 10 µl of sorting buffer. We then added the 10 µl of cell suspension and 21.7 µl of nuclease free water, which results in the same total volume of cell suspension/water that the user guide recommends. After library construction, libraries were sequenced by Novogene on a Novoseq S4 sequencer and FASTQ files were generated by Novogene. Cell Ranger v.3.0.2 was used to generate gene-expression matrices for CSF cells. Reads from the 10X v.2 5' paired library were mapped to the human genome build GRCh38 3.0.0. The 5' gene-expression libraries were then analyzed with the Cell Ranger count pipeline and the resulting expression matrix was used for further analysis in the Seurat package v.3.0.

Seurat clustering of CSF cells

Individual sample expression matrices were loaded into R using the function Read10x under the Matrix package v.1.2-15. The expression matrix for each sample was merged into one Seurat object using the CreateSeuratObject and MergeSeurat functions. The Seurat package v.3.0 was used for filtering, variable gene selection, normalization, scaling, dimensionality reduction, clustering and visualization. Genes were excluded if they were expressed in fewer than 10 cells and cells were excluded if they expressed fewer

than 200 genes. Cells that expressed more than 1,600 genes, more than 6,000 UMIs and more than 10% mitochondrial genes were excluded from the analysis. The sctransform normalization method was used to normalize, scale, select variable genes and regress out sequencing and experimental batch, mitochondrial mapping percentage, number of UMIs, and number of genes. After filtering and normalization, there were 26,797 cells and 14,953 genes. Following PCA, 5 principle components were selected for clustering tSNE dimensionality reduction.

Narcolepsy patient TCR sequences

All Narcolepsy patient TCR sequences were acquired from the Latorre D, et al 2018 study (15).

Calculation of Levenshtein similarities

L-sim scores were calculated using the levenshteinSim function in the RecordLinkage package for R (24). L-sim calculation incorporates the Levenshtein distance algorithm, which quantifies the number of edits, deletions, or insertions required to make two strings identical. L-sim includes a string length normalization and transformation of the final value to be between 0 and 1, with 1 representing identical TCR sequences:

Levenshtein Similarity

$$1 - \frac{\mathrm{lev}_{a,b}(i,j)}{\max(A,B)}$$

Levenshtein Distance

$$\mathrm{lev}_{a,b}(i,j) = \begin{cases} \max(\mathrm{i,j}) & \text{if } \min(i,j) = 0, \\ \min \begin{cases} \mathrm{lev}_{a,b}(i-1,j)+1) \\ \mathrm{lev}_{a,b}(i,j-1)+1) \\ \mathrm{lev}_{a,b}(i-1,j-1)+1_{a_i \neq b_j}) \end{cases} & \text{otherwise.} \end{cases}$$

TCR network visualization

TCR networks that show connections between similar TCRs organized by patient IDs and diagnosis groups were generated using the qgraph function in the qgraph package for R (25). Analysis from Figure 3b-d include only clonal TCRs with unambiguous alpha and beta chains. Supplementary Figures 4 and 5 include only clonal TCRs with unambiguous beta chains.

TCR motif analysis

A custom script was used to identify motifs in our TCR beta chains. Only clonal TCRs with unambiguous beta chains were included for analysis. Identified motifs were searched in the McPAS-TCR database, a manually curated database of TCR sequences found to be associated with pathological conditions in mice and humans (21).

List Of Abbreviations

T cell receptor = TCR; CSF = cerebrospinal fluid; scRNA-TCRseq = single cell RNA-TCR sequencing; Alzheimer's disease = AD; Mild cognitive impairment = MCI; Parkinson's disease = PD; Healthy control = HC; Natural killer = NK; Dendritic cells = DCs; Levenshtein similarity = L-sim; Epstein-Barr virus = EBV

Declarations

Ethics approval and consent to participate

Collection of CSF was approved by the Institutional Review Board of each university, and written consent was obtained from all subjects

Consent for publication

Not applicable

Availability of data and materials

The scRNAseq datasets analyzed during the current study are available in the Gene Expression Omnibus (GEO) repository under accession number GSE134578. The fresh cell counts and flow cytometry data generated in the current study are available from the corresponding author upon request. The R scripts for TCR similarity analysis are available on github:
https://github.com/hamiltonoh/TCR_similarity_analysis.git

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by a National Institute of Neurologic Disease and Stroke K99/R00 Pathway to Independence Award (NS112458-01A1) (D.G.), an Irene Diamond Fund/AFAR Postdoctoral Transition Award in Aging (D.G.), a National Institutes of Health National Institute on Aging (NIA) F32 Fellowship (AG055255-01A1) (D.G.), the NOMIS Foundation (T.W-C.), the Stanford Brain Rejuvenation Project (an initiative of the Stanford Neurosciences Institute), and the NIA funded Stanford Alzheimer's Disease Research Center P50AG047366 (V.W.H.).

Author's contributions

D.G. and O.L. designed experiments. H.O., O.L. and D.G. prepared the manuscript. D.G. conducted flow cytometry analysis and sorting. H.O. and O.L. analyzed scRNAseq data. B.L. designed and performed bioinformatics analysis. All authors edited, read and approved the final manuscript.

Acknowledgements

We thank the entire Stanford Alzheimer's Disease Research Center team as well as G. Kerchner and S. Sha for CSF collection.

References

1. Han S, Lin YC, Wu T, Salgado AD, Mexhitaj I, Wuest SC, et al. Comprehensive immunophenotyping of cerebrospinal fluid cells in patients with neuroimmunological diseases. J Immunol. 2014;192(6):2551-63.
2. Perrin RJ, Fagan AM, Holtzman DM. Multimodal techniques for diagnosis and prognosis of Alzheimer's disease. Nature. 2009;461(7266):916-22.
3. Blennow K, Zetterberg H. The Past and the Future of Alzheimer's Disease Fluid Biomarkers. J Alzheimers Dis. 2018;62(3):1125-40.
4. Ransohoff RM, Engelhardt B. The anatomical and cellular basis of immune surveillance in the central nervous system. Nat Rev Immunol. 2012;12(9):623-35.
5. Schafflick D, Xu CA, Hartlehnert M, Cole M, Schulte-Mecklenbeck A, Lautwein T, et al. Integrated single cell analysis of blood and cerebrospinal fluid leukocytes in multiple sclerosis. Nat Commun. 2020;11(1):247.
6. Aly L, Yousef S, Schippling S, Jelcic I, Breiden P, Matschke J, et al. Central role of JC virus-specific CD4+ lymphocytes in progressive multi-focal leucoencephalopathy-immune reconstitution inflammatory syndrome. Brain. 2011;134(Pt 9):2687-702.
7. Lossius A, Johansen JN, Vartdal F, Robins H, Jurate Saltyte B, Holmoy T, et al. High-throughput sequencing of TCR repertoires in multiple sclerosis reveals intrathecal enrichment of EBV-reactive CD8+ T cells. Eur J Immunol. 2014;44(11):3439-52.
8. Salou M, Garcia A, Michel L, Gainche-Salmon A, Loussouarn D, Nicol B, et al. Expanded CD8 T-cell sharing between periphery and CNS in multiple sclerosis. Ann Clin Transl Neurol. 2015;2(6):609-22.
9. Johansen JN, Vartdal F, Desmarais C, Tutturen AE, de Souza GA, Lossius A, et al. Intrathecal BCR transcriptome in multiple sclerosis versus other neuroinflammation: Equally diverse and compartmentalized, but more mutated, biased and overlapping with the proteome. Clin Immunol. 2015;160(2):211-25.
10. Planas R, Metz I, Ortiz Y, Vilarrasa N, Jelcic I, Salinas-Riester G, et al. Central role of Th2/Tc2 lymphocytes in pattern II multiple sclerosis lesions. Ann Clin Transl Neurol. 2015;2(9):875-93.
11. Hoglund RA, Lossius A, Johansen JN, Homan J, Benth JS, Robins H, et al. In Silico Prediction Analysis of Idiotope-Driven T-B Cell Collaboration in Multiple Sclerosis. Front Immunol. 2017;8:1255.
12. Rathbone E, Durant L, Kinsella J, Parker AR, Hassan-Smith G, Douglas MR, et al. Cerebrospinal fluid immunoglobulin light chain ratios predict disease progression in multiple sclerosis. J Neurol Neurosurg Psychiatry. 2018;89(10):1044-9.
13. Beltran E, Gerdes LA, Hansen J, Flierl-Hecht A, Krebs S, Blum H, et al. Early adaptive immune activation detected in monozygotic twins with prodromal multiple sclerosis. J Clin Invest. 2019;129(11):4758-68.
14. Tomescu-Baciu A, Johansen JN, Holmoy T, Greiff V, Stensland M, de Souza GA, et al. Persistence of intrathecal oligoclonal B cells and IgG in multiple sclerosis. J Neuroimmunol. 2019;333:576966.
15. Latorre D, Kallweit U, Armentani E, Foglierini M, Mele F, Cassotta A, et al. T cells in patients with narcolepsy target self-antigens of hypocretin neurons. Nature. 2018;562(7725):63-8.

16. Schneider-Hohendorf T, Mohan H, Bien CG, Breuer J, Becker A, Gorlich D, et al. CD8(+) T-cell pathogenicity in Rasmussen encephalitis elucidated by large-scale T-cell receptor sequencing. Nat Commun. 2016;7:11153.
17. Sulzer D, Alcalay RN, Garretti F, Cote L, Kanter E, Agin-Liebes J, et al. T cells from patients with Parkinson's disease recognize α-synuclein peptides. Nature. 2017;546(7660):656-61.
18. Gate D, Saligrama N, Leventhal O, Yang AC, Unger MS, Middeldorp J, et al. Clonally expanded CD8 T cells patrol the cerebrospinal fluid in Alzheimer's disease. Nature. 2020;577(7790):399-404.
19. Albert ML, Austin LM, Darnell RB. Detection and treatment of activated T cells in the cerebrospinal fluid of patients with paraneoplastic cerebellar degeneration. Ann Neurol. 2000;47(1):9-17.
20. Hummert MW, Alvermann S, Gingele S, Gross CC, Wiendl H, Mirenska A, et al. Immunophenotyping of cerebrospinal fluid cells by Chipcytometry. J Neuroinflammation. 2018;15(1):160.
21. Tickotsky N, Sagiv T, Prilusky J, Shifrut E, Friedman N. McPAS-TCR: a manually curated catalogue of pathology-associated T cell receptor sequences. Bioinformatics. 2017;33(18):2924-9.
22. Kivisakk P, Mahad DJ, Callahan MK, Trebst C, Tucky B, Wei T, et al. Human cerebrospinal fluid central memory CD4+ T cells: evidence for trafficking through choroid plexus and meninges via P-selectin. Proc Natl Acad Sci U S A. 2003;100(14):8389-94.
23. de Graaf MT, de Jongste AH, Kraan J, Boonstra JG, Sillevis Smitt PA, Gratama JW. Flow cytometric characterization of cerebrospinal fluid cells. Cytometry B Clin Cytom. 2011;80(5):271-81.
24. Sariyar M, Borg A. The RecordLinkage Package: Detecting Errors in Data. The R Journal. 2010;2(2).
25. Epskamp S, Cramer AOJ, Waldorp LJ, Schmittmann VD, Borsboom D. qgraph: Network Visualizations of Relationships in Psychometric Data. Journal of Statistical Software; Vol 1, Issue 4 (2012). 2012.

Tables

Table 1. Studies focused on CSF cells.

An exhaustive search was performed on pubmed.gov to include all studies focused on CSF cell composition in disease.

Author	Year	Journal	Title	Cell preservation method	Method of analysis	Disorder studied
Lueg et al.	2015	Neurobiology of Aging	Clinical relevance of specific T-cell activation in the blood and cerebrospinal fluid of patients with mild Alzheimer's disease	CSF cells processed within 20 minutes	flow cytometry	AD & MCI
Tomescu-Baciu et al.	2019	Journal of Neuroimmunology	Persistence of intrathecal oligoclonal B cells and IgG in multiple sclerosis	CSF cells processed immediately	quantitative mass spectrometry of oligoclonal bands and immunoglobulin heavy-chain from cDNA sequencing	B cells and IgG in MS patients at different time points
Giunti et al.	2003	Journal of Leukocyte Biology	Phenotypic and functional analysis of T cells homing into the CSF of subjects with inflammatory diseases of the CNS	unclear	flow cytometry & culture of T cell clones from CSF (maintained by PGA stimulation)	inflammatory neurological diseases (including MS)
Hauser et al.	1983	Neurology	CSF cells in multiple sclerosis Monoclonal antibody analysis and relationship to peripheral blood T-cell subsets	fresh CSF cells processed within 4 hours of lumbar puncture	fluorescence microscopy	MS
Calebresi et al.	1998	Journal of Neuroimmunology	Cytokine gene expression in cells derived from CSF of multiple sclerosis patients	CSF cells processed immediately	RNA extraction + RT-PCR	MS
Aly et al.	2011	Brain: A Journal of Neurology	Central role of JC virus-specific CD4+ lymphocytes in progressive multi-focal leucoencephalopathy-immune reconstitution inflammatory syndrome	PHA expanded CSF cells	proliferation assay (3H-thymidine incorporation) of cells stimulated with peptides	MS
Lossius et al.	2014	European Journal of Immunology	High-throughput sequencing of TCR repertoires in multiple sclerosis reveals intrathecal enrichment of EBV-reactive CD8+ T	pellets snap frozen in liquid nitrogen & genomic DNA extracted	TCRβ sequencing	MS
Salou et al.	2015	Annals of Clinical and Translational Neurology	Expanded CD8 T-cell sharing between periphery and CNS in multiple sclerosis	CSF cells stored frozen in Trizol & total RNA extracted(CSF obatined from cisterna magna)	TCRβ sequencing	MS
Johansen et al.	2015	Clinical Immunology	Intrathecal BCR transcriptome in multiple sclerosis versus other neuroinflammation: Equally diverse and compartmentalized, but more mutated, biased and overlapping with the proteome	CSF cells processed within 20 minutes	sequencing of IGHV and mass spectrometry	MS
Planas et al.	2015	Annals of Clinical and Translational Neurology	Central role of Th2/Tc2 lymphocytes in pattern II multiple sclerosis lesions	CSF cells expanded in culture for 2 weeks	CSF cells expanded in vitro and then FACs sorted based on Vβ	MS
Hoglund et al.	2017	Frontiers in Immunology	In Silico Prediction Analysis of Idiotope-Driven T–B Cell Collaboration in Multiple Sclerosis	CSF cells processed immediately	sequencing IGH-VDJ region with Adaptive	MS
Rathbone et al.	2018	Journal Neurol Neurosurg Psychiatry	Cerebrospinal fluid immunoglobulin light chain ratios predict disease progression in multiple sclerosis	CSF cells processed immediately	flow cytometry & sequencing of IGKV and IGLV with adaptive	MS
Schafflick et al.	2020	Nature communications	Integrated single cell analysis of blood and cerebrospinal fluid leukocytes in multiple sclerosis	processed and kept at 4 °C	single-cell RNA sequencing (10x)	MS
Hummert et al.	2018	Journal of Neuroinflammation	Immunophenotyping of cerebrospinal fluid cells by Chipcytometry	fresh CSF cells fixed with 4% phosphate-buffered formaldehyde	Chipcytometry	MS/CIS, OIND (infectious & autoimmune), tumor, NIND, NND, unspecified
Beltran et al.	2019	The Journal of Clinical Investigation	Early adaptive immune activation detected in monozygotic twins with prodromal multiple sclerosis	Fresh CSF and blood samples processed within 1 hour after collection	single-cell RNA sequencing (Smart-Seq2)	MS-discordant monozygotic twin pairs
Latorre et al.	2018	Nature	T cells in patients with narcolepsy target self-antigens of hypocretin neurons	CSF cells expanded in culture within a few hours of sampling (15 days in culture)	polyclonal expansion of CSF cells followed by TCR Vβ sequencing	Narcolepsy
Kivisakk et al.	2003	PNAS	Human cerebrospinal fluid central memory CD4+ T cells: Evidence for trafficking through choroid plexus and meninges via P-selectin	Fresh CSF cells processed within 20 minutes of sampling	flow cytometry	Noninflammatory neurological diseases
Albert et al.	2001	Annals of Neurology	Detection and treatment of activated T cells in the cerebrospinal fluid of patients with paraneoplastic cerebellar degeneration	unclear	flow cytometry	paraneoplastic cerebellar degeneration
Schneider-Hohendorf et al.	2016	nature communications	CD8+ T-cell pathogenicity in Rasmussen encephalitis elucidated by large-scale T-cell receptor sequencing	unclear	TCR sequencing with Adaptive	Rasmussen encephalitis
Han et al.	2014	Journal of Immunology	Comprehensive Immunophenotyping of Cerebrospinal Fluid Cells in Patients with Neuroimmunological Diseases	CSF cells processed immediately	flow cytometry	neuroimmunological disorders
Lanz et al.	2019	frontiers in Immunology	Single-Cell High-Throughput Technologies in Cerebrospinal Fluid Research and Diagnostics	Review	NA	NA
Glantz et al.	1998	Cancer	Cerebrospinal fluid cytology in patients with cancer: minimizing false- negative results	fresh cells or kept at 4◦C for 48 hours	flow cytometry	cancer
Pittman et al.	2013	Arch Pathol Lab Med	Utility of flow cytometry of cerebrospinal fluid as a screening tool in the diagnosis of central nervous system lymphoma	unclear	flow cytometry	CNS lymphoma
Svenningsson	1995	Neuroimmunology	Lymphocyte phenotype and subset distribution in normal cerebrospinal fluid.	cells analyzed fresh and fized before flow analysis	flow cytometry	Healthy
de Graaf et al.	2011	Cytometry B Clin Cytom	Central memory CD4+ T cells dominate the normal cerebrospinal fluid.	fresh CSF cells (within 6 hours of lumbar puncture)	flow cytometry	neurological diseases
Enose-Akahata	2018	PLoS Pathog	Immunophenotypic characterization of CSF B cells in virus- associated neuroinflammatory diseases	unclear	flow cytometry	virus-associated neuroinflammatory disease
Gross et al.	2016	Proc Natl Acad Sci USA.	Impaired NK-mediated regulation of T-cell activity in multiple sclerosis is reconstituted by IL-2 receptor modulation.	Fresh CSF cells processed within 20 minutes of sampling	flow cytometry	MS

Kraan et al.	2008	Curr Protoc Cytom	Flow cytometric immunophenotyping of cerebrospinal fluid.	fresh CSF cells	flow cytometry	patients with clinical suspicion of a hematopoietic or lymphoid malignancy
Cepok et al.	2005	Brain	Short-lived plasma blasts are the main B cell effector subset during the course of multiple sclerosis	fresh CSF cells	flow cytometry	MS
Schirmer et al.	2015	Open Forum infectious disease	Extensive recruitment of plasma blasts to the cerebrospinal fluid in toscana virus encephalitis.	fresh CSF cells	flow cytometry, microscopic cell counting and May-Grünwald cell stain	virus encephalitis
Oreja-Guevara	1998	Acta Neurol Scand	Analysis of lymphocyte subpopulations in cerebrospinal fluid and peripheral blood in patients with multiple sclerosis and inflammatory diseases of the nervous system.	fresh CSF cells	flow cytometry	MS & nervous system inflammatory diseases
Kowarik et al.	2014	J Neurology	Immune cell subtyping in the cerebrospinal fluid of patients with neurological diseases.	fresh CSF cells	flow cytometry	neurological diseases
Brucklacher-Waldert	2009	Brain	Phenotypical and functional characterization of T helper 17 cells in multiple sclerosis.	freshly used and expanded in culture by PHA stimulation	TCRβ sequencing of in vitro clones & flow cytometry	MS
Huang et al	2009	Annals of Neurology	Specific central nervous system recruitment of HLA-G(+) regulatory T cells in multiple sclerosis.	freshly used and put into culture	flow cytometry	MS
Mullen et al.	2012	J Neuroimmunology	Phenotypical and functional characterization of T helper 17 cells in multiple sclerosis.	freshly used and expanded in culture by PHA stimulation	flow cytometry	inflammatory and non-inflammatory neurological diseases
Schirmer et al.	2013	JAMA Neurol	Enriched CD161high CCR6+ gammadelta T cells in the cerebrospinal fluid of patients with multiple sclerosis.	fresh CSF cells	flow cytometry	MS
Pranzatelli et al.	2018	Clin Exp Immunol	Cerebrospinal fluid gammadelta T cell frequency is age-related: a case-control study of 435 children with inflammatory and non-inflammatory neurological disorders.	fresh CSF cells	flow cytometry	children with inflammatory and non-inflammatory neurological disorders
Hamann et al.	2013	Journal of Neuroimmunology	Characterization of natural killer cells in paired CSF and blood samples during neuroinflammation.	unclear	flow cytometry	neuroinflammation
Cepok et al.	2001	Brain	Patterns of cerebrospinal fluid pathology correlate with disease progression in multiple sclerosis.	fresh CSF cells	flow cytometry	MS
Harrer et al.	2015	Clin Exp Immunol	High interindividual variability in the CD4/CD8T cell ratio and natalizumab concentration levels in the cerebrospinal fluid of patients with multiple sclerosis.	fresh CSF cells	flow cytometry	MS
Schulte-Mecklenbeck	2019	Eur J Neurol	Immunophenotyping of cerebrospinal fluid cells in ischaemic stroke	CSF cells processed within 20 minutes	flow cytometry	ischemic stroke
Ligocki	2015	ASN Neuro	A distinct class of antibodies may be an indicator of gray matter autoimmunity in early and established relapsing remitting multiple sclerosis patients.	no mention of storing cells so most likely analyzed fresh	single-cell sorting of CD19+ B cells and subsequnt VH & VL sequencing	MS
Rounds	2015	Gene	MSPrecise: a molecular diagnostic test for multiple sclerosis using next generation sequencing.	stored at -80	VH4 gene sequencing of extracted gDNA	RRMS and other neurological disease (OND) patient
Palanichamy	2014	Sci Tranls Med	Immunoglobulin class-switched B cells form an active immune axis between CNS and periphery in multiple sclerosis.	analyzed cells fresh (lysed them immediately to extract RNA)	Ig-VH sequencing from isolated RNA	MS
Stern et al.	2014	Sci Tranls Med	B cells populating the multiple sclerosis brain mature in the draining cervical lymph nodes.	stored at -80	Immunoglobulin (Ig) variable region heavy (VH) and light chain (VL) sequencing from extracted RNA	MS
Lovato et al.	2011	Brain	Related B cell clones populate the meninges and parenchyma of patients with multiple sclerosis.	stored at -80	IG variable region heavy chain sequencing from extracted RNA	MS
Obermeier et al.	2008	Nature Medicine	Matching of oligoclonal immunoglobulin transcriptomes and proteomes of cerebrospinal fluid in multiple sclerosis.	unclear, but only extractd RNA from cells	IgG-H, Ig-λ and Ig-κ sequencing from extracted RNA	MS
Cameron et al.	2009	J Neuroimmunology	Potential of a unique antibody gene signature to predict conversion to clinically definite multiple sclerosis.	analyzed cells fresh	single-cell antibody sequencing	MS
von Budingen	2012	J Clin Invest	B cell exchange across the blood-brain barrier in multiple sclerosis.	CSF cells were spun down and frozen at –80°C.	deep repertoire sequencing of IgG heavy chain variable region genes (*IgG-VH*) from extracted RNA	MS
de Paula Alves Sousa	2016	Annals of Clinical and Translational Neurology	Intrathecal T-cell clonal expansions in patients with multiple sclerosis.	CSF cells were also cryopreserved in RNA*later* solution & stored at -80 until RNA extraction	TCRB sequencing from extracted RNA	MS
Planas et al.	2018	Frontiers in Immunology	Detailed characterization of t cell receptor repertoires in multiple sclerosis brain lesions.	phytohemagglutinin-expanded CSF T cells (in vitro)	T cell receptor β-chain variable gene (TRBV) sequencing (-seq) of genomic (g)DNA with	MS
Greenfield	2019	JCI Insight	Longitudinally persistent cerebrospinal fluid B cells can resist treatment in multiple sclerosis.	stored 7 of 20 CSF samples and 1 of 20 PBMC samples at –80°C as unsorted cell pellets and the remaining were analyzed fresh	Ig heavy chain variable region repertoire sequencing on B cells (flow sorted & bulk)	MS
Irani et al.	2015	Lancet	The active intrathecal B-cell response in LGI1-antibody encephalitis.	not described	PCR amplification and next generation deep immune repertoire sequencing of immunoglobulin (Ig) heavy chain variable regions (VH)	LG11-antibody encephalitis
Kreye et al.	2016	Brain	Human cerebrospinal fluid monoclonal N-methyl-D-aspartate receptor autoantibodies are sufficient for encephalitis pathogenesis.	cells resuspended in 500 µl freezing medium [45% RPMI, 45% foetal calf serum, 10% dimethylsulphoxide (DMSO)] and stored at −80°C	flow cytometric single cell sorting -> sequencing of IGH, IHL, IGK	encephalitis
Farhadian	2018	JCI Insight	Single-cell RNA sequencing reveals microglia-like cells in cerebrospinal fluid during virologically suppressed HIV	fresh CSF and blood	scRNA-seq	HIV
Jordao et al.	2019	Science	Single-cell profiling identifies myeloid cell subsets with distinct fates during neuroinflammation.	Did not isolate CSF Cells	scRNA-seq	EAE (mouse)

Table 2. Metadata for samples used in scRNA-TCRseq

Subject characteristics for CSF samples utilized in this study.

SAMPLE	AGE	SEX	CONDITION	PROCESSING BATCH	SEQUENCING BATCH
CSF01	88	M	HC	1	1
CSF02	77	M	AD	1	1
CSF03	71	F	HC	1	1
CSF04	74	M	HC	1	1
CSF05	72	M	MCI	1	1
CSF06	64	F	AD	1	1
CSF07	57	M	PD	1	1
CSF08	74	M	HC	1	1
CSF09	83	F	AD	2	2
CSF10	66	F	PD	2	2
CSF11	66	M	MCI	2	2
CSF12	73	F	HC	2	2
CSF13	76	F	HC	2	2
CSF14	65	F	PD	2	2
CSF15	48	M	MCI	2	2
CSF16	67	F	HC	2	2
CSF17	78	M	MCI	3	2
CSF18	58	F	AD	3	2
CSF19	83	F	MCI	3	2
CSF20	66	M	HC	3	2
CSF21	63	M	PD	3	2
CSF22	77	M	PD	3	2
CSF23	69	M	HC	3	2
CSF24	62	F	PD	3	2

Supplemental Legends

Supplementary Figure 1. Identification of blood contamination in CSF

a) Representative CSF pellet with visible blood contamination. Samples contaminated with blood should be discarded for molecular analysis of CSF cells.

Supplementary Figure 2. Visualization of sample metadata.

a) Quality control metrics of each sample after removing low quality samples CSF11 and CSF22 **b)** tSNE colored by sample. **c)** tSNE colored by diagnosis. **d)** tSNE colored by processing batch. **e)** tSNE colored by sequencing batch. **f)** Quantification of average cell type distribution per diagnosis based on Seurat clustering.

Supplementary Figure 3. Levenshtein Similarity distributions and Identical TCRb Network.

a) Distributions of L-sim scores of clonal full length TCRαβ and TCRβ sequences **b)** Network displaying connections between samples with identical TCRb sequences (L-sim = 1.0). Network includes only clonal TCRs with unambiguous CDR3b sequences.

Supplementary Figure 4. Networks of narcolepsy CSF T cells showing shared TCRb sequences.

a) TCRb network displaying connections between narcolepsy patient samples with identical $CD4^+$ T cell TCRb Narcolepsy patient (NP) nodes contain clonal CSF $CD4^+$ T cell TCRs, healthy control (HC) nodes contain clonal CSF T cells from Supplementary Figure 3b, while HCRT and TRIB2 nodes contain TCRs from peripheral blood derived $CD4^+$ T cells that were experimentally determined to be specific for HCRT and TRIB2, respectively. Table below shows metadata for highlighted identical TCRs. Note that there are additional TCRs shared among narcolepsy patients, yet it is unknown whether these TCRs recognize

HCRT or another antigen. **b)** TCRb network displaying connections between narcolepsy patient CSF samples with identical $CD8^+$ T cell TCRb sequences. All nodes contain clonal CSF $CD8^+$ T cell TCRs. HC samples from Supplementary Figure 3b were used. Table below shows metadata for highlighted identical TCRs.

Figures

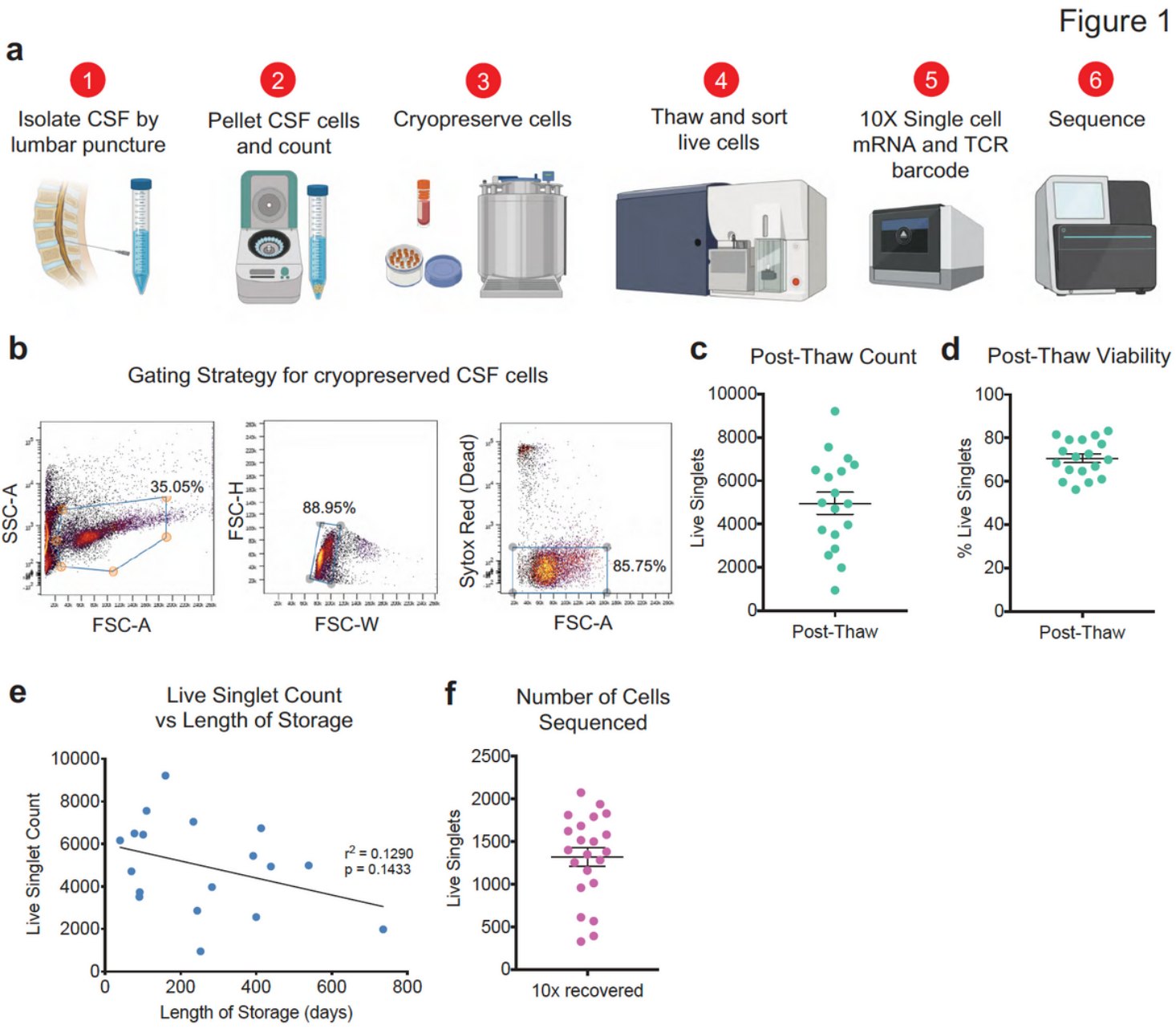

Figure 1

Cryopreservation of human cerebrospinal fluid cells retains cellular viability. a) Workflow for cryopreservation of CSF cells for scRNA-TCRseq. b) Gating strategy for sorting live cryopreserved cells by flow cytometry. c) Quantification of live cell count (live singlets) from flow cytometry sorting following

cell thawing. Mean ± s.e.m. d) Quantification of viability (percentage of live singlets out of all singlets) from sorting by flow cytometry. Mean ± s.e.m. e) Linear regression of live cell count post thaw versus length of storage. f) Quantification of the number of CSF cells captured for sequencing.

Figure 2

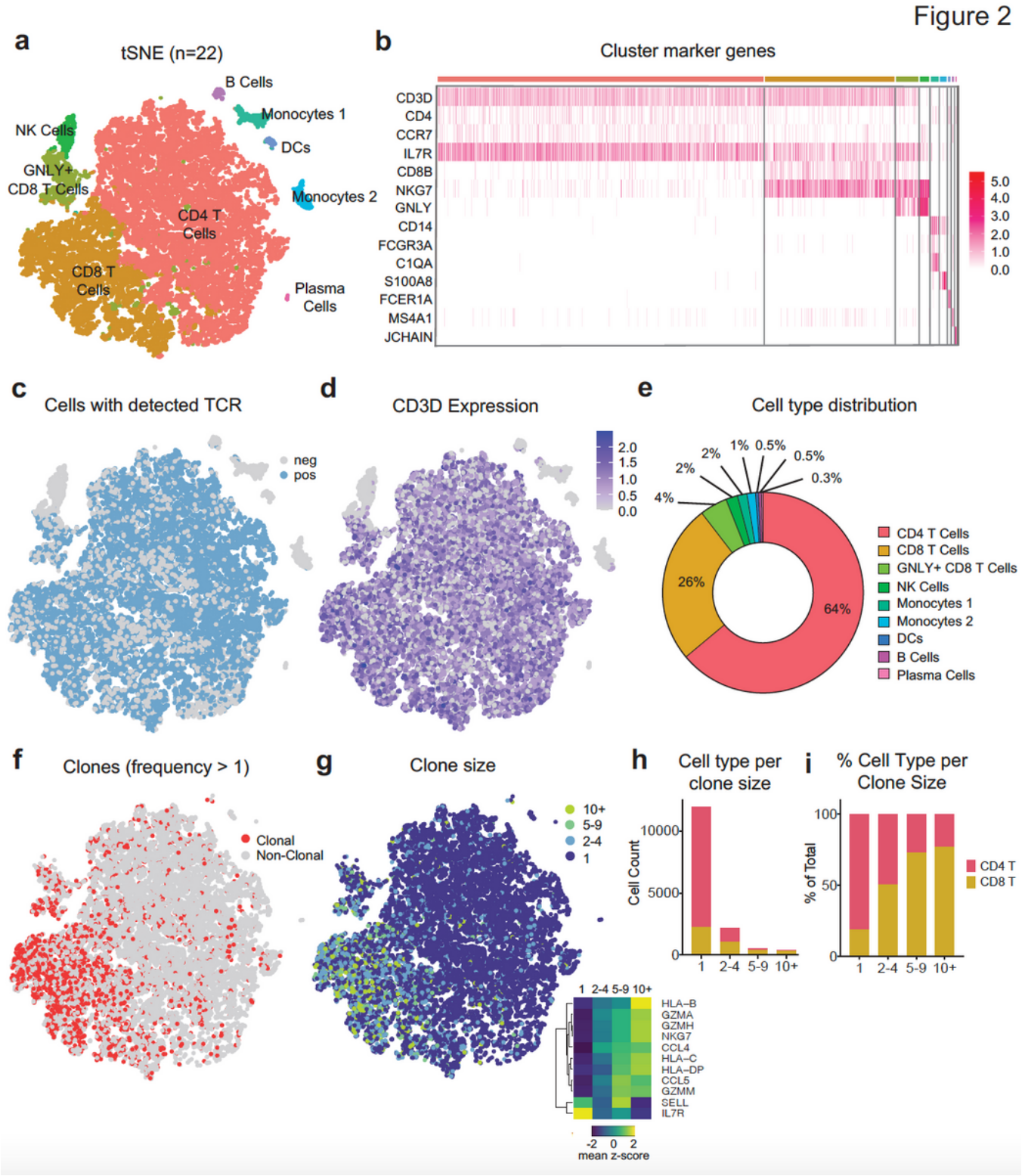

Figure 2

scRNA-TCRseq of human cerebrospinal fluid reveals clonally expanded CD8+ T cells. a) Seurat dimensionality reduction and clustering of 22 CSF samples that passed quality control displayed on tSNE

(GNLY: Granulysin; DCs: Dendritic Cells) b) Heatmap of cluster marker genes used to annotate clusters. c) Cells with detected TCR displayed on tSNE. d) CD3D expression displayed on tSNE. e) Quantification of average cell type distribution based on Seurat clustering. f) Clones – cells with TCR sequences shared with other cells – displayed on tSNE. Only cells with detected TCRs are included. g) Clones of different sizes displayed on tSNE. Only cells with detected TCRs are included; lower right heatmap shows significant differentially expressed genes between clone size bins h) Quantification of number of T cell types per clone size. Only cells with detected TCR included. GNLY+CD8+ T Cells and CD8+ T cells were combined as CD8+ T cells. i) Quantification of % T cell types per clone size.

Figure 3

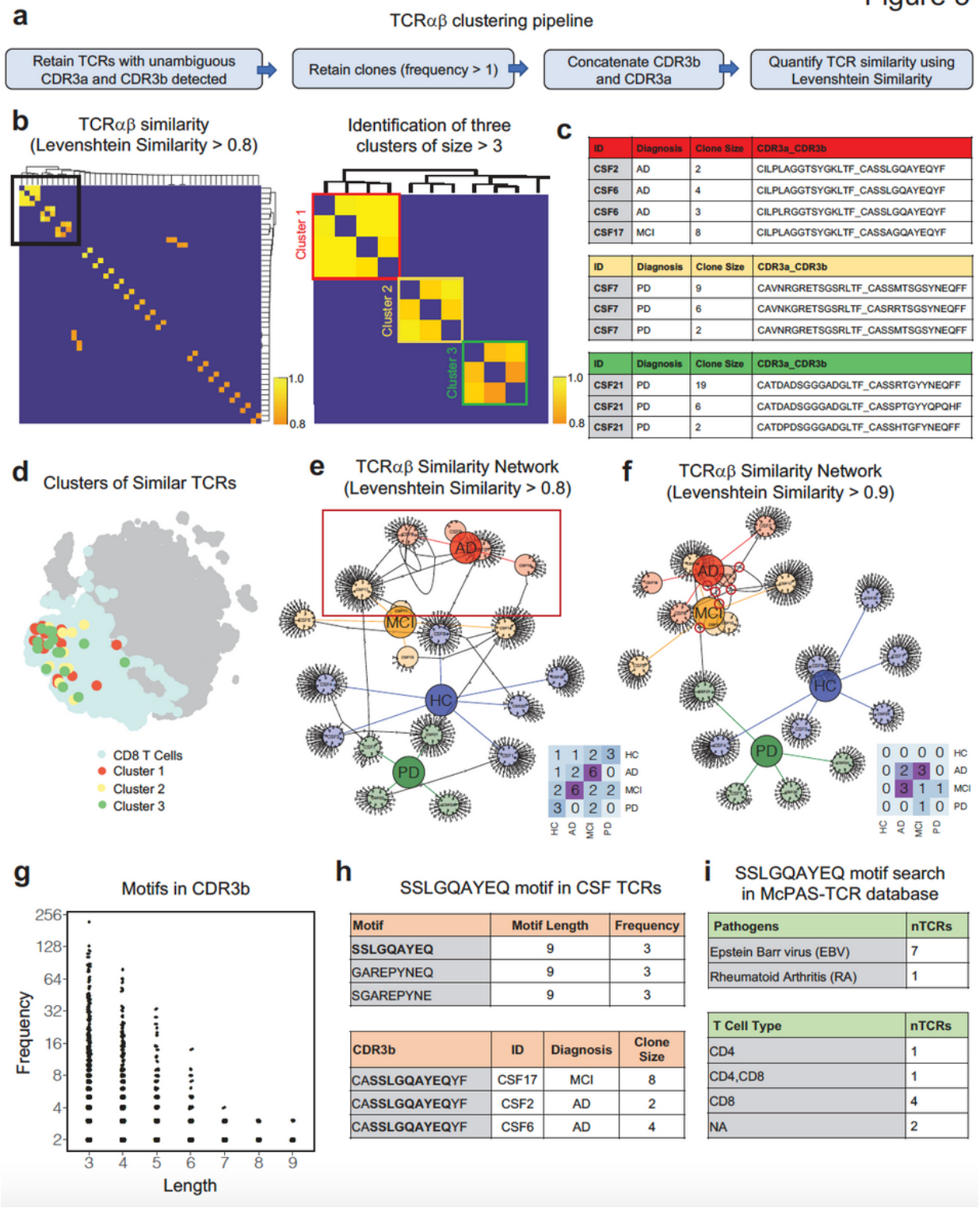

Figure 3

Analysis of T cell receptor sequence similarity within and between neurodegenerative disease groups. a) Workflow for quantifying TCRαβ similarity. Clonal TCRs with unambiguous CDR3a and CDR3b sequences were retained for analysis. L-sim values were then computed between all possible combinations of TCRαβ sequences. b) At left: heatmap highlighting TCR combinations with L-sim score > 0.8. At right: inset of heatmap to show three clusters of similar TCRs. Color bar represents L-sim score. c) Metadata of

clustered TCRs shown in b). Note the similarity of TCRs within disease groups. d) Three clusters of similar TCRs identified in c) highlighted on tSNE. Clusters of TCRs overlap with CD8+ T cells. e) TCR network displaying connections between samples with similar sequences (L-sim > 0.8) identified in b). Each node is a unique patient sample with each small circle sprouting off a node representing a unique clonal TCR. Nodes are colored by disease group; lower right heatmap shows number of similar TCRs between unique samples per disease group f) TCR network (L-sim > 0.9) g) Quantification of shared motifs present in TCRβ sequences. Clonal TCRs with unambiguous CDR3b sequences were retained for analysis. The first and last two amino acids were removed from the analysis, since these regions are highly conserved h) Table of motifs of length 9. The SSLGQAYEQ motif is highlighted and metadata corresponding to patient samples is shown. i) Upper: table of pathogens specific to SSLGQAYEQ motif based on search in public McPAS-TCR database. Lower: table of SSLGQAYEQ motif-containing T cell types based on public McPAS-TCR database search.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementalFigures.pdf