CSF-resident CD4+ T-cells display a distinct gene expression profile with relevance to immune surveillance and multiple sclerosis

Short title:

Gene expression profiling in CSF CD4+ T-cells

James Hrastelj; Cardiff University, Cardiff, UK. ORCID 0000-0002-7991-5259.

Robert Andrews; Cardiff University, Cardiff, UK. ORCID 0000-0002-3491-2361.

Samantha Loveless; Cardiff University, Cardiff, UK.

Joanne Morgan; Cardiff University, Cardiff, UK. ORCID 0000-0003-3091-0040.

S. Mark Bishop; Cardiff University, Cardiff, UK. ORCID 0000-0001-7774-6925.

Nicholas J. Bray; Cardiff University, Cardiff, UK. ORCID 0000-0002-4357-574X.

Nigel M. Williams; Cardiff University, Cardiff, UK. ORCID 0000-0003-1177-6931.

Neil P. Robertson; Neuroscience Research Offices, B4-C4 link corridor, University Hospital of Wales, Cardiff. CF14 4XW. RobertsonNP@cardiff.ac.uk. ORCID 0000-0002-5409-4909.

Abstract

The CNS has traditionally been considered an immune privileged site, but is now understood to have a system of immune surveillance, predominantly involving CD4$^+$ T-cells. Identifying functional differences between CNS and blood CD4$^+$ T-cells therefore has relevance to CNS immune surveillance as well as to neurological conditions such as multiple sclerosis, in which CD4$^+$ T-cells play a central role. Here, CD4$^+$ T-cells were purified from CSF and blood from 21 patients with newly diagnosed treatment-naïve multiple sclerosis and 20 individuals with non-inflammatory disorders using fluorescence-activated cell sorting, and their transcriptomes were profiled by RNA sequencing. Paired comparisons between CD4$^+$ T-cells from CSF and blood identified 5,156 differentially expressed genes in controls and 4,263 differentially expressed in multiple sclerosis patients at false discovery rate < 5%. Differential expression analysis of CD4$^+$ T-cells collected from the CSF highlighted genes involved in migration, activation, cholesterol biosynthesis and signalling, including those with known relevance to multiple sclerosis pathogenesis and treatment. Expression of markers of CD4$^+$ T-cell subtypes suggested an increased proportion of Th1 and Th17 cells in CSF. Gene ontology terms significant only in multiple sclerosis were predominantly those involved in cellular proliferation. A two-way comparison of CSF versus blood CD4$^+$ T-cells in multiple sclerosis compared with non-inflammatory disorder controls identified four significant genes at false discovery rate < 5% (*CYP51A1*, *LRRD1*, *YES1* and *PASK*), further implicating cholesterol biosynthesis and migration mechanisms. Analysis of CSF CD4$^+$ T-cells in an extended cohort of multiple sclerosis cases (total N = 41) compared with non-inflammatory disorder controls (total N = 38) identified 140 differentially expressed genes at false discovery rate < 5%, many of which have known relevance to multiple sclerosis, including *XBP1*, *BHLHE40*, *CD40LG*, *DPP4* and *ITGB1*. This study provides the largest transcriptomic analysis of purified cell subpopulations in CSF to date and has relevance for the understanding of CNS immune surveillance, as well as multiple sclerosis pathogenesis and treatment discovery.

Key words

Multiple sclerosis, gene expression, CSF, CD4$^+$ T-cells, immune surveillance.

Abbreviations

DAVID = Database for Annotation, Visualisation and Integrated Discovery; eQTL = expression quantitative trait locus; FACS = fluorescence-activated cell sorting; FDR = false discovery rate; GO = gene ontology; MDS = multi-dimensional scaling; NID = non-inflammatory disorder; PBS = phosphate-buffered saline; RIN = RNA integrity number; RNA-seq = RNA sequencing; T_{EM} = effector memory T-cells; WNRTB = Welsh Neuroscience Research Tissue Bank.

Introduction

The CNS has, historically, been considered an immune-privileged site devoid of immunological activity in health, but is now known to be a site of highly regulated immune surveillance and homeostasis. Immune cells have been shown to cross the blood-brain barrier at post-capillary venules and the blood-CSF barrier at the choroid plexus and subarachnoid venules (Korn *et al.*, 2017). Cytological analysis has shown that CSF contains stable populations of immune cells in health and that these have a significantly different composition to those in blood (Chofflon *et al.*, 1989). CD4$^+$ T-cells are the predominant cell type in CSF, accounting for ~70% of cells and the majority of these (~90%) are antigen-experienced memory CD4$^+$ T-cells (de Graaf *et al.*, 2011; Han *et al.*, 2014). It has been proposed that the constellation of adhesion molecules expressed by the CNS endothelium (e.g. selectins and addressins) and their interactions with adhesion molecules expressed by T-cells (e.g. integrins) could be responsible for 'homing' of CNS antigen-specific CD4$^+$ T-cells to the CNS (Korn *et al.*, 2017). Such mechanisms are likely to be relevant to the pathogenesis of multiple sclerosis, as multiple lines of evidence, including genetic, immunological and animal model data, implicate CD4$^+$ T-cells in the condition (Dendrou *et al.*, 2015). Furthermore, one of the most potent disease-modifying treatments for multiple sclerosis, natalizumab, is a monoclonal antibody against integrin-α4, an adhesion molecule that facilitates CD4$^+$ T-cell extravasation into the CNS. Thus, identifying differences between CD4$^+$ T-cells in the CNS and blood could provide novel treatment targets for the disorder.

Characterisation of CD4$^+$ T-cells within the CNS in health and disease is challenged by the difficulty in obtaining sizeable cohorts, isolating sufficient numbers of cells and performing timely functional analysis. Genome-wide gene expression (or 'transcriptomic') analyses can provide valuable insights into cell state and function, but while several such analyses have been performed on CD4$^+$ T-cells from peripheral blood (Hu *et al*., 2014; Raj *et al*., 2014; Fahr *et al*., 2015; Davies *et al*., 2016; Hellberg *et al*., 2016; Brorson *et al*., 2019), few studies have investigated gene expression in CSF CD4$^+$ T-cells. To date, CSF analyses have either been restricted to a limited number of assayed genes, used unsorted CSF cells or used small patient cohorts (Brynedal *et al.*, 2010; Veroni *et al.*, 2015; Greenfield *et al.*, 2019). Although single-cell gene expression analyses have great potential for characterising CSF immune cells, costs and technical challenges have so far restricted sample sizes to no more than 12 participants

(Farhadian *et al.*, 2018; Esaulova *et al.*, 2020; Schafflick *et al.*, 2020). Large single-cell studies would allow high-resolution analysis of the roles of functional subtypes of CSF cells; however the extent of transcriptome coverage currently lags behind that provided by standard 'bulk' RNA-sequencing (Chen *et al*., 2019). An alternative, more powerful strategy for detecting differentially expressed genes is to perform transcriptomic analyses on purified CSF CD4$^+$ T-cell populations from large cohorts. In this study, we apply whole transcriptome sequencing to fluorescence-activated cell sorted CD4$^+$ T-cells from the CSF and blood of large cohorts of multiple sclerosis patients and non-inflammatory disorder (NID) controls, identifying thousands of genes that are differentially expressed between CD4$^+$ cells from CSF and blood, as well as genes that are differentially expressed in CSF CD4$^+$ cells in multiple sclerosis.

Materials and methods

Patient recruitment and sample acquisition

Potential participants were identified from patients referred for diagnostic lumbar punctures as part of their routine clinical care to the Neurology Ambulatory Unit at the University Hospital of Wales, Cardiff. Participants were recruited to the South Wales Multiple Sclerosis Epidemiology Project (South East Wales REC panel C, ref 05/WSE03/111) or Welsh Neuroscience Research Tissue Bank (WNRTB; Wale REC panel 3, ref 19/WA/0058) project. Written consent was obtained from each participant and lumbar puncture and venepuncture were performed according to standard clinical practice. 12ml blood and 10ml CSF was collected from each participant concomittantly and transferred on ice to the laboratory.

RNA sequencing (RNA-seq) data were available for a final cohort of 79 participants: 41 with multiple sclerosis (21 of which had both CSF and blood data) and 38 with NID (20 of which had both CSF and blood data). The diagnoses of the NID group included idiopathic intracranial hypertension (14), functional disorder (5), cerebral small vessel disease (5), migraine (3) and a range of other non-inflammatory disorders (supplementary table 1). The demographic characteristics of the included patients are shown in supplementary table 2. The NID controls were more frequently female than patients with multiple sclerosis (85.0% and 65.0%,

respectively; p=0.05), but were well matched for age at sampling and ethnicity. Diagnosis of multiple sclerosis was made according to the McDonald criteria appropriate at the time of diagnosis. None of the multiple sclerosis patients had experienced clinical relapse in the four weeks preceding sampling and no patient had been treated with immunomodulatory drugs. Disease-specific characteristics of the patients with multiple sclerosis are shown in supplementary table 3.

Fluorescence-activated cell sorting

Fluorescence-activated cell sorting (FACS) uses fluorescently labelled antibodies against specific cell surface molecules to count and sort specific cell types into separate tubes. Labelled antibodies were chosen to isolate pure populations of live $CD19^-CD14^-CD3^+CD4^+$ T-cells (supplementary table 4 and supplementary figures 1 and 2).

Within one hour of sample collection, CSF was centrifuged at 200 x g for 10 minutes at 4°C. At room temperature, 1μL LIVE/DEAD® ViViD aqua stain (Invitrogen/ThermoFisher, Paisley, UK; pre-diluted) was added to 39μL PBS and the mixture was vortexed and centrifuged down. CSF supernatant was transferred into a 50ml tube and placed on ice. The cell pellet was resuspended in 50μL of phosphate-buffered saline (PBS) and transferred into a FACS tube (BD Biosciences, Oxford, UK). At room temperature, 8μL of diluted LIVE/DEAD® ViViD aqua stain mixture was added to the cell suspension and incubated at room temperature for 10 mins, shielded from the light. Each labelled antibody was then added to the CSF, which was incubated at 4°C for 20 minutes.

Compensation tubes were prepared containing antibody-binding and inert polystyrene beads (BD Biosciences, Oxford, UK). 50μL of bead suspension was added to five FACS tubes. Each fluorescently labelled antibody was added to a separate bead suspension tube and one bead suspension tube was left unstained. The emission range of the three labels used to exclude cells overlap almost completely (LIVE/DEAD®, anti-CD14 and anti-CD19), so only one compensation tube was necessary for all of these. The bead suspension and labelled antibody mixtures were incubated at room temperature for 10 minutes, shielded from the light. Once the CSF incubation was complete, 2ml PBS was added and the mixture was centrifuged at 200

x *g* for three minutes to wash the cells. The supernatant was removed and the cell pellet was resuspended in 250µL PBS and placed on ice.

2ml of the blood sample was poured into a 15ml tube, PBS was added to a total volume of 10ml and the suspension was mixed gently. 15ml of density gradient cell separation media (Ficoll-Paque™ PLUS, GE Healthcare, Little Chalfont, UK) was added to two SepMate™ tubes (STEMCELL Technologies, Cambridge, UK). Half of the diluted blood sample was then carefully added to each SepMate™ tube to form a layer on the surface of the Ficoll-Paque™. The tubes were centrifuged at 1200 x *g* for 15 minutes at room temperature. The supernatant, containing the layer of mononuclear cells, in each tube was poured into a 50ml tube. PBS was added to a total volume of 30ml in each tube. The cell suspensions were centrifuged at 300 x *g* for 10 minutes at 4°C. The supernatant was poured away and the cell pellets were resuspended in 1ml PBS and transferred to a single FACS tube. The cell suspension then underwent the same staining protocol as the CSF cell suspension (above), including staining with LIVE/DEAD®, anti-CD14 V500, anti-CD19 V500, anti-CD3 APC-H7, anti-CD4 PE-Cy5.5 and anti-CD8 BV711. The washed cell pellet was resuspended in 3ml PBS and placed on ice.

Cell sorting was peformed using the FACSaria Fusion. First, the compensation tubes were loaded and any overlaps in fluorescence emission ranges for the stains/antibodies were adjusted for. Next, the stained CSF cell suspension was loaded and the first 200 cells were used to apply the compensated gates. The rest of the sample was then sorted into purified CD4[+] T-cell populations, which were collected into genomic DNA extraction columns (AllPrep Micro kit, Qiagen, Manchester, UK) containing 350µL RLT lysis buffer (AllPrep Micro kit, Qiagen, Manchester, UK) and 1% beta-mercaptoethanol (irreversible RNase inhibitor; Invitrogen, Paisley, UK), and placed on ice. The number of CSF CD4[+] T-cells sorted ranged from 2-6073 cells/ml of CSF (median 108.1 cells/ml, mean 320.6 cells/ml) (supplementary figure 3). After the CSF cell sorting was complete, the stained mononuclear cell suspension was loaded. Up to 150,000 purified CD4[+] T-cells were collected into a 15ml tube containing 350µL RLT lysis buffer and 1% beta-mercaptoethanol, and placed on ice.

Extraction of nucleic acids, library preparation and sequencing

Extraction and purification of nucleic acids was performed using the Qiagen AllPrep Micro kit (Qiagen, Manchester, UK). RNA was eluted in RNase-free water and stored at -80°C. RNA quality and quantity was measured using the Agilent Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). The range of RNA concentrations from CSF-resident CD4$^+$ T-cells was 56-948pg/µL (median 85 pgµ/L, mean 151pg/µL). In multiple sclerosis, the mean RNA integrity number (RIN) for the samples that were included in the paired analysis was 8.4 for CSF and 7.7 for blood (t-test comparison p=0.18). In NID controls, the mean RIN for the samples that were included in the paired analysis was 6.8 for CSF and 6.3 for NID controls (t-test comparison p=0.50). In the extended CSF cohort, the mean RIN was 6.7 in multiple sclerosis and 5.5 in NID controls (t-test comparison p=0.06).

RNA-seq libraries were prepared from 500pg RNA from each sample using the Tecan Ovation SoLo RNA-seq System (Tecan Genomics, The Netherlands), a system designed for low RNA input libraries, according to the manufacturer's protocol. In order to minimise batch effects, each batch was designed to contain paired CSF and blood RNA samples from two patients with multiple sclerosis and two NID control patients. The RNA-seq libraries underwent QC using the Agilent Bioanalyzer and were then sequenced to a median depth of 157,417,402 reads using the Illumina HiSeq4000.

Mapping, read counting and quality control

FASTQ files containing raw RNA-seq data underwent QC using FastQC. FastQC reports for all samples were inspected for abnormalities both before and after read trimming using Trimmomatic (Bolger *et al*., 2014) to remove adapter sequences and bases with low sequence quality scores. Reads were mapped to the human genome (GRCh38, gene build 91) using STAR (Dobin *et al*., 2013). MarkDuplicates (within the Picard suite; Broad Institute) was used to identify duplicate reads and mapping statistics were generated using BamTools (Barnett *et al*., 2011). A file containing raw read counts for each gene for each sample was created using featureCounts (Liao *et al*., 2014). Normalisation was performed during subsequent analyses.

Differential gene expression and gene ontology analysis

Experimental design

The differential gene expression analysis had four components. The first analysis compared CSF CD4[+] T-cells with blood CD4[+] T-cells in 20 NID controls (paired analysis) with RIN as a covariate. This sought to identify differentially expressed genes in the absence of CNS inflammation that could be important in immune surveillance of the CNS by CD4[+] T-cells. The second analysis compared CSF CD4[+] T-cells with blood CD4[+] T-cells in 21 multiple sclerosis patients (paired analysis) with RIN as a covariate. This sought to identify differentially expressed genes in multiple sclerosis. The third analysis was a two-way paired analysis of CSF CD4[+] T-cells vs blood CD4[+] T-cells in these 21 cases compared with the 20 controls. This sought to identify genes differentially expressed in CSF CD4[+] T-cells when compared with blood CD4[+] T-cells, specifically in multiple sclerosis. The final analysis compared multiple sclerosis CSF CD4[+] T-cells (n=41, including the 21 cases included in previous analyses) with NID control CSF CD4[+] T-cells (n=38, including the 20 controls used in the previous analyses) (group-wise comparison) with RIN, batch, age at sampling and sex as covariates. This analysis sought to identify genes differentially expressed between CSF CD4[+] T-cells in multiple sclerosis and NID controls.

As CSF and blood samples for each patient were taken at the same time and each batch was arranged to minimise batch effects, the first three analyses were paired to increase power to detect differences. This design controlled for patient- and sample-specific factors that influence gene expression e.g. age, sex, batch etc. Mapping statistics for all cohorts are shown in supplementary tables 5 and 6.

Differential expression analyses

All analyses of gene expression were completed using the Bioconductor R package, edgeR (Robinson *et al*., 2010), employing multi-dimensional scaling to identify samples with outlying expression data (supplementary figures 4 and 5, supplementary tables 7 and 8). Data for genes

with low expression were removed using filterByExpr (Chen *et al*., 2016), leaving 10,457 genes for differential analysis.

Gene ontology analysis

Gene ontology (GO) analysis was performed using the online Database for Annotation, Visualisation and Integrated Discovery (DAVID) v6.8 (Huang *et al*., 2009). The list of expressed genes derived from filterByExpr was used as the background gene list. Significantly differentially expressed genes at a false discovery rate (FDR) < 5% were uploaded to DAVID and pathways were considered significant at FDR < 5%.

Data availability

Gene expression data will be made available via the European Genome-phenome Archive (https://www.ebi.ac.uk/ega/home).

Results

CSF CD4[+] T-cells display a distinct gene expression profile to blood CD4[+] T-cells, which is similar in non-inflammatory controls and multiple sclerosis and is predominated by migration molecules

The differential gene expression analysis between CSF CD4[+] T-cells and blood CD4[+] T-cells in NID controls identified 5,156 significantly differentially expressed genes at FDR <5% (figure 1 and supplementary table 9): 2,556 up-regulated and 2,600 down-regulated. The differential gene expression analysis between CSF CD4[+] T-cells and blood CD4[+] T-cells in patients with multiple sclerosis identified 4,263 differentially expressed genes at FDR <5% (figure 2 and supplementary table 10): 2,102 up-regulated and 2,161 down-regulated. Of the 4,263 genes found to be significantly differentially expressed between CSF and blood in multiple sclerosis, 3,511 (82%) were also significantly differentially expressed in NID controls.
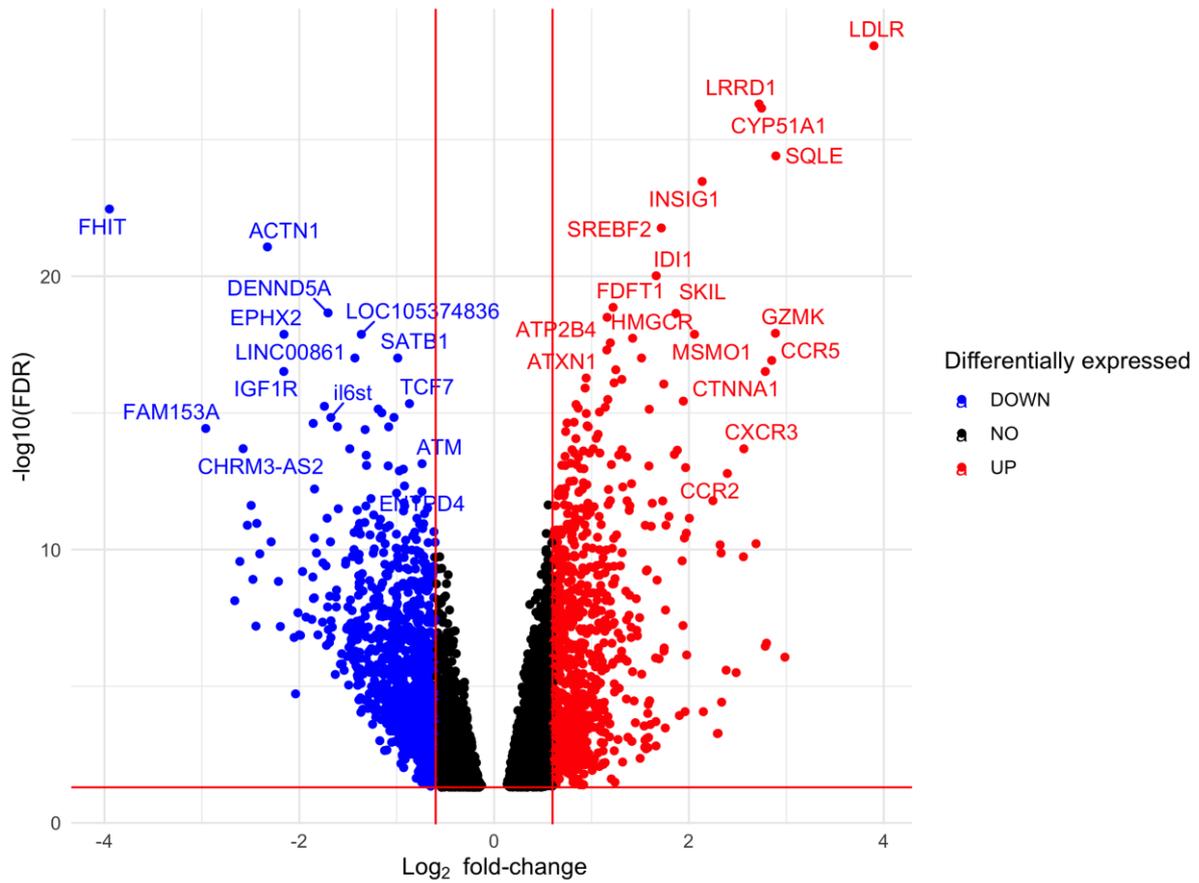
Figure 1. Volcano plot showing differentially expressed genes between CSF and blood CD4⁺ T-cells in non-inflammatory controls. Log_2 fold-change thresholds of +/-0.6 and FDR 0.05 are drawn as red lines for visualisation purposes.
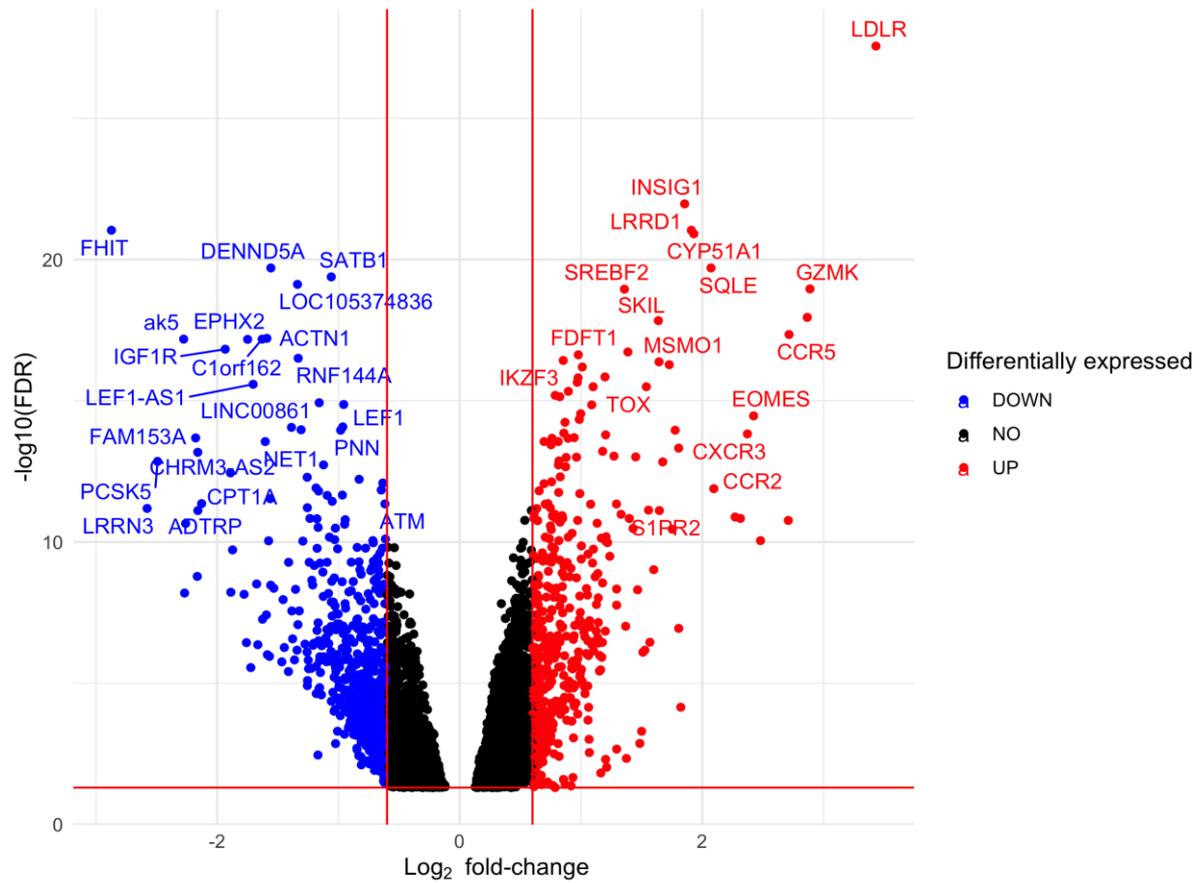
Figure 2. Volcano plot showing differentially expressed genes between CSF and blood CD4+ T-cells in multiple sclerosis. Log$_2$ fold-change thresholds of +/-0.6 and FDR 0.05 are drawn as red lines for visualisation purposes.

The significantly differentially expressed genes in both NID controls and multiple sclerosis were predominantly genes involved in cellular migration and activation. Interestingly, genes associated with Th1 phenotype (e.g. *CXCR3*, *CCR5; supplementary table 11; Sallusto et al., 1998) were significantly up-regulated in CSF. *CCR4*, a gene expressed by Th2 cells (Sallusto et al., 1998), was up-regulated in CSF ($\log_2$ fold-change 0.93 in NID and 1.05 in multiple sclerosis), but to a lesser extent than Th1-related genes (*CCR5* $\log_2$ fold change 2.85 in NID and 2.72 in multiple sclerosis) and other Th2-related genes were down-regulated in CSF (e.g. *IL4RA*, *IRF4*, *STAT5* and *STAT6*). Genes expressed by Th17 cells (e.g. *KLRB1;* Kamishikiryo et al., 2011) and tissue-resident T-cells (e.g. *CCL5*, *LGALS1* and *PTGER2*) were significantly up-regulated in CSF CD4[+] T-cells. Furthermore, genes associated with effector functions were also significantly up-regulated in CSF CD4[+] T-cells, including *GZMK*, *GZMA* and *CST7*.

GO analysis of the 5,156 significantly differentially expressed genes between blood and CSF CD4[+] cells in NID controls identified 61 significant GO terms at FDR <5% (supplementary table 12). GO analysis of the 4,263 significantly differentially expressed genes between blood and CSF CD4[+] cells in multiple sclerosis identified 176 significant GO terms at FDR <5% (supplementary table 13). The GO term 'Movement of cell or subcellular component' was the most significant term in both NID controls (fold enrichment 1.23, FDR 5 x 10^{-7}) and multiple sclerosis (fold enrichment 1.32, FDR 2.7 x 10^{-10}). Fifty-eight of the 61 significant GO terms in NID controls were also significant in multiple sclerosis (table 1), leaving 118 GO terms significant only in multiple sclerosis (table 2) and three GO terms significant only in NID controls (supplementary table 14). The GO terms only significant in multiple sclerosis were predominated by those involved in cellular proliferation. There was consistency in the top GO terms when considering all significant (FDR < 0.05) genes, or taking only the 1000, 500 and 100 most significant genes, with the 100 most significantly differentially expressed genes producing pronounced fold-change enrichments (supplementary table 15).

GO term	Fold enrichment	FDR (NID)
Movement of cell or subcellular component	1.32	1.3×10^{-9}
Single organismal cell-cell adhesion	1.43	1.1×10^{-7}
Cell migration	1.35	2.0×10^{-7}
Cell activation	1.36	2.0×10^{-7}
Locomotion	1.31	2.0×10^{-7}
Single organism cell adhesion	1.39	3.5×10^{-7}
Immune response	1.27	4.5×10^{-7}
Cell motility	1.31	7.3×10^{-7}
Localization of cell	1.31	7.3×10^{-7}
Biological adhesion	1.25	2.1×10^{-6}

Table 1. Ten most significant gene ontology terms common to both multiple sclerosis and non-inflammatory disorder controls.

GO term	Fold enrichment	FDR
Mononuclear cell proliferation	1.62	3.5×10^{-6}
Leukocyte proliferation	1.61	3.6×10^{-6}
Lymphocyte proliferation	1.61	7.2×10^{-6}
Leukocyte activation	1.34	8.9×10^{-6}
Regulation of mononuclear cell proliferation	1.68	1.3×10^{-5}
Regulation of leukocyte proliferation	1.66	1.5×10^{-5}
Regulation of lymphocyte proliferation	1.67	1.7×10^{-5}
Regulation of cell adhesion	1.34	1.0×10^{-4}
Regulation of cell proliferation	1.22	1.3×10^{-4}
Regulation of cell activation	1.38	1.9×10^{-4}

Table 2. Gene ontology terms significant only in multiple sclerosis (ten most significant shown).

Two-way paired analysis of CSF and blood CD4[+] T-cells in multiple sclerosis compared with non-inflammatory controls identifies four genes that further implicate cholesterol biosynthesis and migration mechanisms

Gene expression differences specific to CSF CD4[+] T-cells from patients with multiple sclerosis could be relevant to pathogenesis. It is possible that CD4[+] T-cells in multiple sclerosis could exploit disease-specific extravasation mechanisms, are more likely to extravasate or display pro-inflammatory tendencies. To investigate this, a two-way paired differential gene expression analysis was performed between CSF CD4[+] T-cells and blood CD4[+] T-cells in patients with multiple sclerosis compared with NID controls. This analysis sought to identify genes where the difference in expression between CSF and blood CD4[+] T-cells differed significantly between multiple sclerosis and NID controls. Four genes were identified that were differentially expressed in multiple sclerosis CSF CD4[+] T-cells at FDR <5% (table 3): one up-regulated and three down-regulated (supplementary figure 6).

Gene symbol	Molecule	Log fold-change	P value	FDR
LRRD1	Leucine-rich repeat and death domain-containing protein 1	-0.81	0.000003	0.02
CYP51A1	Lanosterol 14α-demethylase	-0.81	0.000004	0.02
PASK	PAS domain-containing serine-threonine-protein kinase	0.62	0.000007	0.02
Yes1	Tyrosine-protein kinase Yes	-1.14	0.000009	0.02

Table 3. Genes significantly differentially expressed in CSF CD4[+] T-cells compared with blood CD4[+] T-cells in patients with multiple sclerosis compared with NID controls.

CSF CD4⁺ T-cells in patients with multiple sclerosis compared with non-inflammatory controls identifies 140 differentially expressed genes

To seek further genes relevant to multiple sclerosis, a differential expression analysis was performed between an extended cohort of CSF CD4+ T-cells in 41 multiple sclerosis samples vs 38 NID control samples, which identified 140 significant genes at FDR <5% (figure 3 and supplementary table 16): 115 up-regulated and 25 down-regulated.
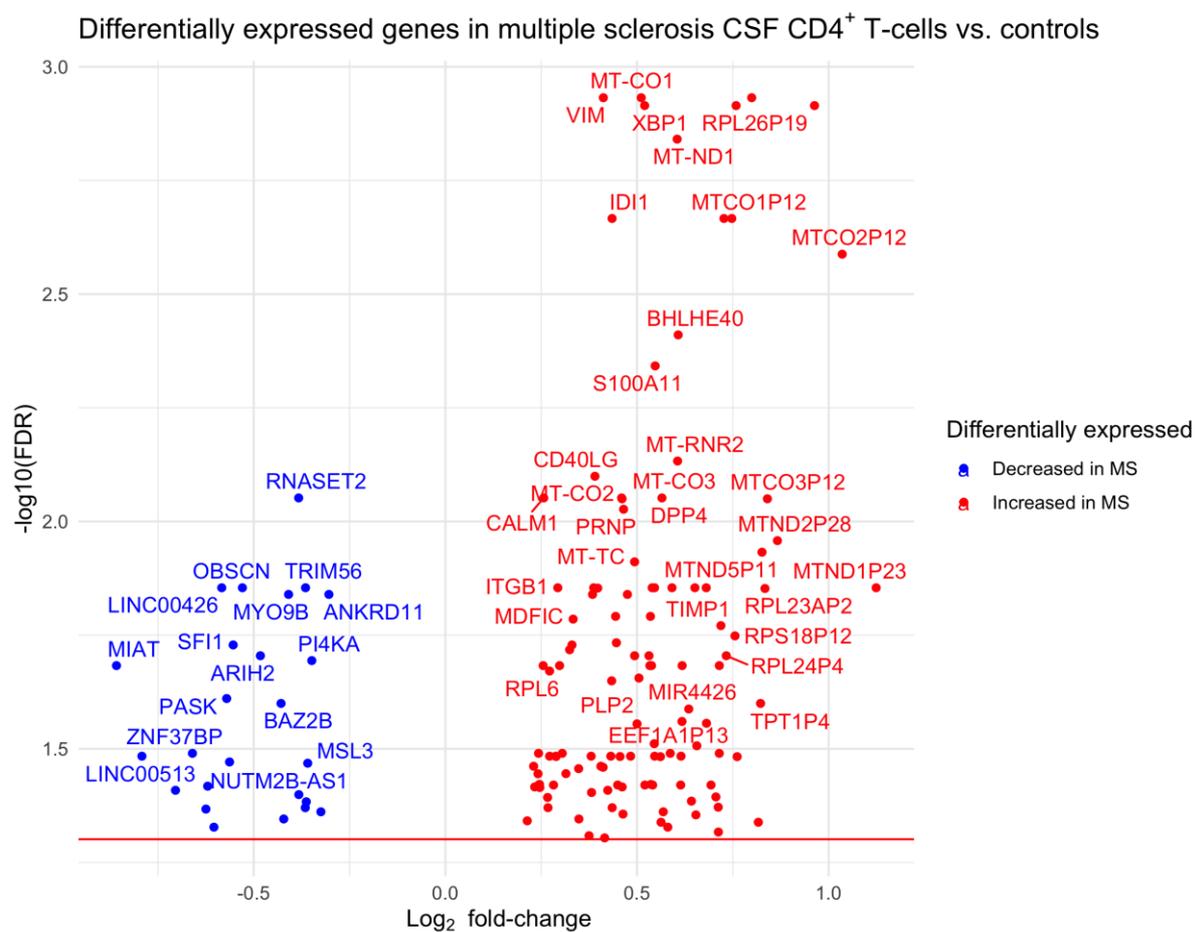

Figure 3. Volcano plot showing differentially expressed genes between multiple sclerosis and non-inflammatory control CSF CD4+ T-cells.

GO analysis was performed of the 140 significantly differentially expressed genes. This identified 31 significant GO terms at FDR <5% (supplementary table 17), which broadly relate to mitochondrial function, with many genes involved in oxidative phosphorylation.

Discussion

Immune surveillance of the CNS plays an important role in protection from infection and maintaining health (Svenningsson *et al*. 1993; Kivisakk *et al*., 2002; Engelhardt *et al*., 2016). The majority of cells found in CSF are CD4$^+$ T-cells (Han *et al*., 2014) and the ways in which they differ from those in blood are likely to be important in the pathogenesis of multiple sclerosis and other CNS inflammatory disorders, but are incompletely understood. Here, we report thousands of differentially expressed genes in CSF CD4$^+$ T-cells compared with blood CD4$^+$ T-cells, irrespective of diagnosis. The results suggest substantial differences in activation status, expression of adhesion molecules and functionality of CD4$^+$ T-cells within the CSF when compared with those in blood. The results also suggest that functional differences between CSF and blood CD4$^+$ T-cells are far greater than differences between CSF CD4$^+$ T-cells in multiple sclerosis and NID controls. Furthermore, the results also highlight genes and cellular processes that may be relevant to multiple sclerosis pathogenesis and treatment target discovery.

It has been reported that CSF cells differ from those in blood in several ways. CD4$^+$ T-cells make up a higher proportion of total CSF cells compared with blood and these cells are larger than those in blood (Han *et al*., 2014). Several studies have reported that, in health, CSF CD4$^+$ T-cells express different surface proteins, such as those associated with different functional subtypes. In particular, a high proportion of CSF CD4$^+$ T-cells have been reported to display a memory phenotype, either central memory or effector memory (Svenningsson *et al*., 1993; Kivisakk *et al*., 2003; Mullen *et al*., 2012). The data we report here suggest that CSF CD4$^+$ T-cells from both multiple sclerosis and NID controls are antigen-experienced. The most significantly differentially expressed gene in both multiple sclerosis and NID controls was *LDLR*, the receptor for low density lipoprotein (LDL). The LDL receptor mediates endocytosis of cholesterol-rich LDL and has been shown to be up-regulated on the surface of T cells following mitogenic stimulation (Cuthbert *et al*., 1990; Suzuki *et al*., 1990). Activated CD4$^+$ T-cells undergo rapid proliferation as part of the adaptive immune response. Cholesterol and other sterols are required to build membranes and for intracellular signalling (Kidani *et al*., 2013), so the up-regulation of sterol synthesis and cholesterol uptake observed in CSF CD4$^+$ T-cells here is consistent with proliferating cells (Twiner *et al*., 2008).

Several studies of CNS inflammatory disorders have shown increased proportions of T-cells displaying a memory phenotype (Vretham *et al*., 1998; Giunti *et al*., 2003; Kivisakk *et al*., 2004; Okuda *et al*., 2005). Previous studies have reported that the majority of CD4$^+$ T-cells in inflamed brain parenchyma are T$_{EM}$ (effector memory) cells (Herich *et al*., 2019), which lack expression of the chemokine receptor *CCR7* and are capable of immediate effector functions. In our data *CCR7* was significantly downregulated in CSF CD4$^+$ T-cells, suggesting enhanced effector functionality. Other markers of effector function (*GZMK*, *GZMA*, *CST7*) were also up-regulated in CSF CD4$^+$ T-cells in this study. These data replicate two recent single cell transcriptomic studies in small numbers of patients, which reported that CSF CD4$^+$ T-cells predominantly express markers associated with effector functions (Pappalardo *et al*., 2020; Schafflick *et al*., 2020). Furthermore, expression of *CCR5*, a Th1 cell marker, was significantly higher in CSF CD4$^+$ T-cells and it has been reported that immune surveillance of brain parenchyma is performed by T$_{EM}$ cells expressing *GZMK* and high levels of *CCR5* (Herich *et al*., 2019).

Lymphocyte homing mechanisms that permit CNS-specific CD4$^+$ T-cell extravasation contain potential targets for treating CNS inflammatory disorders, such as multiple sclerosis. One of the most effective disease modifying drugs for multiple sclerosis, natalizumab, is a monoclonal antibody against integrin-α4 (ITGA4). ITGA4 is expressed on the surface of antigen-experienced CD4$^+$ T-cells. By interfering with the interaction between ITGA4 and its receptor (VCAM-1), natalizumab significantly reduces extravasation of CD4$^+$ T-cells into the CNS. In this study *ITGA4* expression was significantly increased in CSF CD4$^+$ T-cells when compared with blood CD4$^+$ T-cells (log fold change 1.7; FDR = 6.7×10^{-16} and 6.5×10^{-16} in NID controls and multiple sclerosis, respectively). Other molecules involved in cellular migration reported here may contain novel treatment targets for multiple sclerosis and further study is warranted.

The 118 GO terms significant only in multiple sclerosis were predominated by those involved in cellular proliferation. On activation, naïve CD4$^+$ T-cells proliferate, and the local cytokine milieu determines the phenotype of the differentiated cells (Smith-Garvin *et al*., 2009). CSF CD4$^+$ T-cell counts in multiple sclerosis were significantly higher than those of NID controls in this study (median in multiple sclerosis and NID controls 298.9 cells/ml and 81.6 cells/ml, respectively; p=1.4×10^{-6}) and this is consistent with published data (Heming *et al*., 2020). Whilst increased blood-brain- or blood-CSF-barrier permeability in multiple sclerosis is likely

to be contributory, increased CD4[+] T-cell proliferation could also be relevant. Indeed, differences in activation responses and clonal expansion of CD4[+] T-cells has been implicated in multiple sclerosis pathogenesis (Hellberg *et al*., 2016; Sousa *et al*., 2016).

The two-way analysis to identify genes differentially expressed in CSF vs. blood in multiple sclerosis vs. non-inflammatory controls yielded four significant genes: *LRRD1, CYP51A1, PASK* and *Yes1*. Within the human genome, *LRRD1* and *CYP51A1* are in close proximity (within ~30kb). *CYP51A1* encodes lanosterol 14α-demethylase, which is an enzyme that catalyses a critical step in cholesterol synthesis in T-cells (Lepesheva *et al*., 2007; Bietz *et al*., 2017). Cholesterol synthesis intermediates have been shown to play important roles in regulating T-cell differentiation and inflammatory responses (Jetten *et al*., 2020). The effect of pharmacological manipulation of cholesterol biosynthesis in multiple sclerosis warrants further study. *Yes1* is a Src-family kinase involved in T-cell migration through interaction with CXCL12 (Varrin-Doyer *et al*., 2009), which is up-regulated in multiple sclerosis lesions (Krumbholz *et al*., 2006).

The direct comparison of gene expression between multiple sclerosis and non-inflammatory control CSF CD4[+] T-cells yielded 140 significant genes and GO analysis identified 31 enriched terms. The fourth most significant gene, *XBP1* (X-box binding protein 1), is a transcription factor that regulates expression of HLA class II genes (Ono *et al*., 1991; Iwakoshi *et al*., 2003; Bettigole *et al*., 2015). HLA class II is the major multiple sclerosis genetic risk locus and is expressed on activated CD4[+] T-cells (Ko *et al*., 1979). In this study *HLA-DRB1* was found to be significantly up-regulated in CSF CD4[+] T-cells in both multiple sclerosis and NID controls. The function of HLA class II molecules in this context is debated and may involve activation or apoptosis of other CD4[+] T-cells (Holling *et al*., 2004) and warrants further investigation. The significant genes included many with plausible links with multiple sclerosis pathogenesis, such as *BHLHE40* (Lin *et al*., 2014)*, CD40LG* (AnzGene consortium, 2009), *DPP4* and *ITGB1*. The GO analysis identified 31 enriched terms, which were all related to mitochondrial function, and there is growing evidence of mitochondrial dysfunction in multiple sclerosis (De Barcellos *et al*., 2019).

To our knowledge, this is the largest transcriptomic study in purified subpopulations of CSF cells in health or disease to date. Although single cell analysis has the potential to provide more definitive conclusions about the relative proportions of functional subtypes of CD4[+] T-

cells present in CSF, the current costs are prohibitive and restrict analysis to patient cohorts much smaller than that in this study. Whilst CD4$^+$ T-cells play a key role in multiple sclerosis pathogenesis other immune cells are also important. The majority of CSF cells are CD4$^+$ T-cells, but CSF also contains CD8$^+$ T-cells and a limited number of other cell types. Analysis of other cell types could be informative, but the technical challenge of performing RNA-seq on sorted CSF cells is restricting. In this study, a large cohort of individuals with non-inflammatory disorders were used as controls. The non-inflammatory disorders largely consisted of idiopathic intracranial hypertension, a disorder thought to be due to impaired CSF drainage or over production (Binder *et al*., 2004). A range of other disorders made up the remaining controls. Whilst none of these disorders are thought to be primarily inflammatory, it is possible that the function of CSF immune cells is abnormal in some of these individuals. CSF from health individuals would be an ideal control, but it is an extremely scarce resource. In addition, inclusion of controls with non-multiple sclerosis CNS inflammatory disorders could identify genes that are more specific to multiple sclerosis pathogenesis, but the low prevalence of such disorders would have restricted the group size significantly. Finally, it is possible that the gene expression differences between multiple sclerosis and NID controls were a consequence of disease. The aetiological significance of observed gene expression differences could be explored by expression quantitative trait locus (eQTL) analyses to identify gene expression differences that are additionally associated with genetic risk for multiple sclerosis.

In summary, we report significant differences in gene expression between CD4$^+$ T-cells from CSF and blood, and between multiple sclerosis and NID controls. Major differences were identified in activation and migration mechanisms in CD4$^+$ T-cells, as well as cholesterol uptake and biosynthesis, which plays a crucial role in orchestrating CD4$^+$ T-cell differentiation into effector cells. The results implicate an array of adhesion molecules and chemokine receptors in extravasation of CD4$^+$ T-cells into the CNS. Identified genes warrant further investigation as potential therapeutic targets for multiple sclerosis.

Acknowledgements

Competing interests

Neil Robertson has received honoraria and/or support to attend educational meetings from Biogen, Novartis, Sanofi Genzyme, Celgene and Roche. His institution has also received research support from Biogen, Novartis and Genzyme. James Hrastelj has received support to attend an educational meeting by Novartis. The other authors have no competing interests.

References

ANZgene consortium. Genome-wide association study identifies new multiple sclerosis susceptibility loci on chromosomes 12 and 20. *Nat Genet* 2009; 41(7):824-8.

De Barcelos IP, Troxwell RM, Graves JS. Mitochondrial dysfunction in multiple sclerosis. *Biology* 2019; 8(2):37.

Barnett DW, Garrison EK, Quinlan AR, Strömberg MP et al. BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics* 2011; 27(12):1691-2.

Bettigole SE, Lis R, Adoro S, Lee AH et al. The transcription factor XBP1 is selectively required for eosinophil differentiation. *Nat Immunol* 2015; 16(8):829-37.

Bietz A, Zhu H, Xue M, Xu C. Cholesterol metabolism in T cells. [Review]. *Front Immunol* 2017; 8:1664.

Binder DK, Horton JC, Lawton MT, McDermott MW. Idiopathic intracranial hypertension. *Neurosurgery* 2004; 54(3):538-51.

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequencing data. *Bioinformatics* 2014; 30(15):2114-20.

Brorson IS, Eriksson A, Leikfoss IS, Celius EG et al. No differential gene expression for CD4$^+$ T-cells of MS patients and healthy controls. *Mult Scler J Exp Transl Clin* 2019; 5(2): 2055217319856903.

Brynedal B, Khademi M, Wallström E, Hillert J et al. Gene expression profiling in multiple sclerosis: A disease of the central nervous system, but with relapses triggered in the periphery? *Neurobio Dis* 2010; 37:613-21.

Chen G, Ning B and Shi T. Single-cell RNA-seq technologies and related computational data analysis. [Review]. *Front Genet* 2019; 10:317.

Chen Y, Lun ATL, Smyth GK. From reads to genes to pathways: differential expression analysis of RNA-seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. *F1000Research* 2016; 5:1438.

Chofflon M, Gonzalez V, Weiner HL, Hafler DA. Inflammatory cerebrospinal fluid T cells have activation requirements characteristic of CD4$^+$CD45RA$^-$ T cells. *Eur J Immunol* 1989; 19:1791-5.

Cuthbert JA, Lipsky PE. Mitogenic stimulation alters the regulation of LDL receptor gene expression in human lymphocytes. *J Lipid Res* 1990; 31(11):2067-78.

Davies JL, Thompson S, Kaur-Sandhu H, Sawcer S et al. Increased *THEMIS* first exon usage in CD4+ T-cells is associated with a genotype that is protective against multiple sclerosis. *PLoS ONE* 2016; 11(7):e0158327.

de Graaf MT, Sillevis Smitt PAE, Luitwieler RL, van Velzen C et al. Central memory CD4$^+$ T cells dominate the normal cerebrospinal fluid. *Clinical Cytometry Part B* 2011; 80B:43-50.

Dendrou CA, Fugger L, Friese MA. Immunopathology of multiple sclerosis. [Review]. *Nat Rev Immunol* 2015; 15:545-58.

Dobin A, Davis CA, Schlesinger F, Drenkow J et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013; 29(1):15-21.

Engelhardt B, Carare RO, Bechmann I, Flügel A et al. Vascular, glial, and lymphatic immune gateways of the central nervous system. [Review]. *Acta Neuropathol* 2016; 132:317-38.

Esaulova E, Cantoni C, Shchukina I, Zaitsev K et al. Single-cell RNA-seq analysis of human CSF microglia and myeloid cells in neuroinflammation. *Neurol Neuroimmunol Neuroinflam* 2020; 7(4):e732.

Fahr KKW, Marson A, Zhu J, Kleinewietfeld M et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* 2015; 518:337-43.

Farhadian SF, Mehta SS, Zografou C, Robertson K et al. Single-cell RNA sequencing reveals microglia-like cells in cerebrospinal fluid during virologically suppressed HIV. *JCI Insight* 2018; 3(18):e121718.

Giunti D, Borsellino G, Benelli R, Marchese M et al. Phenotypic and functional analysis of T cells homing into the CSF of subjects with inflammatory diseases of the CNS. *J Leukocyte Biol* 2003; 73(5):584-90.

Greenfield AL, Dandekar R, Ramesh A, Eggers EL et al. Longitudinally persistent cerebrospinal fluid B cells can resist treatment in multiple sclerosis. *JCI Insight* 2019; 4(6):e216599.

Han S, Lin YC, Wu T, Salgado AD et al. Comprehensive immunophenotyping of cerebrospinal fluid cells in patients with neuroimmunological diseases. *J Immunol* 2014; 192:2551-63.

Hellberg S, Eklund D, Gawel DR, Köpsén M et al. Dynamic response genes in CD4$^+$ T-cells reveal a network of interactive proteins that classifies disease activity in multiple sclerosis. *Cell Rep* 2016; 16:2928-39.

Heming M, Lohmann L, Schulte-Mecklenbeck A, Brix T et al. Leukocyte profiles in blood and CSF distinguish neurosarcoidosis from multiple sclerosis. *J Neuroimmunol* 2020; 341:577171.

Herich S, Schneider-Hohendorf T, Rohlman A, Ghadiri M et al. Human CCR5^{high} effector memory cells perform CNS parenchymal immune surveillance via GZMK-mediated transendothelial diapedesis. *Brain* 2019; 142:3411-27.

Holling TM, Schooten E, van Den Elsen JP. Function and regulation of MHC class II molecules in T-lymphocytes: of mice and men. [Review]. *Hum Immunol* 2004; 65(4):282-90.

Hu X, Kim H, Raj T, Brennan PJ et al. Regulation of gene expression in autoimmune disease loci and the genetic basis of proliferation in CD4$^+$ effector memory T cells. *PLoS Genet* 2014; 10(6):e1004404.

Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009; 4(1):44-57.

Iwakoshi NN, Lee AH, Vallabhajosyula P, Otipoby KL et al. Plasma cell differentiation and the unfolded protein response intersect at the transcription factor XBP-1. *Nat Immunol* 2003; 4(4):321-9.

Jetten AM, Cook DN. (Inverse) agonists of retinoic acid-related orphan receptor γ: regulation of immune responses, inflammation and autoimmune disease. [Review]. *Annu Rev Pharmacol Toxicol* 2020; 60:371-90.

Kamishikiryo J, Fukuhara H, Okabe Y, Kuroki K et al. Molecular basis for LLT1 protein recognition by human CD161 protein (NKRP1A/KLRB1). *J Biol Chem* 2011; 286(27):23832-30.

Kidani Y, Elsaesser H, Hock MB, Vergnes L et al. Sterol regulatory element-binding proteins are essential for the metabolic programming of effector T cells and adaptive immunity. *Nat Immunol* 2013; 14(5):489-99.

Kivisakk P, Trebst C, Liu Z, Tucky BH et al. T-cells in the cerebrospinal fluid express a similar repertoire of inflammatory chemokine receptors in the absence or presence of CNS inflammation: implications for CNS trafficking. *Clin Exp Immunol* 2002; 129:510-8.

Kivisakk P, Mahad DJ, Callahan MK, Trebst C et al. Human cerebrospinal fluid central memory CD4[+] T-cells: evidence for trafficking through the choroid plexus and meninges via P-selectin. *Proc Nat Acad Sci USA* 2003; 100(14):8389-94.

Kivisakk P, Mahad DJ, Callahan MK, Sikora K et al. Expression of CCR7 in multiple sclerosis: implications for CNS immunity. *Ann Neurol* 2004; 55(5):627-38.

Ko HS, Fu SM, Winchester RJ, Yu DT et al. Ia determinants on stimulated human T lymphocytes. Occurrence on mitogen- and antigen-activated T cells. *J Exp Med* 1979; 150:246.

Korn T and Kallies A. T cell responses in the central nervous system. [Review]. *Nat Rev Immunol* 2017; 17:179-94.

Krumbholz M, Theil D, Cepok S, Hemmer B et al. Chemokines in multiple sclerosis: CXCL12 and CXCL13 up-regulation is differentially linked to CNS immune cell recruitment. *Brain* 2006; 129(1):200-11.

Lepesheva GI, Waterman MR. Sterol 14alpha-demethylase cytochrome P450 (CYP51), a P450 in all biological kingdoms. [Review]. *Biochimica et Biophysica Acta* 2007; 1770(3):467-77.

Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 2014; 30(7):923-30.

Lin CC, Bradstreet TR, Schwarzkopf EA, Sim J et al. Bhlhe40 cytokine production by T cells and is essential for pathogenicity in autoimmune neuroinflammation. *Nat Commun* 2014; 5:3551.

Mullen KM, Gocke AR, Allie R, Ntranos A et al. Expression of CCR7 and CD45RA in CD4[+] and CD8[+] subsets in cerebrospinal fluid of 134 patients with inflammatory and non-inflammatory neurological diseases. *J Neuroimmunol* 2012; 249(1-2):86-92.

Okuda Y, Okuda M, Apatoff BR, Posnett DN. The activation of memory CD4$^+$ T-cells and CD8$^+$ T-cells in patients with multiple sclerosis . *J Neurol Sci* 2005; 235(1-2):11-7.

Ono SJ, Liou HC, Davidon R, Strominger JL et al. Human X-box-binding protein 1 is required for the transcription of a subset of human class II major histocompatibility genes and forms a heterodimer with c-fos. *Proc Natl Acad Sci U S A* 1991; 88(10):4309-12.

Pappalardo JL, Zhang L, Pecsok MK, Perlman K et al. Transcriptomic and clonal characterization of T cells in the human central nervous system. *Sci Immunol* 2020; 5:eabb8786.

Raj T, Rothamel K, Mostafavi S, Ye C et al. Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. *Science* 2014; 344:519-23.

Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010; 26(1):139-40.

Sallusto F, Lenig D, Mackay CR and Lanzavecchia A. Flexible programs of chemokine receptor expression on human polarized T helper 1 and 2 lymphocytes. *J Exp Med* 1998; 187(6):875-83.

Schafflick D, Xu CA, Hartlehnert M, Cole M et al. Integrated single cell analysis of blood and cerebrospinal fluid leukocytes in multiple sclerosis. *Nat Commun* 2020; 11:247.
Smith-Garvin JE, Koretzky GA, Jordan MS. T cell activation. [Review]. *Ann Rev Immunol* 2009; 27:591-619.

Sousa APA, Johnson KR, Nicholas R, Darko S et al. Intrathecal T-cell clonal expansions in patients with multiple sclerosis. *Ann Clin Trans Neurol* 2016; 3(6):422-33.

Suzuki K, Hara M, Kitani A, Harigai M et al. Augmentation of LDL receptor activities on lymphocytes by interleukin-2 and anti-CD3 antibody: a flow cytometric analysis. *Biochim Biophys Acta* 1990; 1042(3):352-8.

Svenningsson A, Hansson GK, Andersen O, Andersson R et al. Adhesion molecule expression on cerebrospinal fluid T lymphocytes: Evidence for common recruitment mechanisms in multiple sclerosis, aseptic meningitis and normal controls. *Ann Neurol* 1993; 34:155-61.

Twiner MJ, Ryan JC, Morey JS, Smit KJ et al. Transcriptional profiling and inhibition of cholesterol biosynthesis in human T lymphocyte cells by the marine toxin azaspiracid. *Genomics* 2008; 91:289-300.

Varrin-Doyer M, Vincent P, Cavagna S, Auvergnon N et al. Phosphorylation of collapsing response mediator protein 2 on Tyr-479 regulates CXCL12-induced T lymphocyte migration. *J Biol Chem* 2009; 284:13265-76.

Veroni C, Marnetto F, Granieri L, Bertolotto A et al. Immune and Epstein-Barr virus gene expression in cerebrospinal fluid and peripheral blood mononuclear cells from patients with relapsing-remitting multiple sclerosis. *J Neuroinflamm* 2015; 12:132.

Vretham M, Dahle C, Ekerflt C, Forsberg P et al. CD4 and CD8 lymphocyte subsets in cerebrospinal fluid and peripheral blood from patients with multiple sclerosis, meningitis and normal controls. *Acta Neurol Scand* 1998; 97(4):215-20.