




# Expression profiling of cerebrospinal fluid identifies dysregulated antiviral mechanisms in multiple sclerosis

 Maria Ban,<sup>1,†</sup> Danila Bredikhin,<sup>2,3,†</sup> Yuanhua Huang,<sup>1,4,†,‡</sup> Marc Jan Bonder,<sup>2,3,§</sup> Kania Katarzyna,<sup>5</sup> Amanda J. Oliver,<sup>6</sup> Nicola K. Wilson,<sup>7,8</sup> Paul Coupland,<sup>5</sup> James Hadfield,<sup>5</sup> Berthold Göttgens,<sup>7,8</sup> Elo Madisson,<sup>4,6</sup> Oliver Stegle<sup>2,3,4,†</sup> and Stephen Sawcer<sup>1,†</sup>

<sup>†</sup>These authors contributed equally to this work.

See Schwab (<https://doi.org/10.1093/brain/awae006>) for a scientific commentary on this article.

Despite the overwhelming evidence that multiple sclerosis is an autoimmune disease, relatively little is known about the precise nature of the immune dysregulation underlying the development of the disease.

Reasoning that the CSF from patients might be enriched for cells relevant in pathogenesis, we have completed a high-resolution single-cell analysis of 96 732 CSF cells collected from 33 patients with multiple sclerosis ( $n = 48\ 675$ ) and 48 patients with other neurological diseases ( $n = 48\ 057$ ).

Completing comprehensive cell type annotation, we identified a rare population of CD8+ T cells, characterized by the upregulation of inhibitory receptors, increased in patients with multiple sclerosis. Applying a Multi-Omics Factor Analysis to these single-cell data further revealed that activity in pathways responsible for controlling inflammatory and type 1 interferon responses are altered in multiple sclerosis in both T cells and myeloid cells. We also undertook a systematic search for expression quantitative trait loci in the CSF cells. Of particular interest were two expression quantitative trait loci in CD8+ T cells that were fine mapped to multiple sclerosis susceptibility variants in the viral control genes *ZC3HAV1* (rs10271373) and *IFITM2* (rs1059091). Further analysis suggests that these associations likely reflect genetic effects on RNA splicing and cell-type specific gene expression respectively.

Collectively, our study suggests that alterations in viral control mechanisms might be important in the development of multiple sclerosis.

- 1 Department of Clinical Neurosciences, University of Cambridge, Cambridge CB2 0QQ, UK
- 2 European Molecular Biology Laboratory, Genome Biology Unit, 69117 Heidelberg, Germany
- 3 Division of Computational Genomics and Systems Genetics, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany
- 4 European Molecular Biology Laboratory, European Bioinformatics Institute, Cambridge CB10 1SD, UK
- 5 University of Cambridge, CRUK Cambridge Institute, Cambridge CB2 0RE, UK
- 6 Wellcome Sanger Institute, Wellcome Genome Campus, Cambridge CB10 1SA, UK
- 7 Department of Haematology, University of Cambridge, Cambridge CB2 0AW, UK
- 8 Wellcome-Medical Research Council Cambridge Stem Cell Institute, University of Cambridge, Cambridge CB2 0AW, UK

‡ Present address: School of Biomedical Sciences, University of Hong Kong, Hong Kong SAR, China

§ Present address: Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

Received August 04, 2023. Revised November 06, 2023. Accepted November 18, 2023. Advance access publication December 1, 2023

© The Author(s) 2023. Published by Oxford University Press on behalf of the Guarantors of Brain.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Correspondence to: Maria Ban  
Department of Clinical Neurosciences  
University of Cambridge, Box 165  
Cambridge Biomedical Campus  
Hills Road, Cambridge CB2 0QQ, UK  
E-mail: mb531@medschl.cam.ac.uk

**Keywords:** multiple sclerosis; scRNA-seq; eQTL; CSF; ZC3HAV1; IFITM2

## Introduction

Multiple sclerosis is a chronic autoimmune disease of the CNS which is pathologically characterized by patchy inflammatory demyelination and progressive neurodegeneration. Genome-wide association studies (GWAS) have identified over 200 susceptibility variants, the majority of which map to regulatory regions markedly enriched for genes with immunological functions,<sup>1</sup> collectively suggesting a role for dysregulation in immune cell pathways in disease susceptibility. To date though, the analysis of circulating peripheral immune cells has provided limited insights into the aetiology of multiple sclerosis. Given the immune privilege of the CNS, we reasoned that the immune cells in the CSF are likely to be enriched for cells of relevance to multiple sclerosis and might therefore provide novel insights that might otherwise be obscured by non-multiple sclerosis relevant parts of the circulating immune system.

Abnormalities in the CSF, including an elevated lymphocyte count and the presence of oligoclonal immunoglobulin bands indicative of chronic immune activation, are hallmark features of the disease and are used to aid diagnosis.<sup>2</sup> Immunophenotyping of CSF cells using flow cytometry has shown that patients with multiple sclerosis have increased proportions of B and plasma cells and decreased proportions of monocytes compared to healthy individuals or individuals with other neurological diseases.<sup>3,4</sup> Furthermore, it has also been shown that cells within the CSF of multiple sclerosis patients are predominantly of a memory phenotype and have a more activated profile.<sup>4</sup> Single-cell RNA sequencing (scRNA-seq) studies have corroborated these early findings and also refined specific CD4<sup>+</sup> T cell subsets that have increased frequencies in patients with multiple sclerosis including T regulatory cells and T follicular helper cells, and have further proposed novel cell subtypes as potential key players in multiple sclerosis pathogenesis which require further investigation.<sup>5–8</sup> However, these studies have lacked the power to explore the effects of known multiple sclerosis susceptibility variants in the CSF.

Here we use scRNA-seq to provide a high-resolution examination of CSF cell expression in 81 samples collected from patients with multiple sclerosis ( $n = 33$ ) and from patients with other neurological conditions ( $n = 48$ ). Following cell type annotation, we identified several cell-type-specific regulatory programs that correlated with disease status. Furthermore, we also undertook the first CSF-based single-cell expression quantitative trait loci (eQTL) analysis, which identified eQTL effects for two multiple sclerosis susceptibility variants in CD8<sup>+</sup> T cells, the risk allele rs10271373\_A increasing expression of the zinc finger CCCH-type antiviral protein 1 (ZC3HAV1) and rs1059091\_A reducing the expression of interferon-induced transmembrane protein 2 (IFITM2). Overall, our observations suggest a possible dysregulation in the control of viral responses in multiple sclerosis patients.

## Materials and methods

### Patient recruitment

Our study was approved by the National Research Ethics Committee (Service South Central—Berkshire; 15/SC/0087) and all study subjects gave valid fully informed written consent. In total, 88 subjects were recruited from the Cambridge University Hospital (Addenbrooke's) neurosciences department. Patients were recruited from amongst those attending the department for a lumbar puncture as part of the investigation and treatment of their neurological condition. For seven of the recruited patients (three multiple sclerosis; two idiopathic intracranial hypertension; one non-inflammatory neurological disease; and one other inflammatory neurological disease), the CSF sample was unusable for technical reasons. Multiple sclerosis was diagnosed in line with standard clinical criteria.<sup>9</sup> All but two of our multiple sclerosis patients were newly diagnosed and therefore not on disease modifying treatment at the time of sampling. The two multiple sclerosis patients with an established diagnosis were both on treatment with natalizumab (Tysabri®). Disease activity in multiple sclerosis was determined by the primary clinical team on the basis of clinical and radiological features; the disease being judged to be inactive, active or highly active. [Supplementary Table 1](#) shows the demographics, clinical and laboratory hospital measured CSF features for the 81 study subjects with RNA sequencing data that passed quality control and were included in the analysis. In the UK, lumbar puncture is not routinely performed in the investigation of suspected multiple sclerosis and therefore tends to be employed in the context of less typical cases. This selection of subjects for the investigation likely explains the older than average age at diagnosis of our multiple sclerosis cases, and the absence of the usual excess of females. The other inflammatory neurological diseases (OIND) include chronic inflammatory demyelinating polyneuropathy, neurosarcoidosis and transverse myelitis, while the non-inflammatory neurological diseases (NIND) include cerebrovascular disease, functional neurological disorder, low-grade glioma, migraine, myelopathy, neuropathy, seizure, trigeminal neuralgia and idiopathic intracranial hypertension (IIH)—the latter were considered as a separate group as their CSF would not be expected to differ from normality other than in terms of pressure.

### Cell isolation and single cell library generation

All samples reached the lab within 1 hour of collection and were either processed immediately or frozen for later pooling. To each sample an equivalent volume of X-VIVO™ 10 Serum-free Hematopoietic Cell Medium (Lonza) was added to maintain cell viability. Owing to the low concentration of cells in the CSF, each sample was first concentrated by centrifugation at 300g for 10 min, the supernatant removed leaving approximately a 200 µl volume.

A manual cell count was completed on the concentrated sample using a Neubauer haemocytometer and cell viability assessed using Trypan Blue exclusion dye. The cell suspension was then centrifuged at 300g for another 10 min to further concentrate the sample to a final volume of 32 µl ready to be loaded onto a 10x Chromium Single Cell Controller using the Chromium Single Cell Gene Expression 3' v2 kit (10x Genomics). Sequencing was completed on an Illumina HiSeq 4000 at the Cancer Research Institute, Cambridge University sequencing hub using locally optimized protocols and settings.

### Genotype imputation

All 81 donors were genotyped using the Illumina Global Screening Array. Imputation was performed against the Haplotype Reference Consortium (HRC)<sup>10</sup> data using EAGLE2 through the Sanger Imputation server (<https://imputation.sanger.ac.uk>). Genotypes were converted to hard call binary plink format using GenotypeHarmonizer,<sup>11</sup> before performing eQTL mapping. During the conversion we removed genotype calls with a posterior probability below 0.4, single nucleotide polymorphisms with a MACH  $R^2 < 0.6$  and/or a Hardy–Weinberg equilibrium below  $1.0 \times 10^{-4}$ .

### Sequencing data preprocessing

The raw scRNA-seq reads in FASTQ format from each sequencing run were processed using Cell Ranger v3.1 (10x Genomics) and aligned to the human genome (Homo\_sapiens.GRCh38.dna.primary\_assembly.fa) using the default Cell Ranger settings and annotated using Homo\_sapiens.GRCh38.93.filtered.gtf. Cells were called using both Cell Ranger and emptyDrops,<sup>12</sup> with the union of cells called used for all downstream analysis. Among the 71 sequencing runs, five of these sequencing runs were multiplexed with two to four donors per run, in total 15 donors were included in multiplexed runs. To demultiplex these runs, we first used cellSNP v0.1.7<sup>13</sup> to genotype each cell, followed by Vireo v0.2.1<sup>14</sup> to assign cells to each multiplexed individual based on known genotypes. We used two separate strategies to detect doublets. First, for each multiplexed sequencing run, Vireo<sup>14</sup> identifies the cross-donor doublets through their mixed genotypes. Second, we also detected cross-cell type doublets from the transcriptome profiles on each sequencing run by a consensus method combining doubletFinder<sup>15</sup> and Scrublet.<sup>16</sup>

### Cell type annotation

The raw count matrix was corrected for ambient RNA using the soupX tool<sup>17</sup> in R using standard settings. Count matrix normalization and log-transformation, clustering and cell type annotation were performed using the package scanpy<sup>18</sup> in Python. Batch effects were reduced by using the BBKNN tool.<sup>19</sup> Clusters were annotated to cell types according to known canonical marker genes. A sequential clustering method was then used for finer annotation of larger clusters. Clusters where there were markers from multiple lineages indicating contamination, markers of dividing cells, low unique molecular identifier (UMI) counts or high mitochondrial percentage were all removed prior to downstream analysis. Cell types were annotated to the finest achievable level. Related clusters were later combined into a higher level annotation that we used for downstream analysis that required a larger number of cells for each analysis. Differential cell type composition between groups was analysed using the DCATS R package (v0.99.7).

### Factor decomposition using multi-omics factor analysis

Multi-omics factor analysis (MOFA+)<sup>20</sup> was applied to the normalized counts of highly-variable genes provided with cell type labels and set to learn 30 factors. Gene enrichment analysis was performed as implemented in the MOFA2 R package using the C5 ontology gene sets from the MSIGDB database.<sup>21,22</sup> Factor weights and the DoRothEA database of transcription factor targets<sup>23</sup> were then used to identify transcription factor regulons enriched in each factor. Factor weights for transcription factor targets were aggregated per factor and per transcription factor, which can be represented as a matrix product of factor weights and binary transcription factor—gene target relations. In order to test for the significance of the associations, for each MOFA+ factor and transcription factor, the same number of targets was sampled randomly 10 000 times in order to calculate an empirical *P*-value.

### Viral genome alignment

In this study, we performed a cell-resolution viral genome alignment. First, we compiled a list of 833 viral sequences, containing 762 viruses collected from VirTect,<sup>24</sup> and seven other viruses and 64 consensus sequences of human endogenous retroviruses (HERVs) collected by Vargiu et al.<sup>25</sup> Then for each of our scRNA-seq batches, we extracted the reads that were uncounted or unmapped for the human transcriptome in the CellRanger step by using 'samtools view possorted\_genome\_bam.bam | grep -v 'xf:i:25'>xf\_non25.sam'. Next, we aligned these reads to the 833 viral sequences via STARSolo<sup>26</sup> using CellRanger's bam file as input and assigned each read to a specific cell using the cell-associated barcode list for each sample (see the 'Sequencing data preprocessing' section) as a whitelist. A cell-by-virus UMI count matrix is returned and by concatenating all batches we can perform downstream analysis. For reproducibility and public reuse, we have packaged all these analysis steps at <https://github.com/huangyh09/ViralScan>. To compare the donor-level prevalence between multiple sclerosis and non-multiple sclerosis samples, we performed a logistic regression to test if the coefficient of 'is\_MS' is significantly non-zero with a model:  $\text{Virus\_present} \sim \text{is\_MS} + \text{total\_UMIs} + \text{intercept}$ . To compare the cell-level prevalence between multiple sclerosis and non-multiple sclerosis samples, we performed a linear regression to test if the coefficient of 'is\_MS' is significantly non-zero with a model:  $\text{viral\_proportion} \sim \text{is\_MS} + \text{total\_UMIs\_log} + \text{total\_cells\_log} + \text{intercept}$ .

### Expression quantitative trait loci mapping

For eQTL mapping, we followed the best practices for single-cell eQTL mapping outlined by Cuomo et al.<sup>27</sup> Specifically, we did single cell-based normalization using sctransform<sup>28</sup> and afterwards mean aggregated the cells to donor level gene-expression. The eQTLs were mapped per cell type using a linear mixed model implemented in LIMIX v2 (<https://github.com/limix/limix>). As outlined in Cuomo et al.,<sup>27</sup> we leveraged two random effects, one to account for population structure and one for donor wise read-depth (encoded as 1/*n* reads per donor). In the analysis of each cell type, donors with fewer than two cells of that type were dropped, making the number of donors, and therefore power, different per cell type. Owing to the limited sample size of our study, we chose to only map eQTLs for protein coding genes that had non-zero expression in at least 10% of the donors and chose to filter SNPs on a minor

allele frequency of 10% and used a cis-eQTL window of 100 kb either side of the gene.

Multiple testing correction was performed as in Cuomo *et al.*<sup>27</sup> Specifically, we used 1000 permutations per QTL, and first correct per gene for testing multiple variants per gene followed by two different approaches to correct for the number of genes per eQTL map. For genes included in the eQTLGen<sup>29</sup> set, we applied the conditional false discovery rate (cFDR) approach (conditioning on the eQTLGen summary statistics); for the remaining genes, we used the Storey Q value approach. FDR was controlled at FDR 10% and cFDR 10%. Effects were considered to be replicated in other tissues when the direction of effect was concordant and there was nominal statistical significance in the second tissue.

For fine-mapping, we applied SusieR v0.11.92<sup>30</sup> using the default settings in each cell type separately following the guidelines provided by the authors. We annotated the eQTL signals based on fine-mapping results and specifically annotated signals overlapping with multiple sclerosis GWAS variants.

### Detection of allele specific expression

To perform cell type-specific allele specific expression (ASE), we first genotyped each individual cell with cell-snp-lite v1.2.0<sup>13</sup> and aggregated expression data for each cell type to generate pseudo-bulk level data. We focused on the 8046 SNPs across 512 cis-genes of multiple sclerosis disease variants. Where there were multiple coding SNPs in a gene, we only considered the one with the maximum number of donors with both heterozygous genotypes and scRNA-seq data for that gene. Then for each SNP in each cell type, we use all informative donors to perform a beta-binomial regression with AOD R package (v1.3.1) and test if the mean allele frequency is 0.5, equivalent to testing if the intercept of the regression is 0. The *P*-value calculated with the Wald test is obtained.

### Analysis of splicing and transcript usage

We first simplified the transcript annotation of the gene *ZC3HAV1* into two major transcripts with distinct exons in the 3' end. Then using BRIE v2.2.0<sup>31</sup> we quantified the transcript compatible counts for each individual cell with brie-count using default parameters. A beta-binomial regression (from AOD R package) was used to calculate the *P*-value to test if there is a genetic effect on the proportion of transcript usage.

## Results

### The cellular architecture of CSF

We undertook single-cell transcriptome profiling of CSF cells from 33 multiple sclerosis patients and 48 with other neurological conditions using the 10x Genomics 3' v2 platform (Fig. 1A and Supplementary Table 1). Following quality control, doublet removal and computational adjustment of ambient RNA (see the 'Materials and methods' section), a total of 96 732 cells were available for downstream analysis. On average, we captured 1194 quality control (qc)-passing cells per patient (range 21–10 491 cells per patient; Supplementary Table 2); 48 675 cells from patients with multiple sclerosis, 18 524 from patients with OIND, 13 847 from patients with NIND and 15 686 from patients with IIH.

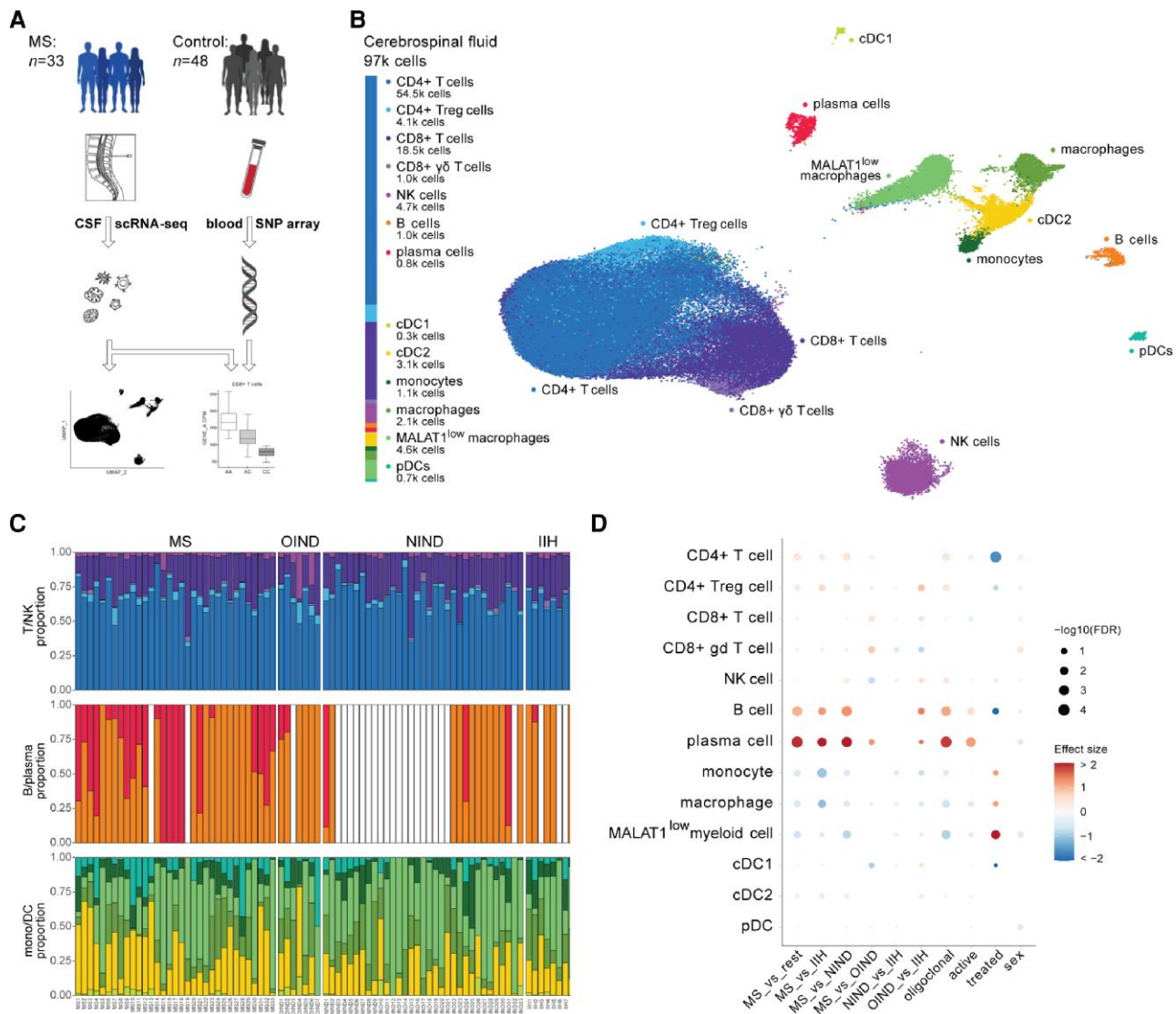
Through dimensionality reduction, iterative unsupervised clustering and inspection of canonical marker gene expression, we annotated cells into 13 major immune populations which were used

for all downstream analysis (Fig. 1B; see the 'Materials and methods' section). Most of our multiple sclerosis patients (*n* = 31) were treatment-naïve, the other two patients receiving treatment with Tysabri, a monoclonal antibody against  $\alpha$ 4-integrin which reduces the migration of leucocytes through the blood–brain barrier.<sup>32</sup> This treatment is known to alter cell type composition in the CSF<sup>7</sup>; therefore, the Tysabri-treated patients were excluded from the cell type proportion analysis. Comparison of the immune cell type proportions between disease groups was concordant with previous findings,<sup>3–7</sup> with multiple sclerosis patients having an overrepresentation of plasma cells, and an underrepresentation of monocytes and macrophages in comparison to patients with IIH and NIND (Fig. 1C and D). An increase in B cells and T regulatory cells was also seen in multiple sclerosis patients when compared to IIH and NIND but not OIND patients. Next, we assessed the relationship between cell type proportions within multiple sclerosis patients based on the clinical annotations of these samples (Fig. 1D). B and plasma cells were enriched in oligoclonal-positive multiple sclerosis patients and those with active disease. Whereas the Tysabri-treated patients showed a reduction of CD4+ T cell and B cell proportion when compared to treatment-naïve multiple sclerosis patients, correlating with previous findings of altered immune cell composition in the CSF of Tysabri-treated patients.<sup>7</sup> A high proportion of macrophage-like cells characterized by very low expression of the long non-coding genes *MALAT1* and *NEAT1* was also observed in the Tysabri-treated multiple sclerosis patients (Fig. 1D). Published data would suggest that this population could be immunosuppressive as knockdown of *MALAT1* has been shown to reduce inflammatory injury following lung transplant through inhibiting the neutrophil infiltration and activation,<sup>33</sup> while *NEAT1* knockdown was shown to lead to increased viral loads.<sup>34</sup>

Our higher resolution annotation identified many of the known low frequency cell subsets, including AXL+ dendritic cells, ACY3+ dendritic cells and SPP1+ macrophages (Supplementary Figs 1 and 2), and further identified a less well defined rare population of CD8+ T cells expressing cytotoxic markers (*GZMA*, *GZMK*) and characterized by upregulation of the co-stimulatory marker *CD27*, upregulation of inhibitory receptors (*HAVCR2*, *TIGIT*) and downregulation of chemokines (*CCL4*, *CCL5*). This uncommon cell subtype was enriched in multiple sclerosis patients (Supplementary Fig. 3), however given the low frequency of these cells additional data will be required to confirm these initial findings. The patient specific high resolution cell type proportions are provided in Supplementary Table 2.

### Differential regulation of inflammatory and type I IFN gene expression programs in multiple sclerosis

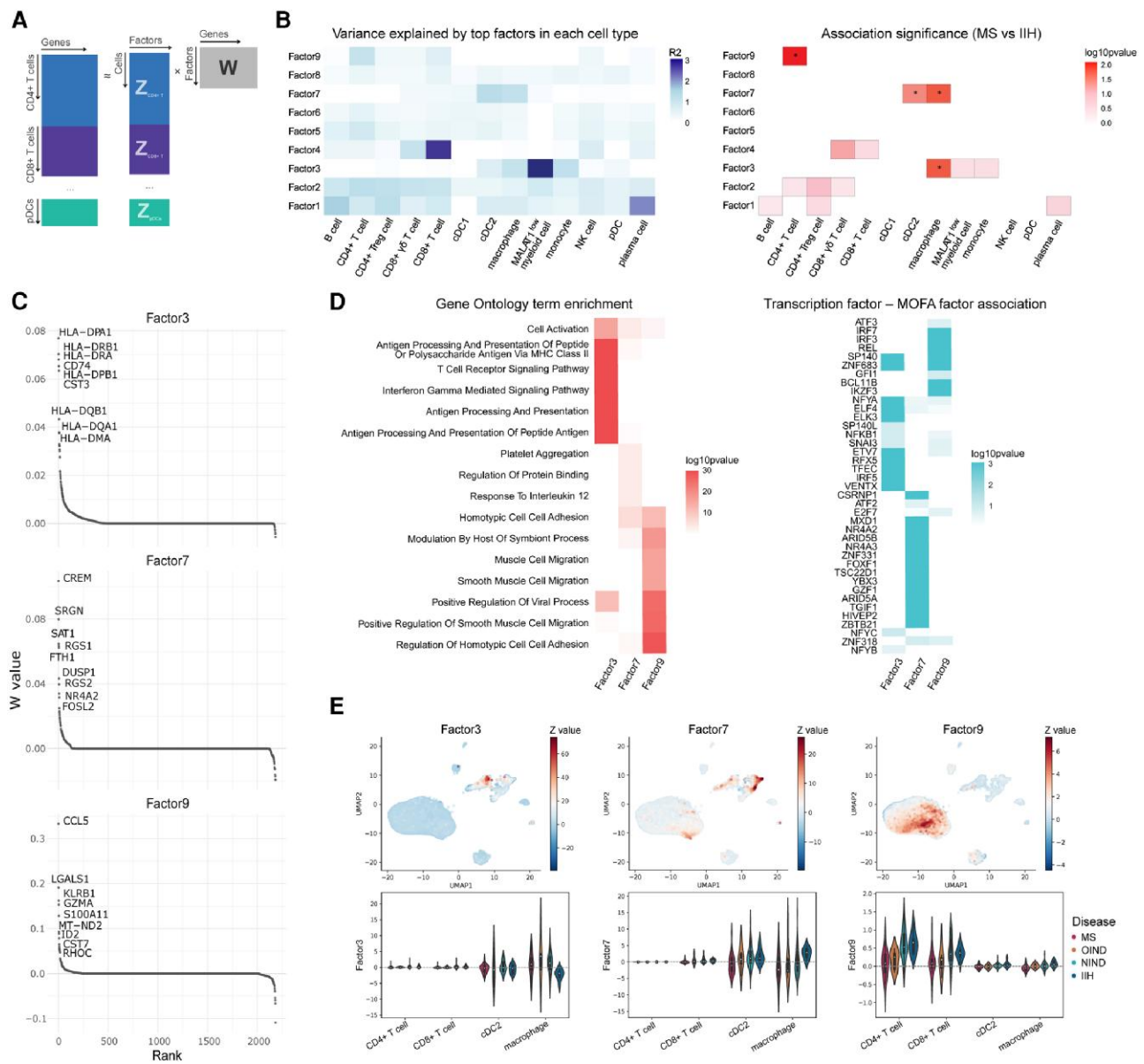
Next, we set out to characterize cell type-specific differences in function between multiple sclerosis cases and controls as reflected in gene expression. Given the limited power of conventional cell type-specific differential expression analysis of individuals we applied MOFA+<sup>20,35</sup> to increase statistical power. This approach identifies sets of co-regulated genes (factors) that contribute substantially to gene expression variation within different cell types (Fig. 2A and Supplementary Table 3; see the 'Materials and methods' section). MOFA+ identified nine major factors, of which we focused on the 14 factor-cell type pairs that explained at least 1% of the variance in expression, and tested these for association with multiple sclerosis using the IIH patients as controls (see the 'Materials and methods' section). At a FDR of <10%, using linear



**Figure 1** Study overview and characterization of cellular architecture of CSF. (A) Illustration of the experimental design and sampling scheme, in which droplet-based single cell sequencing was performed in 33 multiple sclerosis patients and 48 controls using the 10x Chromium Single Cell 3' Solution V2. (B) Uniform manifold approximation and projection (UMAP) representation of single-cell transcriptomes of 96 732 CSF cells collected from all 81 donors, with the colours denoting the major cell types defined by iterative clustering and the expression of canonical marker genes. (C) Cell type proportions in three groups of cells across samples [T/natural killer (NK) at the top, B/plasma in the middle and monocytes/dendritic cells (DCs) at the bottom], ordered by sample group [multiple sclerosis (MS), other inflammatory neurological disease (OIND), non-inflammatory neurological disease (NIND) and idiopathic intracranial hypertension (IIH)]. Vertical bars correspond to individual samples. White indicates missing information (no cells of that type sampled in that individual). The colours correspond to the major cell types identified in C. (D) Differential cell abundance in pairwise comparisons between: MS versus non-MS; MS versus IIH; MS versus NIND; MS versus OIND; NIND versus IIH; OIND versus IIH; oligoclonal positive MS versus oligoclonal negative MS; active MS versus inactive MS; Tysabri-treated MS versus non-treated MS; female MS versus male MS patients. Dot colour denotes effect size estimates and dot size denotes the significance of the respective comparison. DC = dendritic cell; cDC1 = myeloid/conventional DC1; cDC2 = myeloid/conventional DC2; pDC = plasmacytoid DC; SNP = single nucleotide polymorphism.

regression and permutation-based statistics, we identified three multiple sclerosis-associated factors: factor 9 in CD4+ T cells, factor 3 in macrophages and factor 7 in macrophages and dendritic cells (Fig. 2B), the top-weighted genes for each factor are shown in Fig. 2C. To annotate these factors, we considered two complementary strategies. First, we performed a gene set enrichment analysis (GSEA) for each factor, weighting genes by the absolute value of their factor weights. Second, we searched for transcription factors (TFs) with regulons significantly enriched for genes with high factor weights (Fig. 2D), hypothesizing that these TFs might drive the co-regulated expression.<sup>36</sup>

The most statistically significant association with disease was observed for factor 9 in CD4+ T cells ( $P = 0.007$ ; permutation test at the donor level), where patients with multiple sclerosis had reduced expression compared to IIH (Fig. 2E). The genes with the highest weights for this factor, including *CCL5*, *LGALS1*, *S100A11*, *GZMA* and *GZMK* (Fig. 2C), are characteristic of tissue-resident cytotoxic T effector memory cells<sup>6,37</sup> and overlap with a previously described CD4+ T effector memory population re-expressing CD45RA ( $T_{EMRA}$ ).<sup>38</sup> For this factor, GSEA shows significant enrichment for genes involved in the regulation of homotypic cell-cell adhesion (GO: 0034110; Fig. 2D). In T cells, homotypic cell interactions lead



**Figure 2** Differential activity of gene expression programmes in multiple sclerosis versus controls. (A) Illustration of the cell-type aware group-matrix factorization considered using multi-omics factor analysis (MOFA+). Expression profiles for different cell types are factorized into factors (Z) and weight matrices (W), respectively. (B) Gene expression variance explained by individual factors across cell types (R<sup>2</sup>) (left) and association with disease groups [multiple sclerosis versus idiopathic intracranial hypertension (IIH); linear model] (right). This analysis identifies factor 3 as primarily capturing disease-associated variation in macrophages, factor 7 as explaining variation in cDC2 and macrophages and factor 9 as explaining variation in CD4+ T cells. (C) Top gene loadings of the three disease-associated factors. (D) Left: Gene set enrichment based on factor loadings for the corresponding factors as in C. Right: A complementary strategy for factor annotations based on enrichments with known transcription factor targets. (E) Analysis of factor relevance across cells. Top: Factor activity across cells on the global uniform manifold approximation and projection (UMAP) representation coloured by the factor value weight; bottom: violin plots showing factor values averaged per donor show disease association for individual cell types. DC = dendritic cell; cDC1 = myeloid/conventional DC1; cDC2 = myeloid/conventional DC2; pDC = plasmacytoid DC; IIH = idiopathic intracranial hypertension; MHC = major histocompatibility complex; MS = multiple sclerosis; NK = natural killer cell; NIND = non-inflammatory neurological disease; OIND = other inflammatory neurological disease.

to the release of IL4, IL-10 and TGFβ and the generation of suppressor T cells, providing a negative feedback mechanism that controls immune responses.<sup>39</sup> Several TFs involved in attenuating inflammatory and anti-viral responses have regulons overlapping with the genes showing the highest absolute weightings for factor 9 (Fig. 2D), including BCL11B,<sup>40,41</sup> REL,<sup>42</sup> IKZF3<sup>43</sup> and IRF3/IRF7.<sup>44</sup> The genes with the highest absolute value of weighing for factor 7 are characteristic of anti-inflammatory M2 macrophages, including

CREM,<sup>45</sup> FTH1<sup>46</sup> and NR4A2,<sup>47</sup> and show reduced expression in multiple sclerosis. For factor 3, the genes with the highest absolute value weightings were upregulated in multiple sclerosis in macrophages and include a range of HLA genes and showed GSEA for antigen processing and presentation of peptide or polysaccharide antigen via MHC class II (GO:0002504) (Fig. 2D) as well as significant overlap with the regulon of VENTX which is involved in macrophage differentiation and pro-inflammatory activity.<sup>48</sup>

The regulons from TFs *IRF5* and *ETV7*, which are involved in type I IFN response regulation and are associated with M1 macrophage polarization,<sup>49,50</sup> showed significant overlap for factor 3.

### Alignment of viral genomes is not enriched in patients with multiple sclerosis

Viruses, particularly the Epstein–Barr virus (EBV), have long been implicated in the pathogenesis of multiple sclerosis,<sup>51</sup> and in comparison to healthy individuals the CSF from patients with multiple sclerosis more frequently contains polyspecific immunoglobulins directed against a range of common viral agents; the so-called Measles Rubella Zoster reaction.<sup>52</sup> It therefore seemed logical to examine whether there was evidence of viral genomes in our CSF samples. To do this, we aligned the scRNA-seq reads that were either unmapped to the human transcriptome or were within unannotated regions in our data to 833 viral genomes and calculated the virus-specific UMI count per cell. We detected alignment (defined by at least one UMI detected across any cell type) for 78 of these viral genomes (Fig. 3A). The proportion of cells containing at least one viral UMI of any viral sequence showed no statistically significant difference between participants with multiple sclerosis compared to those without multiple sclerosis (Fig. 3B). Focusing on the specific viral genomes that were prevalent in more than 5% of either the multiple sclerosis or non-multiple sclerosis cohort (Supplementary Table 4 and Supplementary Fig. 4), the vast majority of these 65 well-represented viruses belong to the human endogenous retrovirus (HERV) family, transposable elements derived from retroviral integration into the human genome.<sup>53</sup> Sequences of retroviral origin are considered to constitute up to 8% of the human genome.<sup>54</sup> Despite being viewed as of little functional importance, HERVs are widely expressed in normal tissue particularly in the brain and testis<sup>55</sup> and implicated in diseases including multiple sclerosis.<sup>56</sup> While we observed an increased prevalence of a few HERVs in multiple sclerosis patients (Supplementary Table 4), this increase was not statistically significant (FDR > 10%). Outside of HERVs, the only other viruses that we identified in more than 5% of either the multiple sclerosis or non-multiple sclerosis cohort were the human herpesvirus 1, 2, 6A and 6B, human papillomavirus 71 and Abelson murine leukemia virus. While these were all more prevalent in multiple sclerosis, again this increase was not statistically significant (Supplementary Fig. 4 and Supplementary Table 4).

### Expression quantitative trait loci mapping in CSF cell types

To determine the impact of genetic variation on gene expression in the annotated cell types, we also completed array-based genome-wide genotyping of all individuals included in the study. Following imputation using the Haplotype Reference Consortium<sup>10</sup> and quality control, we retained 9 149 816 common germline variants (minor allele frequency > 10% and MACH R2: 0.6; see the ‘Materials and methods’ section; Supplementary Fig. 5). We completed a genome-wide survey of eQTLs to uncover if there were any novel findings within cells isolated from the CSF. To do this, we applied the single-cell eQTL calling workflow proposed by Cuomo et al.<sup>27</sup> Briefly, this approach maps cis-eQTLs in individual cell types by aggregating expression profiles at a donor level (considering between 6594 and 12 305 protein-coding genes per cell type and variants within 100 kb of each considered gene; see the ‘Materials and methods’ section). In CD4+ and CD8+ T cells,

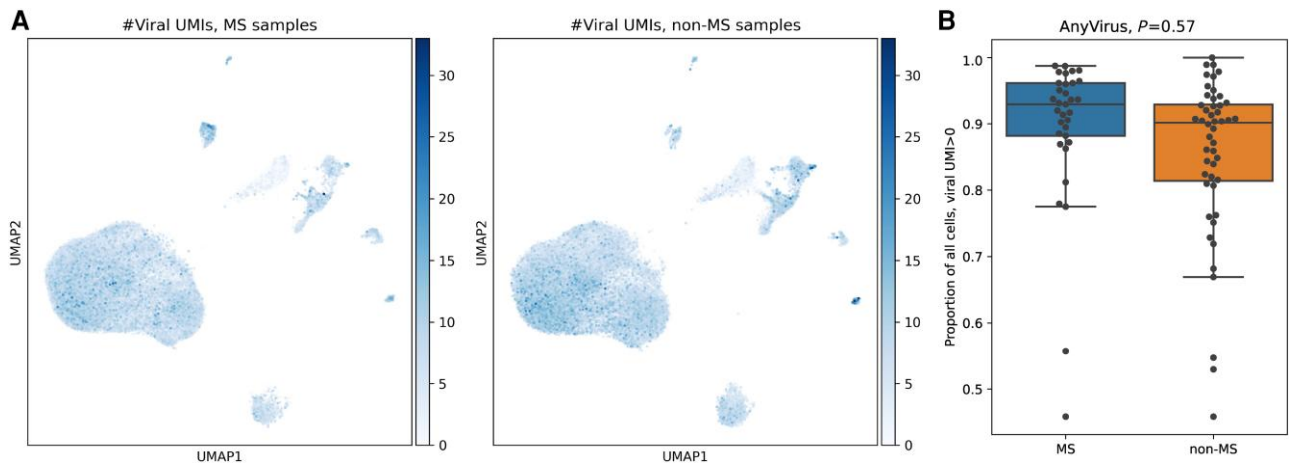
the most abundant cell types, we found 437 and 252 genes with a cis-eQTL, respectively, while for B cells, an uncommon cell type in the CSF, we only identified 25 such effects (Supplementary Tables 5 and 6). As the power for eQTL discovery is affected by the number of cells and individuals assayed, these results were not unexpected.

To assess the validity of our CSF eQTL signals, we sought to replicate these associations in relevant published cell-type specific single-cell<sup>57</sup> and bulk RNA-peripheral blood mononuclear cell<sup>58</sup> expression studies. As the cell types, genes and variants tested were not the same across these studies, we established the rate at which our CSF eQTLs were reported as eQTLs in either of these studies,<sup>57,58</sup> with eQTLs considered to be confirmed if there was the same direction of effect at nominal significance ( $P < 0.05$ ). Overall, 68%–100% of the CSF eQTLs we identified were confirmed in either of the published datasets (Fig. 4A and Supplementary Table 6). The highest replication rates were seen in T and natural killer cells (>82%), whereas cells derived from the myeloid lineage had a slightly lower replication rate (ranging from 69% to 79%), with the lowest replication seen for eQTLs in B cells (68%). Next, we looked for concordance of eQTL signals between CSF cell types (see the ‘Materials and methods’ section) and found that confirmation rates between cell types found in the CSF within our study ranged from 52% in B cells to 84% in CD4+ T regulatory cells (Supplementary Table 6). Given that eQTL effects are known to be influenced by the cell state and environment, we would not have expected complete concordance between blood and CSF or across the different cell types. In summary, the replication rate gives confidence that the eQTL signals described are genuine, despite the moderate sample size of our study.

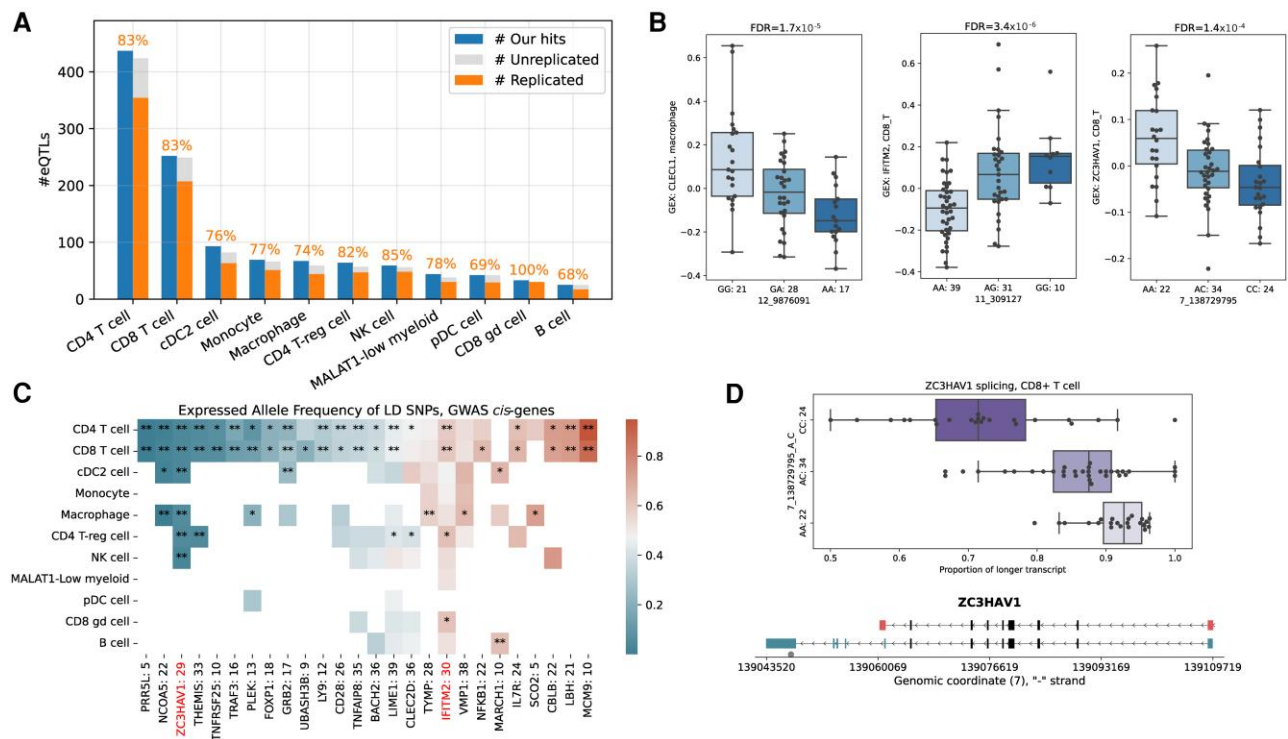
Next, we used fine-mapping to investigate the relationship between the eQTL identified and the 200 multiple sclerosis susceptibility variants identified from GWAS.<sup>1</sup> Using SusieR,<sup>30</sup> we identified credible sets of causal variants for 416 of the 749 cis-eQTLs (Supplementary Table 7; see the ‘Materials and methods’ section); these sets contained a median of 21 variants per eQTL (ranging from 1 to 336). Of these 416 fine-mapped cis-eQTLs, 17 mapped within 100 kb of a known multiple sclerosis risk factor. In three instances, the credible set for the eQTL included the known multiple sclerosis susceptibility variants: rs3764021 for the expression of *CLECL1* in macrophages; rs1059091 for the expression of *IFITM2* in CD8+ T cells; and rs10271373 for the expression of *ZC3HAV1* in CD8+ T cells (Fig. 4B). We further investigated whether there was correlation of the fine-mapped eQTLs with genetic variants that have recently been described as associated with multiple sclerosis progression.<sup>59</sup> Of the 12 progression variants, none mapped to within 100 kb of a fine-mapped cis-eQTL. This is perhaps not surprising, as enrichment analysis suggests the variants associated with progression are enriched in or near genes highly expressed in the brain which are not well captured in cells of the CSF.

### Allele-specific expression is consistent with expression quantitative trait loci mapping

To complement the eQTL analysis and increase our power to identify further cis-eQTL effects that are correlated with multiple sclerosis susceptibility or progression we completed an ASE analysis. No informative SNPs mapped close to the variants associated with multiple sclerosis progression [with linkage disequilibrium (LD)  $R^2 > 0.15$ ]; however, informative SNPs were available in 254 of the genes mapping close to multiple sclerosis susceptibility variants (with LD  $R^2 > 0.15$ ). Statistically significant evidence of ASE was seen for 40 of these genes in at least one cell type (Wald test in



**Figure 3** Viral transcript expression in the CSF. (A) Uniform manifold approximation and projection (UMAP) plots colour-coded for the total unique molecular identifier (UMI) counts per cell for sequence reads mapped to any of the 833 viral sequences analysed. *Left*: UMAP plot for the 48 675 cells from multiple sclerosis patients; *right*: UMAP plot for the 48 057 cells from all non-multiple sclerosis patients. (B) Proportion of cells that contain at least one viral UMI in multiple sclerosis patients versus controls, no statistically significant difference was found between the groups ( $P > 0.05$ ). MS = multiple sclerosis.



**Figure 4** Genetic effects on gene expression and co-localization with multiple sclerosis variants. (A) Number of genes with a *cis*-expression quantitative trait loci (eQTL) detected across 11 cell types (blue bar). For genes that could be probed for replication (see the 'Materials and methods' section), the number of eQTLs that were consistent (orange bar) or discordant (grey bar) with previous studies in either peripheral blood mononuclear cells<sup>58</sup> or cell type specific single cell data<sup>57</sup> is indicated. Values above the bar denote the replication rates. (B) Scatter plot of gene expression, stratified by eQTL variant status, for the three genes with a cell-type-specific *cis*-eQTL fine mapped to established multiple sclerosis susceptibility variants. (C) Heatmap indicating allele-specific expression of the minor allele for the 26 single nucleotide polymorphisms (SNPs) in high linkage disequilibrium (LD) with multiple sclerosis-associated variants in different cell types with red indicating higher expression of the minor allele. The number of donors with heterozygous genotypes is indicated next to the gene name. White indicates a missing value where the gene was expressed < 2 heterozygous donors. Single asterisk = false discovery rate (FDR) < 0.01; two asterisks = FDR < 0.0001 (Wald test, beta-binomial regression). (D) *Top*: Expression scatter plot showing the proportion of the long and short ZC3HAV1 transcripts according to rs10271373 genotype; *bottom*: graphical representation of the two main ZC3HAV1 transcripts. DC = dendritic cell; cDC2 = myeloid/conventional DC2; pDC = plasmacytoid DC; GWAS = genome-wide association study; NK = natural killer cell.



beta-binomial regression;  $FDR < 0.01$ ; [Supplementary Table 8](#)), including 26 genes that have ASE in at least one cell type other than CD4+ T cells ([Fig. 4C](#)). Interestingly, in keeping with the eQTL analysis, we found evidence of ASE for both ZC3HAV1 and IFITM2. In both genes the multiple sclerosis associated SNPs are located within transcribed regions of the gene [3' untranslated region (UTR) and missense variant, respectively] and so were directly assessable for ASE. For ZC3HAV1, while the most statistically significant effect was seen in CD8+ T cells (correlating with the eQTL analysis), ASE was detected across most cell types with the same direction of effect. Meanwhile the ASE for rs1059091 in IFITM2 was strongest in CD8+ T cells but also evident in CD4+ T cells alone, suggesting that the effects observed may be specific to these cell types. This aligns with publicly available data in peripheral blood mononuclear cells which demonstrate this SNP as an eQTL for IFITM2 in CD8+ naïve T cells.<sup>60</sup>

Interestingly, we noted that the pattern of extensive ASE within ZC3HAV1 was specifically localized to the 3'UTR of the longer ZC3HAV1 transcript (ENST00000464606.5), which contains the multiple sclerosis-associated SNP rs10271373 as well as seven other SNPs in high linkage disequilibrium which all showed ASE. Alternative polyadenylation of ZC3HAV1 results in multiple isoforms differing at the 3' end of the gene. To further characterize whether the observed ZC3HAV1 eQTL was a splice variant we used the BRIE software<sup>31</sup> to quantify the proportion of the captured isoforms of ZC3HAV1 (ENST00000464606.5/ENST00000680309.1 versus ENST00000471652.1). This analysis revealed that the eQTL influence of rs10271373 was mediated almost exclusively through an effect on the longer transcripts (ENST00000464606.5/ENST00000680309.1), which includes three additional exons (exon 10–12) and an alternate 3'UTR ([Fig. 4D](#)). The proportion of the longer transcript is significantly decreased from 91.3% in homozygotes for the multiple sclerosis risk allele to 73.8% in homozygotes for the protective allele ( $P < 1 \times 10^{-16}$ , Wald test in beta-binomial regression). The transcript usage does not differ between multiple sclerosis and non-multiple sclerosis samples in any cell type ( $P = 0.42$ ).

## Discussion

Here we report the results from our comprehensive single-cell resolution analysis of CSF from patients with multiple sclerosis and other neurological diseases. Our analysis confirms the well-established cellular composition of CSF in patients with multiple sclerosis and highlights inflammatory responses altered in multiple sclerosis. We further describe the first single-cell eQTL analysis of CSF and identify two eQTLs in CD8+ T cells correlated with multiple sclerosis susceptibility, both in genes related to controlling viral responses—ZC3HAV1 and IFITM2.

Our eQTL analysis indicates that the multiple sclerosis susceptibility variant rs10271373 influences the expression of ZC3HAV1 in CD8+ T cells, a gene with four isoforms,<sup>61</sup> two of which, ZAP-L and ZAP-S,<sup>62</sup> are suggested to be important in resistance to particular viruses.<sup>62–65</sup> This eQTL effect appears to be driven by an underlying splicing QTL, with increased usage of the ZAP-L isoform in individuals that carry the multiple sclerosis risk allele (rs10271373\_A). This longer isoform has an extended C-terminal region which contains a poly (ADP ribose) polymerase (PARP)-like domain<sup>62</sup> and contains a CaaX prenylation motif. Farnesyl modification of the cysteine residue of this motif enhances the antiviral activity through altering the subcellular location of ZAP-L to the endocytic and lysosomal compartments for viral interaction.<sup>66</sup>

Interestingly, this isoform is known to be under positive selection<sup>62</sup> and contains EBV miRNA binding sites within its 3'UTR.<sup>67</sup> We further identified a second eQTL that fine-mapped to a multiple sclerosis susceptibility variant (rs1059091) for another antiviral gene, IFITM2, where reduced expression of the gene in CD8+ T cells was observed in individuals carrying the risk allele (rs1059091\_A), supporting a previous finding that observed this eQTL in CD8+ naïve T cells only,<sup>60</sup> suggesting this eQTL is likely to be cell-type specific. IFITM2 is believed to inhibit viral entry, likely through effects on the endocytic pathway<sup>68</sup> and has also been implicated in Th2 T cell differentiation.<sup>69</sup> Given the location of IFITM2 within the IFITM cluster on chromosome 11, along with transcript-specific genotypic effects of rs1059091 on IFITM2 expression in response to particular viral strains,<sup>70</sup> further investigation is warranted to investigate this complex regulatory region.

Given the likely importance of viral infection in the pathogenesis of multiple sclerosis,<sup>51</sup> especially the role of EBV,<sup>71</sup> it seemed logical to check our samples for the presence of viruses. While we did not identify an enrichment of any one virus in the CSF of patients with multiple sclerosis, patients with multiple sclerosis did have an increased proportion of a rare CD8+ T cell population where the top defining markers, including the inhibitory receptors HAVCR2 and TIGIT, overlap previously described viral specific exhausted-like CD8+ cells,<sup>72,73</sup> which are known to arise following chronic antigenic stimulation.<sup>74</sup> The expression of HAVCR2 and TIGIT in CD8+ T cells however is also present on several other CD8+ T cell phenotypes, including activated cells,<sup>75</sup> and tissue resident memory cells.<sup>76</sup> Given that exhausted T cells<sup>77</sup> and CD8+ tissue resident memory cells<sup>7,8,78</sup> have all been implicated in multiple sclerosis, careful phenotyping of this cell population at the epigenetic and protein level will be essential to establish the function of these cells and in turn the mechanism by which they may be of relevance in the disease process.

Using a MOFA approach, we identified several factors showing altered expression in multiple sclerosis. The most statistically significant being the decreased activity of genes controlling anti-inflammatory and anti-viral type I IFN responses in the T cells. The genes with the highest MOFA weighting for this factor also overlap with those characterizing tissue-resident cytotoxic T cell effector memory<sup>6,37</sup> or T<sub>EMRA</sub> cells.<sup>38</sup> This finding is consistent with previous work, emphasizing the importance in multiple sclerosis of specific T helper cell subsets,<sup>79</sup> alongside the protective effects of type I IFNs in CNS autoimmunity and the known benefit of using the type I IFN—IFN $\beta$ —as a treatment in multiple sclerosis.<sup>80</sup> Whether our observation that these genes show reduced expression in multiple sclerosis reflects a depletion of a specific cell subset or rather reflects a shift in function across a wider range of CD4+ T cells remains to be established. The limited ability to resolve cell subtypes in unstimulated CD4+ T single-cell data is well established.<sup>81</sup> While reference sets have proven useful in partially overcoming this issue,<sup>82</sup> development of disease-relevant reference sets will be required to allow meaningful cross-study comparisons of disease-specific effects within T cell subtypes.

As well as changes in T cells we also identified two disease-relevant MOFA factors in the macrophage population, one related to anti-inflammatory M2 polarization and the other related to antigen presentation. Macrophages have been shown to be of importance in the mouse model of multiple sclerosis—experimental autoimmune encephalomyelitis<sup>83</sup>—human multiple sclerosis pathology,<sup>84</sup> disease progression,<sup>85</sup> remyelination<sup>86</sup> and in the neuro-inflammation present in many degenerative neurological diseases.<sup>87</sup> Our MOFA data indicate that in multiple sclerosis there

is reduced polarization towards the M2 phenotype and increased expression of the machinery of antigen presentation, likely indicating that the MOFA has identified factors indicative of an altered phenotypic cell state in multiple sclerosis.

Given the extreme functional plasticity of immune cells in response to environmental stimuli, it is difficult to disentangle whether the phenotypes we have captured represent the cause or the consequence of disease. However, our single cell analysis has identified genes and pathways critical in both adaptive and innate immune cells, emphasizing the importance of both arms of the immune system in multiple sclerosis. Furthermore, our finding that the multiple sclerosis susceptibility variants rs10271373 and rs1059091 are eQTLs for antiviral genes in CSF CD8+ T cells supports the hypothesis that dysregulation in viral control mechanisms is involved in the development of multiple sclerosis.

## Data availability

These data have been deposited in the European Genome-phenome Archive (EGA), which is hosted by the EBI and the CRG, under accession number EGAS00001007478. For reproducibility, the data preprocessing scripts and analysis notebooks can be found at <https://github.com/huangyh09/MSclerosisSrc>. For the eQTL analysis, we used the standard pipeline at: [https://github.com/single-cell-genetics/limix\\_qtl](https://github.com/single-cell-genetics/limix_qtl).

## Acknowledgements

We thank all the volunteers for their participation in the study and the staff on the University of Cambridge Department of Clinical Neuroscience Programmed Investigation Unit for all their help.

## Funding

This work was supported by funding from the Multiple Sclerosis Society in the UK, grant reference 43. We acknowledge support from the NIHR Cambridge Biomedical Research Centre. D.B. was supported by the EMBL International PhD Programme and Darwin Trust Fellowship. Y.H. was supported by the EBPOD fellowship from the University of Cambridge and EMBL-EBI. E.M. was supported by the ESPOD fellowship from the Wellcome Sanger Institute and EMBL-EBI. Core support grants by Wellcome and Medical Research Council (MRC) to the Wellcome-MRC Cambridge Stem Cell Institute 203151/Z/16/Z. Work in the Gottgens Laboratory is funded by grants from Wellcome (206328/Z/17/Z), Blood Cancer UK (18002). This research was funded in whole, or in part, by the Wellcome Trust. For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

## Competing interests

O.S. is a paid advisor of Insitro. Inc. The remaining authors report no competing interests.

## Supplementary material

Supplementary material is available at *Brain* online.

## References

1. International Multiple Sclerosis Genetics Consortium. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science*. 2019;365:eaav7188.
2. Deisenhammer F, Zetterberg H, Fitzner B, Zettl UK. The cerebrospinal fluid in multiple sclerosis. *Front Immunol*. 2019;10:726.
3. Alvermann S, Hennig C, Stuve O, Wiendl H, Stangel M. Immunophenotyping of cerebrospinal fluid cells in multiple sclerosis: In search of biomarkers. *JAMA Neurol*. 2014;71:905–912.
4. Han S, Lin YC, Wu T, et al. Comprehensive immunophenotyping of cerebrospinal fluid cells in patients with neuroimmunological diseases. *J Immunol*. 2014;192:2551–2563.
5. Schafflick D, Xu CA, Hartlehnert M, et al. Integrated single cell analysis of blood and cerebrospinal fluid leukocytes in multiple sclerosis. *Nat Commun*. 2020;11:247.
6. Pappalardo JL, Zhang L, Pecsok MK, et al. Transcriptomic and clonal characterization of T cells in the human central nervous system. *Sci Immunol*. 2020;5:eabb8786.
7. Ostkamp P, Deffner M, Schulte-Mecklenbeck A, et al. A single-cell analysis framework allows for characterization of CSF leukocytes and their tissue of origin in multiple sclerosis. *Sci Transl Med*. 2022;14:eadc9778.
8. Beltrán E, Gerdes LA, Hansen J, et al. Early adaptive immune activation detected in monozygotic twins with prodromal multiple sclerosis. *J Clin Invest*. 2019;129:4758–4768.
9. Thompson AJ, Banwell BL, Barkhof F, et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *Lancet Neurol*. 2018;17:162–173.
10. McCarthy S, Das S, Kretzschmar W, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet*. 2016;48:1279–1283.
11. Deelen P, Bonder M, van der Velde K, et al. Genotype harmonizer: Automatic strand alignment and format conversion for genotype data integration. *BMC Res Notes*. 2014;7:901.
12. Lun ATL, Riesenfeld S, Andrews T, et al. EmptyDrops: Distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. *Genome Biol*. 2019;20:63.
13. Huang X, Huang Y. Cellsnr-lite: An efficient tool for genotyping single cells. *Bioinformatics*. 2021;37:4569–44571.
14. Huang Y, McCarthy DJ, Stegle O. Vireo: Bayesian demultiplexing of pooled single-cell RNA-Seq data without genotype reference. *Genome Biol*. 2019;20:273.
15. McGinnis CS, Murrow LM, Gartner ZJ. DoubletFinder: Doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst*. 2019;8:329–337.e4.
16. Wolock SL, Lopez R, Klein AM. Scrublet: Computational identification of cell doublets in single-cell transcriptomic data. *Cell Syst*. 2019;8:281–291.e9.
17. Young MD, Behjati S. SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *Gigascience*. 2020;9:giaa151.
18. Wolf FA, Angerer P, Theis FJ. SCANPY: Large-scale single-cell gene expression data analysis. *Genome Biol*. 2018;19:15.
19. Polański K, Young MD, Miao Z, Meyer KB, Teichmann SA, Park J-E. BBKNN: Fast batch alignment of single cell transcriptomes. *Bioinformatics*. 2020;36:964–965.
20. Argelaguet R, Arnol D, Bredikhin D, et al. MOFA+: A statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biol*. 2020;21:111.
21. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102:15545–15550.

22. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics*. 2011;27:1739-1740.
23. Garcia-Alonso L, Holland CH, Ibrahim MM, Turei D, Saez-Rodriguez J. Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Res*. 2019;29:1363-1375.
24. Khan A, Liu Q, Chen X, et al. Detection of human papillomavirus in cases of head and neck squamous cell carcinoma by RNA-seq and VirTect. *Mol Oncol*. 2019;13:829-839.
25. Vargiu L, Rodriguez-Tomé P, Sperber GO, et al. Classification and characterization of human endogenous retroviruses; mosaic forms are common. *Retrovirology*. 2016;13:7.
26. Kaminow B, Yunusov D, Dobin A. STARsolo: accurate, fast and versatile mapping/quantification of single-cell and single-nucleus RNA-seq data. *bioRxiv*. [Preprint] <https://doi.org/10.1101/2021.05.05.442755>.
27. Cuomo ASE, Alvani G, Azodi CB, single-cell eQTLGen consortium, McCarthy DJ, Bonder MJ. Optimizing expression quantitative trait locus mapping workflows for single-cell studies. *Genome Biol*. 2021;22:188.
28. Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-Seq data using regularized negative binomial regression. *Genome Biol*. 2019;20:296.
29. Vösa U, Claringbould A, Westra H-J, et al. Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat Genet*. 2021;53:1300-1310.
30. Wang G, Sarkar A, Carbonetto P, Stephens M. A simple new approach to variable selection in regression, with application to genetic fine mapping. *Journal of the Royal Statistical Society Series B: Statistical Methodology*. 2020;82:1273-1300.
31. Huang Y, Sanguinetti G. BRIE2: Computational identification of splicing phenotypes from single-cell transcriptomic experiments. *Genome Biol*. 2021;22:251.
32. Polman CH, O'Connor PW, Havrdova E, et al. A randomized, placebo-controlled trial of natalizumab for relapsing multiple sclerosis. *N Engl J Med*. 2006;354:899-910.
33. Wei L, Li J, Han Z, Chen Z, Zhang Q. Silencing of lncRNA MALAT1 prevents inflammatory injury after lung transplant ischemia-reperfusion by downregulation of IL-8 via p300. *Mol Ther Nucleic Acids*. 2019;18:285-297.
34. Zhang Q, Chen C-Y, Yedavalli VSRK, Jeang K-T. NEAT1 long non-coding RNA and paraspeckle bodies modulate HIV-1 posttranscriptional expression. *mBio*. 2013;4:e00596-12.
35. Argelaguet R, Velten B, Arnol D, et al. Multi-omics factor analysis—A framework for unsupervised integration of multi-omics data sets. *Mol Syst Biol*. 2018;14:e8124.
36. Parts L, Stegle O, Winn J, Durbin R. Joint genetic analysis of gene expression data with inferred cellular phenotypes. *PLoS Genet*. 2011;7:e1001276.
37. Szabo PA, Levitin HM, Miron M, et al. Single-cell transcriptomics of human T cells reveals tissue and activation signatures in health and disease. *Nat Commun*. 2019;10:4706.
38. Wang X, Shen X, Chen S, et al. Reinvestigation of classic T cell subsets and identification of novel cell subpopulations by single-cell RNA sequencing. *J Immunol*. 2022;208:396-406.
39. Thummler K, Leipe J, Rammung A, Schulze-Koops H, Skapenko A. Immune regulation by peripheral suppressor T cells induced upon homotypic T cell/T cell interactions. *J Leukoc Biol*. 2010;88:1041-1050.
40. Drashansky TT, Helm E, Huo Z, et al. Bcl11b prevents fatal autoimmunity by promoting T<sub>reg</sub> cell program and constraining innate lineages in T<sub>reg</sub> cells. *Sci Adv*. 2019;5:eaaw0480.
41. Vanvalkenburgh J, Albu DI, Bapanpally C, et al. Critical role of Bcl11b in suppressor function of T regulatory cells and prevention of inflammatory bowel disease. *J Exp Med*. 2011;208:2069-2081.
42. de Jesús TJ, Ramakrishnan P. NF-κB c-Rel dictates the inflammatory threshold by acting as a transcriptional repressor. *iScience*. 2020;23:100876.
43. Ridley ML, Fleskens V, Roberts CA, et al. IKZF3/Aiolos is associated with but not sufficient for the expression of IL-10 by CD4(+) T cells. *J Immunol*. 2020;204:2940-2948.
44. Jefferies CA. Regulating IRFs in IFN driven disease. *Front Immunol*. 2019;10:325.
45. Song G, Shi Y, Zhang M, et al. Global immune characterization of HBV/HCV-related hepatocellular carcinoma identifies macrophage and T-cell subsets associated with disease progression. *Cell Discov*. 2020;6:90.
46. Hu Z-W, Wen Y-H, Ma R-Q, et al. Ferroptosis driver SOCS1 and suppressor FTH1 independently correlate with M1 and M2 macrophage infiltration in head and neck squamous cell carcinoma. *Front Cell Dev Biol*. 2021;9:727762.
47. Mahajan S, Saini A, Chandra V, et al. Nuclear receptor Nr4a2 promotes alternative polarization of macrophages and confers protection in sepsis. *J Biol Chem*. 2015;290:18304-18314.
48. Wu X, Gao H, Ke W, Giese RW, Zhu Z. The homeobox transcription factor VentX controls human macrophage terminal differentiation and proinflammatory activation. *J Clin Invest*. 2011;121:2599-2613.
49. Krausgruber T, Blazek K, Smallie T, et al. IRF5 promotes inflammatory macrophage polarization and TH1-TH17 responses. *Nat Immunol*. 2011;12:231-238.
50. Li H, Zhang Y, Zheng S. Comprehensive analysis identified ETV7 as a potential prognostic biomarker in bladder cancer. *Biomed Res Int*. 2021;2021:8530186.
51. Virtanen J, Jacobson S. Viruses and multiple sclerosis. *CNS Neurol Disord-Dr*. 2012;11:528-544.
52. Brettschneider J, Tumani H, Kiechle U, et al. Igg antibodies against measles, rubella, and varicella zoster virus predict conversion to multiple sclerosis in clinically isolated syndrome. *PLoS One*. 2009;4:e7638.
53. Lower R, Lower J, Kurth R. The viruses in all of US: Characteristics and biological significance of human endogenous retrovirus sequences. *Proc Natl Acad Sci U S A*. 1996;93:5177-5184.
54. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001;409:860-921.
55. She J, Du M, Xu Z, et al. The landscape of hervRNAs transcribed from human endogenous retroviruses across human body sites. *Genome Biol*. 2022;23:231.
56. Morris G, Maes M, Murdjeva M, Puri BK. Do human endogenous retroviruses contribute to multiple sclerosis, and if so, how? *Mol Neurobiol*. 2019;56:2590-2605.
57. Oelen R, de Vries DH, Brugge H, et al. Single-cell RNA-Sequencing of peripheral blood mononuclear cells reveals widespread, context-specific gene expression regulation upon pathogenic exposure. *Nat Commun*. 2022;13:3267.
58. Lepik K, Annilo T, Kukuskina V, et al. C-reactive protein upregulates the whole blood expression of CD59—An integrative analysis. *PLoS Comput Biol*. 2017;13:e1005766.
59. International Multiple Sclerosis Genetics Consortium; MultipleMS Consortium. Locus for severity implicates CNS resilience in progression of multiple sclerosis. *Nature*. 2023;619:323-331.
60. Schmiedel BJ, Singh D, Madrigal A, et al. Impact of genetic polymorphisms on human immune cell gene expression. *Cell*. 2018;175:1701-1715.e16.

61. Li MMH, Aguilar EG, Michailidis E, et al. Characterization of novel splice variants of zinc finger antiviral protein (ZAP). *J Virol.* 2019;93:e00715–19.
62. Kerns JA, Emerman M, Malik HS. Positive selection and increased antiviral activity associated with the PARP-containing isoform of human zinc-finger antiviral protein. *PLoS Genet.* 2008;4:e21.
63. Zimmer MM, Kibe A, Rand U, et al. The short isoform of the host antiviral protein ZAP acts as an inhibitor of SARS-CoV-2 programmed ribosomal frameshifting. *Nat Commun.* 2021;12:7193.
64. Zhu Y, Chen G, Lv F, et al. Zinc-finger antiviral protein inhibits HIV-1 infection by selectively targeting multiply spliced viral mRNAs for degradation. *Proc Natl Acad Sci U S A.* 2011;108:15834–15839.
65. Gonzalez-Perez AC, Stempel M, Wyler E, et al. The zinc finger antiviral protein ZAP restricts human cytomegalovirus and selectively binds and destabilizes viral UL4/UL5 transcripts. *mBio.* 2021;12:e02683–20.
66. Charron G, Li MM, MacDonald MR, Hang HC. Prenylome profiling reveals S-farnesylation is crucial for membrane targeting and antiviral activity of ZAP long-isoform. *Proc Natl Acad Sci U S A.* 2013;110:11085–11090.
67. Afrasiabi A, Fewings NL, Schibeci SD, et al. The interaction of human and Epstein-Barr virus miRNAs with multiple sclerosis risk loci. *Int J Mol Sci.* 2021;22:2927.
68. Narayana SK, Helbig KJ, McCartney EM, et al. The interferon-induced transmembrane proteins, IFITM1, IFITM2, and IFITM3 inhibit hepatitis C virus entry. *J Biol Chem.* 2015;290:25946–25959.
69. Yanez DC, Sahni H, Ross S, et al. IFITM proteins drive type 2 T helper cell differentiation and exacerbate allergic airway inflammation. *Eur J Immunol.* 2019;49:66–78.
70. Wu W-L, Grotefend CR, Tsai M-T, et al. Delta20 IFITM2 differentially restricts X4 and R5 HIV-1. *Proc Natl Acad Sci U S A.* 2017;114:7112–7117.
71. Bjornevik K, Cortese M, Healy BC, et al. Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. *Science.* 2022;375:296–301.
72. Galletti G, De Simone G, Mazza EMC, et al. Two subsets of stem-like CD8(+) memory T cell progenitors with distinct fate commitments in humans. *Nat Immunol.* 2020;21:1552–1562.
73. Kusnadi A, Ramirez-Suastegui C, Fajardo V, et al. Severely ill COVID-19 patients display impaired exhaustion features in SARS-CoV-2-reactive CD8<sup>+</sup> T cells. *Sci Immunol.* 2021;6:eabe4782.
74. McLane LM, Abdel-Hakeem MS, Wherry EJ. CD8 T cell exhaustion during chronic viral infection and cancer. *Annu Rev Immunol.* 2019;37:457–495.
75. Gorman JV, Starbeck-Miller G, Pham N-LL, et al. Tim-3 directly enhances CD8 T cell responses to acute *Listeria monocytogenes* infection. *J Immunol.* 2014;192:3133–3142.
76. Clarke J, Panwar B, Madrigal A, et al. Single-cell transcriptomic analysis of tissue-resident memory T cells in human lung cancer. *J Exp Med.* 2019;216:2128–2149.
77. Pender MP, Csurhes PA, Pfluger CMM, Burrows SR. CD8 T cell deficiency impairs control of Epstein-Barr virus and worsens with age in multiple sclerosis: Figure 1. *J Neurol Neurosurg Psychiatry.* 2012;83:353–354.
78. Franssen NL, Hsiao C-C, van der Poel M, et al. Tissue-resident memory T cells invade the brain parenchyma in multiple sclerosis white matter lesions. *Brain.* 2020;143:1714–1730.
79. Kaskow BJ, Baecher-Allan C. Effector T cells in multiple sclerosis. *Cold Spring Harb Perspect Med.* 2018;8:a029025.
80. Reder AT, Feng X. How type I interferons work in multiple sclerosis and other diseases: Some unexpected mechanisms. *J Interferon Cytokine Res.* 2014;34:589–599.
81. Ding J, Smith SL, Orozco G, Barton A, Eyre S, Martin P. Characterisation of CD4<sup>+</sup> T-cell subtypes using single cell RNA sequencing and the impact of cell number and sequencing depth. *Sci Rep.* 2020;10:19825.
82. Andreatta M, Corria-Osorio J, Muller S, Cubas R, Coukos G, Carmona SJ. Interpretation of T cell states from single-cell transcriptomics data using reference atlases. *Nat Commun.* 2021;12:2965.
83. Leuti A, Talamonti E, Gentile A, et al. Macrophage plasticity and polarization are altered in the experimental model of multiple sclerosis. *Biomolecules.* 2021;11:837.
84. Kuhlmann T, Ludwin S, Prat A, Antel J, Brück W, Lassmann H. An updated histological classification system for multiple sclerosis lesions. *Acta Neuropathol.* 2017;133:13–24.
85. Kamma E, Lasisi W, Libner C, Ng HS, Plemel JR. Central nervous system macrophages in progressive multiple sclerosis: Relationship to neurodegeneration and therapeutics. *J Neuroinflammation.* 2022;19:45.
86. Miron VE, Boyd A, Zhao J-W, et al. M2 microglia and macrophages drive oligodendrocyte differentiation during CNS remyelination. *Nat Neurosci.* 2013;16:1211–1218.
87. Mamma S, Fagone P, Cavalli E, et al. The role of macrophages in neuroinflammatory and neurodegenerative pathways of Alzheimer's disease, amyotrophic lateral sclerosis, and multiple sclerosis: Pathogenetic cellular effectors and potential therapeutic targets. *Int J Mol Sci.* 2018;19:831.